## Deadlines

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Date of assignment</th>
<th>Deadline (midnight 23:59)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HW1</td>
<td>Sep 7</td>
<td>Sep 20</td>
</tr>
<tr>
<td>HW2</td>
<td>Sep 21</td>
<td>Oct 4</td>
</tr>
<tr>
<td>HW3</td>
<td>Oct 5</td>
<td>Oct 18</td>
</tr>
<tr>
<td>Paper summary</td>
<td>Oct 12</td>
<td>Oct 26</td>
</tr>
<tr>
<td>HW4</td>
<td>Oct 19</td>
<td>Nov 1</td>
</tr>
<tr>
<td>HW5</td>
<td>Nov 9</td>
<td>Nov 22</td>
</tr>
<tr>
<td>HW6</td>
<td>Nov 23</td>
<td>Dec 6</td>
</tr>
<tr>
<td>Project</td>
<td><strong>Oct 5 - 7</strong></td>
<td>Dec 14 - 16</td>
</tr>
</tbody>
</table>

Team formation deadline is **Oct 19**
Recap from last time

Supervised Learning

Deep Learning

Unsupervised Learning

Reinforcement Learning
Convolutional Neural Networks handle natural variety in the data

scaling  rotation  original

CNN  →  X

CNN  →  O
Convolutional Neural Networks
compare smaller patterns

This helps to quantify the similarity between slightly different images
CNN Dataset

Batch (size = 4)

Learning Rate

Architecture

Optimizer

Validation

<table>
<thead>
<tr>
<th></th>
<th>Cat</th>
<th>Dog</th>
<th>True</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch #1</td>
<td>0.1</td>
<td>0.9</td>
<td>Dog</td>
<td>0.15</td>
</tr>
<tr>
<td>Batch #2</td>
<td>0.87</td>
<td>0.13</td>
<td>Cat</td>
<td>0.2</td>
</tr>
<tr>
<td>Batch #3</td>
<td>0.75</td>
<td>0.25</td>
<td>Cat</td>
<td>0.41</td>
</tr>
<tr>
<td>Average total</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Training error

<table>
<thead>
<tr>
<th>Batch #1</th>
<th>Batch #2</th>
<th>Batch #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.84</td>
<td>0.82</td>
<td>0.42</td>
</tr>
</tbody>
</table>

https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a
During one epoch we train through the entire training set.

**Epoch**
Epoch #1

Aver. training loss = 0.67
Validation loss = 0.42

Epoch #2

Aver. training loss = 0.54
Validation loss = 0.41

Epoch #3

Aver. training loss = 0.33
Validation loss = 0.37
Epoch #1
Aver. training loss = 0.67
Validation loss = 0.42

Epoch #2
Aver. training loss = 0.54
Validation loss = 0.41

Epoch #3
Aver. training loss = 0.33
Validation loss = 0.37

Epoch #4
Aver. training loss = 0.17
Validation loss = 0.31

Epoch #5
Aver. training loss = 0.09
Validation loss = 0.29
Epoch #5

How to fix this?

Overfitting
There are two main ways of looking at this (A and B)

Epoch #5

Overfitting
There are two main ways of looking at this (A and B)

**way A:** something wrong with a model
There are two main ways of looking at this (A and B):

way A: something wrong with a model

way B: something wrong with a data
There are two main ways of looking at this (A and B).

**way A:** something wrong with a model

**way B:** something wrong with the data

Dataset

Epoch #5

loss

Overfitting
Regularisation methods

Way A methods

Way B methods
Regularisation methods

Way A methods

Way B methods

*These methods are applicable to most ML models not only Deep Neural Networks
Regularisation methods

Way A methods

Way B methods
Loss functions we encountered
Linear Regression

![Graph showing a scatter plot with an overlay of a linear regression line. The x-axis represents the independent variable, and the y-axis represents the dependent variable. The data points are scattered around the line, indicating a linear relationship.]
We want to find a line such that … 

… it minimises the sum of errors 

$$\arg \min = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
Loss functions we encountered

\[ RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

real value

predicted value

Linear Regression
Point coordinates as input

(X, Y)

Input

Hidden Layer

Output Layer

Class Scores

Total Error

Biases

Expected Score

Class 1 Score

Error 1

Total error

Class 2 Score

Error 2

0.0

1.0
### Input

Point coordinates as input

\((X, Y)\)

### Hidden Layer

\(b\)

### Output Layer

\(b\)

### Class Scores

**Expected Score**

- **Class 1 Score**
  - Error 1
  - Total error

- **Class 2 Score**
  - Error 2

### Total Error

\[ E_i = \frac{1}{2} \left( \text{truth} - \text{out}_{oi} \right)^2 \]
Loss functions we encountered

\[ RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

real value
predicted value

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

real value
predicted value

Linear Regression

Neural Network
(Multi Layer Perceptron)
batch #1

Dataset

Batch (size = 4)

CNN

Learning Rate

Architecture

Optimizer

Validation

<table>
<thead>
<tr>
<th>Cat</th>
<th>Dog</th>
<th>True</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.84</td>
<td>0.16</td>
<td>Cat</td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>0.35</td>
<td>Cat</td>
<td></td>
</tr>
<tr>
<td>0.54</td>
<td>0.46</td>
<td>Dog</td>
<td></td>
</tr>
<tr>
<td>0.38</td>
<td>0.62</td>
<td>Cat</td>
<td></td>
</tr>
</tbody>
</table>
Binary cross-entropy error

https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a
Dataset

Batch
(size = 4)

CNN

Validation

Learning Rate
Architecture
Optimizer

<table>
<thead>
<tr>
<th>Cat</th>
<th>Dog</th>
<th>True</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.84</td>
<td>0.16</td>
<td>Cat</td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>0.35</td>
<td>Cat</td>
<td></td>
</tr>
<tr>
<td>0.54</td>
<td>0.46</td>
<td>Dog</td>
<td></td>
</tr>
<tr>
<td>0.38</td>
<td>0.62</td>
<td>Cat</td>
<td></td>
</tr>
</tbody>
</table>

Binary cross-entropy error

\[ \text{Error} = - (y \log(p) + (1 - y) \log(1 - p)) \]

1 for cat & 0 for dog

Probability of class dog

Probability of class cat

https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a
Loss functions we encountered

\[
RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

\[
BCE = -(y \log(p) + (1 - y) \log(1 - p))
\]
Loss functions we encountered

- **Residual Sum of Squares**
  \[ RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

- **Mean Squared Error**
  \[ MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

- **Binary Cross-entropy**
  \[ BCE = - (y \log(p) + (1 - y) \log(1 - p)) \]

Loss (error) functions tell us **how far** we are from the optimal model.
Linear Regression (example)
Linear Regression (example)

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

Line will pass through this point

\[ \bar{x} = 3 \quad \bar{y} = 4 \]

*This follows from the fact that sum of errors has to be zero, more details: https://math.stackexchange.com/questions/494181/why-the-sum-of-residuals-equals-0-when-we-do-a-sample-regression-by-ols*
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

\[ \bar{x} = 3 \quad \bar{y} = 4 \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

\( \bar{x} = 3 \quad \bar{y} = 4 \)
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x - \bar{x}</td>
<td>y - \bar{y}</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

\[ \bar{x} = 3 \quad \bar{y} = 4 \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

\[ \bar{x} = 3 \quad \bar{y} = 4 \]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

\( \bar{x} = 3 \quad \bar{y} = 4 \)
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

\[ x \quad y \quad x - \bar{x} \quad y - \bar{y} \quad (x - \bar{x})^2 \]

\[
\begin{array}{c|c|c|c|c}
 x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 \\
1 & 2 & -2 & -2 & \\
2 & 4 & -1 & 0 & \\
3 & 5 & 0 & 1 & \\
4 & 4 & 1 & 0 & \\
5 & 5 & 2 & 1 & \\
\end{array}
\]

\[ \bar{x} = 3 \quad \bar{y} = 4 \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

\[ \bar{x} = 3 \quad \bar{y} = 4 \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

\[
\begin{array}{c|cc|cc|cc|cc}
 x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
 1 & 2 & -2 & -2 & 4 & 4 \\
 2 & 4 & -1 & 0 & 1 & 0 \\
 3 & 5 & 0 & 1 & 0 & 0 \\
 4 & 4 & 1 & 0 & 1 & 0 \\
 5 & 5 & 2 & 1 & 4 & 2 \\
\end{array}
\]

\[ \bar{x} = 3 \quad \bar{y} = 4 \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

\[
\begin{array}{c|c|c|c|c|c|c|c|c}
 x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
1 & 2 & -2 & -2 & 4 & 4 \\
2 & 4 & -1 & 0 & 1 & 0 \\
3 & 5 & 0 & 1 & 0 & 0 \\
4 & 4 & 1 & 0 & 1 & 0 \\
5 & 5 & 2 & 1 & 4 & 2 \\
\end{array}
\]

\[ \bar{x} = 3 \quad \bar{y} = 4 \]

\[ w_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \]
\[
\hat{y} = w_0 + w_1 x
\]

Recall lecture about **PCA**…

\[
w_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}
\]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

Recall lecture about PCA…

What is this?

\[ w_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \]
How to **interpret** values in **covariance** matrix?

\[
[-2, -1, 0, 1, 2] \quad [-2, 0, 0, 1, 1]
\]

\[
\bar{x} = 0 \quad \bar{y} = 0
\]

\[
cov(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}
\]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

Recall lecture about **PCA**…

\[ \text{Cov}(x,y) \quad \text{covariance} \]

\[ w_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

Recall lecture about PCA…

Covariance:

\[ \text{Cov}(x,y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \]

What about this?
How to interpret values in covariance matrix?

$$\Sigma (x_i - \bar{x})^2 \over n-1$$

Variance: $\sigma = \frac{\sum (x_i - \bar{x})^2}{n - 1}$
\[ \hat{y} = w_0 + w_1 x \]

Recall lecture about **PCA**

\[ w_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \]

\[ \text{Cov}(x,y) \text{ - covariance} \]

\[ \text{Var}(x) \text{ - variance} \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

\[ w_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

\( \bar{x} = 3 \quad \bar{y} = 4 \)
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

\[
\begin{array}{c|c|c|c|c|c|c}
 x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
 1 & 2 & -2 & -2 & 4 & 4 \\
 2 & 4 & -1 & 0 & 1 & 0 \\
 3 & 5 & 0 & 1 & 0 & 0 \\
 4 & 4 & 1 & 0 & 1 & 0 \\
 5 & 5 & 2 & 1 & 4 & 2 \\
\end{array}
\]

\[ \bar{x} = 3 \quad \bar{y} = 4 \quad 10 \]

\[ w_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1x \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - ( \bar{x} )</th>
<th>y - ( \bar{y} )</th>
<th>((x - \bar{x})^2)</th>
<th>((x - \bar{x})(y - \bar{y}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

\( \bar{x} = 3\)  \( \bar{y} = 4\)

\[ w_1 = \frac{6}{\sum (x - \bar{x})^2} \]
Linear Regression (example)

\[ \hat{y} = w_0 + w_1 x \]

\[ \bar{x} = 3 \quad \bar{y} = 4 \]

\[ w_1 = \frac{6}{10} \]
Linear Regression (example)

\[ \hat{y} = w_0 + 0.6x \]

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x - (\bar{x})</td>
<td>y - (\bar{y})</td>
<td>(x - (\bar{x}))^2</td>
<td>(x - (\bar{x}))(y - (\bar{y}))</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

\(\bar{x} = 3\) \(\bar{y} = 4\)

\[ w_1 = 0.6 \]
Linear Regression (example)

\[ \hat{y} = w_0 + 0.6x \]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 3 \quad \bar{y} = 4 \]

\[ w_1 = 0.6 \]
Linear Regression (example)

$$\hat{y} = w_0 + 0.6x$$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
<th>$x - \bar{x}$</th>
<th>$y - \bar{y}$</th>
<th>$(x - \bar{x})^2$</th>
<th>$(x - \bar{x})(y - \bar{y})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

$\bar{x} = 3$ \hspace{1cm} $\bar{y} = 4$

$$w_1 = 0.6$$

$$\hat{y} = w_0 + 0.6 \hat{x}$$
Linear Regression (example)

\[ \hat{y} = w_0 + 0.6 x \]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 3 \quad \bar{y} = 4 \]

\[ w_1 = 0.6 \]

\[ \hat{y} = w_0 + 0.6 \hat{x} \]
Linear Regression (example)

\[ \hat{y} = w_0 + 0.6x \]

\[
\begin{array}{cccccc}
  x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
  1 & 2 & -2 & -2 & 4 & 4 \\
  2 & 4 & -1 & 0 & 1 & 0 \\
  3 & 5 & 0 & 1 & 0 & 0 \\
  4 & 4 & 1 & 0 & 1 & 0 \\
  5 & 5 & 2 & 1 & 4 & 2 \\
\end{array}
\]

\[ \bar{x} = 3 \quad \bar{y} = 4 \]

\[ w_1 = 0.6 \]

\[ 4 = w_0 + 0.6 \hat{x} \]
Linear Regression (example)

\[ \hat{y} = w_0 + 0.6x \]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 3 \quad \bar{y} = 4 \]

\[ w_1 = 0.6 \]

\[ 4 = w_0 + 0.6 \hat{x} \]
Linear Regression (example)

\[ \hat{y} = w_0 + 0.6 \times x \]

![Graph showing linear regression example]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 3 \quad \bar{y} = 4 \]

\[ w_1 = 0.6 \]

\[ 4 = w_0 + 0.6 \times 3 \]
Linear Regression (example)

\[ \hat{y} = w_0 + 0.6x \]

\[ x \quad y \quad x - \bar{x} \quad y - \bar{y} \quad (x - \bar{x})^2 \quad (x - \bar{x})(y - \bar{y}) \]

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

\( \bar{x} = 3 \) \quad \( \bar{y} = 4 \)

\[ w_1 = 0.6 \]

\[ 4 = w_0 + 0.6 \times 3 \]
Linear Regression (example)

\[ \hat{y} = w_0 + 0.6x \]

\[ w_1 = 0.6 \]

\[ 4 = w_0 + 1.8 \]
Linear Regression (example)

\[ \hat{y} = w_0 + 0.6x \]

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
\hline
1 & 2 & -2 & -2 & 4 & 4 \\
2 & 4 & -1 & 0 & 1 & 0 \\
3 & 5 & 0 & 1 & 0 & 0 \\
4 & 4 & 1 & 0 & 1 & 0 \\
5 & 5 & 2 & 1 & 4 & 2 \\
\hline
\end{array}
\]

\( \bar{x} = 3 \quad \bar{y} = 4 \)

\begin{align*}
\hat{w}_1 &= 0.6 \\
\hat{w}_0 &= 4 - 1.8
\end{align*}
Linear Regression (example)

\[ \hat{y} = w_0 + 0.6x \]

\[ \bar{x} = 3 \quad \bar{y} = 4 \]

\[ w_1 = 0.6 \]

\[ w_0 = 2.2 \]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \times \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>x - ( \bar{x} )</th>
<th>y - ( \bar{y} )</th>
<th>(x - ( \bar{x} ))^2</th>
<th>(x - ( \bar{x} ))(y - ( \bar{y} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

\( \bar{x} = 3 \) \( \bar{y} = 4 \)

\[ w_1 = 0.6 \]

\[ w_0 = 2.2 \]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6x \]

What is error function Linear Regression optimises?
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \, x \]

What is error function Linear Regression optimises?

RSS
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \, x \]

What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[ RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6x \]

What is Residual Sum of Squares (RSS) for this line?

\[ RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \, x \]

What is Residual Sum of Squares (RSS) for this line?

\[ RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]
What is Residual Sum of Squares (RSS) for this line?

\[
\text{RSS} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2
\]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[ RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 \]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 =
\]

\[
= (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2
\]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \, x \]

What is Residual Sum of Squares (RSS) for this line?

\[ RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = \]

\[ = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 \]
Linear Regression (example)

\[
\hat{y} = 2.2 + 0.6x
\]

What is Residual Sum of Squares (RSS) for this line?

\[
\text{RSS} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = (2 - \hat{y}_1)^2 + (2 - \hat{y}_2)^2 + (3 - \hat{y}_3)^2 + (4 - \hat{y}_4)^2 + (5 - \hat{y}_5)^2
\]
What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = (2 - \hat{y}_1)^2 + (2 - \hat{y}_2)^2 + (4 - \hat{y}_3)^2 + (5 - \hat{y}_4)^2 + (5 - \hat{y}_5)^2
\]

\[
\hat{y}_1 = ?
\]
We should use the model \( \hat{y} = 2.2 + 0.6x \)

What is Residual Sum of Squares (RSS) for this line?

\[
\text{RSS} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = \]

\[
= (2 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2
\]

\( \hat{y}_1 = ? \)
We should use the model \( \hat{y} = 2.2 + 0.6 \times x \)

What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = (2 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2
\]
We should use the model $\hat{y} = 2.2 + 0.6 \times x$

What is Residual Sum of Squares (RSS) for this line?

$$RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = (2 - \hat{y}_1)^2 + (3 - \hat{y}_2)^2 + (4 - \hat{y}_3)^2 + (5 - \hat{y}_4)^2 + (5 - \hat{y}_5)^2$$

$$\hat{y}_1 = 2.2 + 0.6 \times 1$$
We should use the model \( \hat{y} = 2.2 + 0.6 \times x \)

What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = (2 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2
\]

\( \hat{y}_1 = 2.2 + 0.6 \)
We should use the model $\hat{y} = 2.2 + 0.6 \times x$

What is Residual Sum of Squares (RSS) for this line?

$RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = (2 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2$

$\hat{y}_1 = 2.8$
We should use the model \( \hat{y} = 2.2 + 0.6 \times x \)

What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = (2 - 2.8)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2
\]

\( \hat{y}_1 = 2.8 \)
We should use the model $\hat{y} = 2.2 + 0.6 \times x$

What is Residual Sum of Squares (RSS) for this line?

$$RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = (2 - 2.8)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2$$
We should use the model $\hat{y} = 2.2 + 0.6 \times x$

What is Residual Sum of Squares (RSS) for this line?

$$RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = (-0.8)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2$$
We should use the model $\hat{y} = 2.2 + 0.6 \times x$

- residual (error) is the difference between predicted and real value

$\text{Residual Sum of Squares (RSS)}$ for this line:

$$RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = (-0.8)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2$$
We should use the model \( \hat{y} = 2.2 + 0.6 \times x \)

What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = (-0.8)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2
\]
What is Residual Sum of Squares (RSS) for this line?

We should use the model \( \hat{y} = 2.2 + 0.6 \times x \)

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = \\
= 0.64 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2
\]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = 0.64 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2
\]
\( \hat{y} = 2.2 + 0.6 \times x \)

What is Residual Sum of Squares (RSS) for this line?

}\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 =
\]

\[= 0.64 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2\]
$\hat{y} = 2.2 + 0.6 \times x$

What is Residual Sum of Squares (RSS) for this line?

$RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 =
= 0.64 + (4 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2$

$\hat{y}_2 = 2.2 + 0.6 \times x_2$
\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[ \text{RSS} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = 0.64 + (4 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 \]

\[ \hat{y}_2 = 2.2 + 0.6 \times 2 \]
\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[ RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = 0.64 + (4 - \hat{y}_2)^2 + (\hat{y}_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 \]

\[ \hat{y}_2 = 2.2 + 1.2 \]
\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 =
\]

\[
= 0.64 + (4 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2
\]

\[ \hat{y}_2 = 3.4 \]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[
\text{RSS} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = 0.64 + (4 - 3.4)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2
\]

\[ \hat{y}_2 = 3.4 \]
Linear Regression (example)

\[
\hat{y} = 2.2 + 0.6 \times x
\]

What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = 0.64 + (0.6)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2
\]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[
\text{RSS} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = 0.64 + 0.36 + (y_3-\hat{y}_3)^2 + (y_4-\hat{y}_4)^2 + (y_5-\hat{y}_5)^2
\]
\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = \\
= 0.64 + 0.36 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2
\]
\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[ RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = \]

\[ = 0.64 + 0.36 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 \]
\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[ \text{RSS} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = \]

\[ = 0.64 + 0.36 + (5 - 4)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 \]

\[ \hat{y}_3 = 2.2 + 0.6 \times 3 \]
What is Residual Sum of Squares (RSS) for this line?

\[ RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = 0.64 + 0.36 + (5 - 4)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 \]

\[ \hat{y}_3 = 2.2 + 1.8 \]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 =
\]

\[
= 0.64 + 0.36 + (5 - 4)^2 + (4 - \hat{y}_4)^2 + (\hat{y}_5 - \hat{y}_5)^2
\]

\[
\hat{y}_3 = 4
\]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[
\text{RSS} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = 
\]

\[
= 0.64 + 0.36 + (1)^2 + (4 - \hat{y}_4)^2 + (5 - \hat{y}_5)^2
\]
\[
\hat{y} = 2.2 + 0.6 \times x
\]

What is Residual Sum of Squares (RSS) for this line?

\[
RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = 0.64 + 0.36 + 1 + 0.36 + 0.04
\]
Linear Regression (example)

\[ \hat{y} = 2.2 + 0.6 \times x \]

What is Residual Sum of Squares (RSS) for this line?

\[ RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = 2.4 \]
What is Residual Sum of Squares (RSS) for this line?

\[ RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = 2.4 \]
Linear Regression (example)

\[ \hat{y} = 2.2 + w_1 \times x \quad w_1 = [0.6, 0.7, 0.8, 0.9] \]

What is Residual Sum of Squares (RSS) for this line?

\[ RSS = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 = 2.4 \]
\[ \hat{y} = 2.2 + w_1 \times x \quad w_1 = [0.6, 0.7, 0.8, 0.9] \]

What is Residual Sum of Squares (RSS) for this line?

\[ \text{RSS}(w_1 = 0.6) = 2.4 \]
What is the Residual Sum of Squares (RSS) for this line?

\[ \hat{y} = 2.2 + w_1 \cdot x \quad w_1 = [0.6, 0.7, 0.8, 0.9] \]

\[ \text{RSS}(w_1 = 0.6) = 2.4 \]
\[ \text{RSS}(w_1 = 0.7) = 2.96 \]
\( \hat{y} = 2.2 + w_1 \times x \quad w_1 = [0.6, 0.7, 0.8, 0.9] \)

What is Residual Sum of Squares (RSS) for this line?

- \( \text{RSS}(w_1 = 0.6) = 2.4 \)
- \( \text{RSS}(w_1 = 0.7) = 2.96 \)
- \( \text{RSS}(w_1 = 0.8) = 4.6 \)
\[ \hat{y} = 2.2 + w_1 \times x \quad w_1 = [0.6, 0.7, 0.8, 0.9] \]

What is Residual Sum of Squares (RSS) for this line?

\[
\begin{align*}
\text{RSS}(w_1 = 0.6) &= 2.4 \\
\text{RSS}(w_1 = 0.7) &= 2.96 \\
\text{RSS}(w_1 = 0.8) &= 4.6 \\
\text{RSS}(w_1 = 0.9) &= 7.35
\end{align*}
\]
Loss functions we encountered

<table>
<thead>
<tr>
<th>Residual Sum of Squares</th>
<th>Mean Squared Error</th>
<th>Binary Cross-entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$</td>
<td>$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$</td>
<td>$BCE = -\frac{1}{n} \sum_{i=1}^{n} (y_i \log(p_i) + (1-y_i) \log(1-p_i))$</td>
</tr>
</tbody>
</table>
Loss functions we encountered

Residual Sum of Squares

\[ \text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

Mean Squared Error

\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

Binary Cross-entropy

\[ \text{BCE} = -\frac{1}{n} \sum_{i=1}^{n} (y_i \log(p_i) + (1-y_i)\log(1-p_i)) \]

\[ \text{RSS/MSE/BCE} \]
Loss functions we encountered

**Residual Sum of Squares**

\[ RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

**Mean Squared Error**

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

**Binary Cross-entropy**

\[ BCE = -\frac{1}{n} \sum_{i=1}^{n} (y_i \log(p_i) + (1-y_i) \log(1-p_i)) \]
Loss functions we encountered

**Residual Sum of Squares**

\[ \text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

**Mean Squared Error**

\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

**Binary Cross-entropy**

\[ \text{BCE} = -\frac{1}{n} \sum_{i=1}^{n} (y_i \log(p_i) + (1-y_i) \log(1-p_i)) \]

Adding the sum of squared model weights to the total loss (error) is called **L2 regularisation**
Loss functions we encountered

Residual Sum of Squares

\[ RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

Mean Squared Error

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

Binary Cross-entropy

\[ BCE = -\frac{1}{n} \sum_{i=1}^{n} (y_i \log(p_i) + (1-y_i) \log(1-p_i)) \]

Adding the sum of squared model weights to the total loss (error) is called L2 regularisation
Loss functions we encountered

**Residual Sum of Squares**

\[ RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

**Mean Squared Error**

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

**Binary Cross-entropy**

\[ BCE = -\frac{1}{n} \sum_{i=1}^{n} (y_i \log(p_i) + (1-y_i) \log(1-p_i)) \]

**Error + L2 regularisation**

\[ Error + \lambda \sum_{i} (w_i)^2 \]
Loss functions we encountered

<table>
<thead>
<tr>
<th>Residual Sum of Squares</th>
<th>Mean Squared Error</th>
<th>Binary Cross-entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 ]</td>
<td>[ MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 ]</td>
<td>[ BCE = -\frac{1}{n} \sum_{i=1}^{n} (y_i \log(p_i) + (1-y_i) \log(1-p_i)) ]</td>
</tr>
</tbody>
</table>

\[ \text{Error} + \lambda \sum_i^n (w_i)^2 \]

\[ \text{Error} + \lambda (w_1^2 + w_2^2 + \ldots + w_n^2) \]

sum of model weights squared
Loss functions we encountered

<table>
<thead>
<tr>
<th>Residual Sum of Squares</th>
<th>Mean Squared Error</th>
<th>Binary Cross-entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>( RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 )</td>
<td>( MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 )</td>
<td>( BCE = -\frac{1}{n} \sum_{i=1}^{n} (y_i \log(p_i) + (1-y_i)\log(1-p_i)) )</td>
</tr>
</tbody>
</table>

\[
Error + \lambda \sum_{i} (w_i)^2
\]

\[
Error + \lambda \sum_{i} |w_i|
\]

\( L2 \) regularisation

\( L1 \) regularisation
Loss functions we encountered

Residual Sum of Squares

\[ RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

Mean Squared Error

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

Binary Cross-entropy

\[ BCE = -\frac{1}{n} \sum_{i=1}^{n} (y_i \log(p_i) + (1-y_i)\log(1 - p_i)) \]

Error + \lambda \sum_{i}^{n} (w_i)^2 \quad \text{L2 regularisation}

Error + \lambda \sum_{i}^{n} |w_i| \quad \text{L1 regularisation}
### Loss functions we encountered

<table>
<thead>
<tr>
<th>Loss Function</th>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual Sum of Squares</td>
<td>$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$</td>
<td>Sum of squared differences between actual and predicted values.</td>
</tr>
<tr>
<td>Mean Squared Error</td>
<td>$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$</td>
<td>Average of the squares of the differences between actual and predicted values.</td>
</tr>
<tr>
<td>Binary Cross-entropy</td>
<td>$BCE = -\frac{1}{n} \sum_{i=1}^{n} (y_i \log(p_i) + (1-y_i)\log(1 - p_i))$</td>
<td>Measures how well a model’s predictions approximate the actual outcomes.</td>
</tr>
</tbody>
</table>

**Regularization Terms**

- **L2 regularization**
  
  $$\text{Error} + \lambda \sum_{i} (w_i)^2$$

- **L1 regularization**
  
  $$\text{Error} + \lambda \sum_{i} |w_i|$$

- Sum of absolute values of model weights
Loss functions we encountered

**Residual Sum of Squares**

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

**Mean Squared Error**

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

**Binary Cross-entropy**

$$BCE = -\frac{1}{n} \sum_{i=1}^{n} (y_i \log(p_i) + (1-y_i)\log(1-p_i))$$

Apart from having to **minimise the loss** (error)
regularisation terms are the **additional constraints**
which models must satisfy

https://towardsdatascience.com/intuitions-on-l1-and-l2-regularisation-235f2db4c261
L1 and L2 norms

L2 norm  \[ \|\tilde{w}\|_2 = (w_1^2 + w_2^2 + \ldots + w_n^2)^{\frac{1}{2}} \]
L1 and L2 norms

**L2 norm**  \[ \|\mathbf{w}\|_2 = (w_1^2 + w_2^2 + \ldots + w_n^2)^{\frac{1}{2}} \]

**L2 regularisation** is **squared L2 norm** of vector \( \mathbf{w} \)
L1 and L2 norms

L2 norm \[ \| \tilde{w} \|_2 = (w_1^2 + w_2^2 + \ldots + w_n^2)^{\frac{1}{2}} \]

L2 regularisation is squared L2 norm of vector \( \tilde{w} \)

\[ \| \tilde{w} \|_2^2 = (w_1^2 + w_2^2 + \ldots + w_n^2)^{\frac{1}{2} \times 2} = (w_1^2 + w_2^2 + \ldots + w_n^2) = \sum_{i=1}^{n} (w_i)^2 \]
**L1 and L2 norms**

\[ \| \tilde{w} \|_2 = (w_1^2 + w_2^2 + \ldots + w_n^2)^\frac{1}{2} \]

**L2 norm**

\[ \| \tilde{w} \|_2 = (w_1^2 + w_2^2 + \ldots + w_n^2)^{\frac{1}{2} \cdot 2} = (w_1^2 + w_2^2 + \ldots + w_n^2) = \sum_{i} (w_i)^2 \]

**L2 regularisation** is **squared L2 norm** of vector \( w \)

\[ \| \tilde{w} \|_1 = |w_1| + |w_2| + \ldots + |w_n| \]

**L1 norm**

**L1 regularisation** is just **L1 norm** of vector \( w \)

\[ \| \tilde{w} \|_1 = |w_1| + |w_2| + \ldots + |w_n| = \sum_{i} |w_i| \]
Linear Regression (without regularisation)

\[ \hat{y} = 2.2 + 0.6 \, x \]

\[ \bar{x} = 3 \quad \bar{y} = 4 \]

\[ w_1 = 0.6 \]

\[ w_0 = 2.2 \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

Train:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Validation:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Validation
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

Train:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \bar{x} = ? \quad \bar{y} = ? \quad \sum = ? \quad \sum = ? \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(\bar{x} = 1.5\) \(\bar{y} = 3\)

\(\sum = ?\) \(\sum = ?\)
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

Train:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x - \bar{x}</td>
<td>y - \bar{y}</td>
<td>(x - \bar{x})^2</td>
<td>(x - \bar{x})(y - \bar{y})</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = ? \quad \sum = ? \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

Train:

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
\hline
1 & 2 & -0.5 & -1 & & \\
2 & 4 & 0.5 & 1 & & \\
\hline
\end{array}
\]

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \Sigma = ? \quad \Sigma = ? \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td></td>
</tr>
</tbody>
</table>

\( \bar{x} = 1.5 \)  \( \bar{y} = 3 \)

\[ \sum = ? \]  \[ \sum = ? \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

**Train:**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x - \bar{x}</td>
<td>y - \bar{y}</td>
<td>(x - \bar{x})^2</td>
<td>(x - \bar{x})(y - \bar{y})</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \Sigma = ? \quad \Sigma = ? \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

Train:

\[
\begin{array}{c|c|c}
(x - \bar{x})^2 & 0.25 & 0.25 \\
\hline
\bar{y} & 3 & \\
\sum = 0.5 & \sum = ?
\end{array}
\]

Validation:

\[ \bar{x} = 1.5, \bar{y} = 3 \]

\[ \sum = 0.5 \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

Train:

\[
\begin{array}{cccccc}
\text{x} & \text{y} & \text{x} - \bar{x} & \text{y} - \bar{y} & (\text{x} - \bar{x})^2 & (\text{x} - \bar{x})(\text{y} - \bar{y}) \\
1 & 2 & -0.5 & -1 & 0.25 & 0.5 \\
2 & 4 & 0.5 & 1 & 0.25 & 0.5 \\
\end{array}
\]

\(\bar{x} = 1.5 \quad \bar{y} = 3\)

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

\[ w_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \]

\[
\begin{array}{c|c|c|c|c|c|c}
 x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
\hline
1 & 2 & -0.5 & -1 & 0.25 & 0.5 \\
2 & 4 & 0.5 & 1 & 0.25 & 0.5 \\
\hline
\end{array}
\]

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

\[ w_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \quad \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

\[ w_1 = \frac{1}{\sum (x - \bar{x})^2} \]

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

\[ w_1 = \frac{1}{\sum (x - \bar{x})^2} \sum (x - \bar{x})(y - \bar{y}) \]

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>( x - \bar{x} )</td>
<td>( y - \bar{y} )</td>
<td>( (x - \bar{x})^2 )</td>
<td>( (x - \bar{x})(y - \bar{y}) )</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\( \bar{x} = 1.5 \) \( \bar{y} = 3 \)

\( \sum = 0.5 \) \( \sum = 1 \)
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \]

\[ w_1 = \frac{1}{0.5} \]

\[ \bar{x} = 1.5 \]
\[ \bar{y} = 3 \]

\[ \sum = 0.5 \]
\[ \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + w_1 x \quad w_1 = 2 \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \bar{\sum} = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + 2x \quad w_1 = 2 \]

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + 2 \times x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = w_0 + 2 \times x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ 3 = w_0 + 2 \times 1.5 \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]

The line will go through this point.
Linear Regression (without regularisation)

\[ 3 = w_0 + 3 \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \Sigma = 0.5 \quad \Sigma = 1 \]

Line will go through this point
Linear Regression (without regularisation)

\[ w_0 = 3 - 3 \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ w_0 = 0 \]

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]

\[
\begin{array}{cccccc}
  x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
  1 & 2 & -0.5 & -1 & 0.25 & 0.5 \\
  2 & 4 & 0.5 & 1 & 0.25 & 0.5 \\
\end{array}
\]
Linear Regression (without regularisation)

\[ \hat{y} = 0 + 2 \times x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2x \]

Train:

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
\hline
1 & 2 & -0.5 & -1 & 0.25 & 0.5 \\
2 & 4 & 0.5 & 1 & 0.25 & 0.5 \\
\hline
\end{array}
\]

\( \bar{x} = 1.5 \quad \bar{y} = 3 \)

\( \sum = 0.5 \quad \sum = 1 \)
Linear Regression (without regularisation)

\[ \hat{y} = 2x \]

Train:

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
\hline
1 & 2 & -0.5 & -1 & 0.25 & 0.5 \\
2 & 4 & 0.5 & 1 & 0.25 & 0.5 \\
\hline
\end{array}
\]

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

Error = ?
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

\[ \text{Error} = \text{RSS} \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

\[ \text{Error} = \text{RSS} \]

\[ \text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

Train:

\[
| \begin{array}{|c|c|c|c|c|c|} 
  x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
  1 & 2 & -0.5 & -1 & 0.25 & 0.5 \\
  2 & 4 & 0.5 & 1 & 0.25 & 0.5 \\
  \end{array} |
\]

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]

Error = RSS

\[ RSS = \sum_{i=1}^{2} (y_i - \hat{y}_i)^2 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]

**Error** = RSS

\[ RSS = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

\[ \text{Error} = \text{RSS} \]

\[ \text{RSS} = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 \]

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x - \bar{x}</td>
<td>y - \bar{y}</td>
<td>(x - \bar{x})^2</td>
<td>(x - \bar{x})(y - \bar{y})</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

\[ \text{Error} = \text{RSS} \]

\[ \text{RSS} = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>(x - \bar{x})</th>
<th>(y - \bar{y})</th>
<th>((x - \bar{x})^2)</th>
<th>((x - \bar{x})(y - \bar{y}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

Error = RSS

\[ RSS = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

Train:

\[ x \quad y \quad x - \bar{x} \quad y - \bar{y} \quad (x - \bar{x})^2 \quad (x - \bar{x})(y - \bar{y}) \]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]

\[ \text{Error} = \text{RSS} \]

\[ \text{RSS} = (2 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

\[
\text{Error} = \text{RSS} = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2
\]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \quad \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 0.5 \quad \sum = 1 \]

Error = RSS

\[ RSS = (2 - 2)^2 + (y_2 - \hat{y}_2)^2 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

\[ \text{Error} = \text{RSS} \]

\[ \text{RSS} = (0)^2 + (y_2 - \hat{y}_2)^2 \]

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
\hline
1 & 2 & -0.5 & -1 & 0.25 & 0.5 \\
2 & 4 & 0.5 & 1 & 0.25 & 0.5 \\
\hline
\end{array}
\]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

\[ \text{Error} = \text{RSS} \]

\[ \text{RSS} = (0)^2 + (y_2 - \hat{y}_2)^2 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

\[ \text{Error} = \text{RSS} \]

\[ \text{RSS} = (0)^2 + (4 - \hat{y}_2)^2 \]

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x - \bar{x}</td>
<td>y - \bar{y}</td>
<td>(x - \bar{x})^2</td>
<td>(x - \bar{x})(y - \bar{y})</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

\[ \sum = 3 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

**What is the Error of this model?**

Error = RSS

\[ RSS = (0)^2 + (4 - \hat{y}_2)^2 \]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - (\bar{x})</th>
<th>y - (\bar{y})</th>
<th>((x - \bar{x})^2)</th>
<th>((x - \bar{x})(y - \bar{y}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\(\bar{x} = 1.5\) \(\bar{y} = 3\) 
\[\sum = 0.5 \quad \sum = 1\]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

\[ \bar{x} = 1.5 \]
\[ \bar{y} = 3 \]

\[ \sum = 0.5 \]
\[ \sum = 1 \]

\[ \text{Error} = RSS \]

\[ RSS = (0)^2 + (4 - 4)^2 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

\[ \text{Error} = \text{RSS} \]

\[ \text{RSS} = (0)^2 + (0)^2 \]

### Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - ( \bar{x} )</th>
<th>y - ( \bar{y} )</th>
<th>(x - ( \bar{x} ))^2</th>
<th>(x - ( \bar{x} ))(y - ( \bar{y} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\( \bar{x} = 1.5 \quad \bar{y} = 3 \)

\[ \sum = 0.5 \quad \sum = 1 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

Error = RSS

\[ RSS = 0 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

\[ \text{Error} = \text{RSS} \]

\[ \text{RSS} = 0 \]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

Error = \( RSS = 0 \)

\[
\sum (x - \bar{x})(y - \bar{y}) = 0.5
\]

\[
\sum (x - \bar{x})^2 = 0.5
\]

\[
\bar{x} = 1.5 \quad \bar{y} = 3
\]

\[
RSS = 0
\]
Linear Regression (without regularisation)

\[ \hat{y} = 2 \times x \]

What is the Error of this model?

**Error** = 0
Linear Regression (without regularisation)

\[
\hat{y} = 2 \times x
\]

model #1

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[\bar{x} = 1.5 \quad \bar{y} = 3\]

\[\sum = 0.5 \quad \sum = 1\]

Error \((2 \times x) = 0\)
Linear Regression (without regularisation)

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\bar{x} = 1.5, \bar{y} = 3

\sum x = 3.5, \sum y = 7

\sum (x - \bar{x})^2 = 0.5, \sum (x - \bar{x})(y - \bar{y}) = 1

Error \((2 \times x) = 0\)

model \#1 \(\hat{y} = 2 \times x\)

model \#2 \(\hat{y} = 1.5 \times x\)
Linear Regression (without regularisation)

What about its error?

Error \((2 \times x) = 0\)
Linear Regression (without regularisation)

Train:

Validation

\[ \overline{x} = 1.5 \]

\[ \overline{y} = 3 \]

\[ \sum = 0.5 \]

\[ \sum = 1 \]

\[ \sum (x - \overline{x})(y - \overline{y}) = 0.5 \]

\[ \sum (x - \overline{x})^2 = 1 \]

\[ \sum = 0.5 \]

\[ \sum = 1 \]

\[ \sum (x - \overline{x})(y - \overline{y}) = 0.5 \]

What about its error?

**model #1**

\[ \hat{y} = 2 \times x \]

\[ \text{Error } (2 \times x) = 0 \]

**model #2**

\[ \hat{y} = 1.5 \times x \]

\[ \text{Error } (1.5 \times x) = ? \]
Linear Regression (without regularisation)

Train:

### model #1
\[ \hat{y} = 2 \times x \]

### model #2
\[ \hat{y} = 1.5 \times x \]

What about its error?

**Error** \((2 \times x) = 0\)

**Error** \((1.5 \times x) = \text{RSS}\)

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - (\bar{x})</th>
<th>y - (\bar{y})</th>
<th>((x - \bar{x})^2)</th>
<th>((x - \bar{x})(y - \bar{y}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5, \bar{y} = 3 \]

\[ \sum = 0.5, \sum = 1 \]
Linear Regression (without regularisation)

\[
\hat{y} = 2 \times x
\]

\[
\hat{y} = 1.5 \times x
\]

What about its error?

\[
\text{Error } (2 \times x) = 0
\]

\[
\text{Error } (1.5 \times x) = 1.25
\]
Linear Regression (without regularisation)

Model #1
\[ \hat{y} = 2\times x \]

Model #2
\[ \hat{y} = 1.5\times x \]

We would obviously choose the **first model** as it works better on train.

**Model #1**

Error \((2\times x) = 0\)

**Model #2**

Error \((1.5\times x) = 1.25\)
Linear Regression (without regularisation)

How would the error change if we introduce \textbf{L2} regularisation?

Error $(2 \times x) = 0$

Error $(1.5 \times x) = 1.25$
Linear Regression (with $\textbf{L2}$ regularisation)

How would the error change if we introduce $\textbf{L2}$ regularisation?
Linear Regression (with \textbf{L2} regularisation)

How would the error change if we introduce \textbf{L2} regularisation?

\[ \text{Error} + \lambda \sum_{i=1}^{n} (w_i)^2 \]
Linear Regression (with **L2** regularisation)

How would the error change if we introduce **L2** regularisation?

\[ \text{Error} + \lambda \sum_{i}^{n} (w_i)^2 \]

Assume \( \lambda \) equal to 1
Linear Regression (with **L2** regularisation)

How would the error change if we introduce **L2** regularisation?

\[
Error + \lambda \sum_{i} (w_i)^2
\]

Assume \( \lambda \) equal to 1, in other circumstances it's found via **cross-validation algorithm**
Linear Regression (with \textbf{L2} regularisation)

How would the error change if we introduce \textbf{L2} regularisation?

$$Error + \sum_{i}^{n} (w_i)^2$$
Linear Regression (with L2 regularisation)

model #1
\[ \hat{y} = 2 \times x \]

model #2
\[ \hat{y} = 1.5 \times x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

model #1
Error \((2 \times x) = ?\)

model #2
Error \((1.5 \times x) = ?\)
Linear Regression (with L2 regularisation)

model #1
\[ \hat{y} = 2 \times x \]

model #2
\[ \hat{y} = 1.5 \times x \]

Train:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x - \bar{x}</td>
<td>y - \bar{y}</td>
<td>(x - \bar{x})^2</td>
<td>(x - \bar{x})(y - \bar{y})</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Error (2*x) = RSS + \sum_{i} (w_i)^2

model #2
Error (1.5*x) = RSS + \sum_{i} (w_i)^2
Linear Regression (with $L2$ regularisation)

**Train:**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>$x - \bar{x}$</th>
<th>$y - \bar{y}$</th>
<th>$(x - \bar{x})^2$</th>
<th>$(x - \bar{x})(y - \bar{y})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

**Error** $(2x) = 0 + \sum_i (w_i)^2$

**Error** $(1.5x) = \text{RSS} + \sum_i (w_i)^2$

\[
\hat{y} = 2x \\
\hat{y} = 1.5x
\]
Linear Regression (with L2 regularisation)

\[
\hat{y} = 2 \times x \\
\hat{y} = 1.5 \times x
\]

Train:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[
\text{Error} (2x) = 0 + \sum_{i} (w_i)^2
\]

\[
\text{Error} (1.5x) = 1.25 + \sum_{i} (w_i)^2
\]
Linear Regression (with $L_2$ regularisation)

Model #1
\[ \hat{y} = 2 \times x \]

Model #2
\[ \hat{y} = 1.5 \times x \]

Train:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
<th>$x - \bar{x}$</th>
<th>$y - \bar{y}$</th>
<th>$(x - \bar{x})^2$</th>
<th>$(x - \bar{x})(y - \bar{y})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Error (2\times x) = 0 + (w_1)^2

Error (1.5\times x) = 1.25 + (w_1)^2
Linear Regression (with L2 regularisation)

model #1
\[ \hat{y} = 2 \times x \]
model #2
\[ \hat{y} = 1.5 \times x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5, \quad \bar{y} = 3 \]
\[ \sum = 0.5 \]
\[ \sum = 1 \]

Error (2*\(x\)) = 0 + (w_1)^2

Error (1.5*\(x\)) = 1.25 + (w_1)^2
Linear Regression (with L2 regularisation)

model #1
\[ \hat{y} = 2 \times x \]

model #2
\[ \hat{y} = 1.5 \times x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \quad \sum = 1 \]

**Error** (2\times x) = 0 + (2)^2

**Error** (1.5\times x) = 1.25 + (w_1)^2
Linear Regression (with **L2** regularisation)

- **model #1**
  \[ \hat{y} = 2 \times x \]

- **model #2**
  \[ \hat{y} = 1.5 \times x \]

**Train:**

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

**Error** \( (2 \times x) = 4 \)

- **model #1**

- **model #2**

**Error** \( (1.5 \times x) = 1.25 + (w_1)^2 \)
Linear Regression (with \textbf{L}2 regularisation)

model #1
\[ \hat{y} = 2 \times x \]

model #2
\[ \hat{y} = 1.5 \times x \]

Train:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x - \bar{x}</td>
<td>y - \bar{y}</td>
<td>(x - \bar{x})^2</td>
<td>(x - \bar{x})(y - \bar{y})</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]

model #1
\textbf{Error} \ (2 \times x) = 4

model #2
\textbf{Error} \ (1.5 \times x) = 1.25 + (w_1)^2
Linear Regression (with L2 regularisation)

model #1
\[ \hat{y} = 2 \times x \]
model #2
\[ \hat{y} = 1.5 \times x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - \bar{x}</th>
<th>y - \bar{y}</th>
<th>(x - \bar{x})^2</th>
<th>(x - \bar{x})(y - \bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \]
\[ \sum = 0.5 \quad \sum = 1 \]

Error \((2 \times x) = 4\)

model #2
\[ \text{Error } (1.5 \times x) = 1.25 + (w_1)^2 \]
Linear Regression (with L2 regularisation)

Model #1:
\[ \hat{y} = 2 \times x \]

Model #2:
\[ \hat{y} = 1.5 \times x \]

Train:
\[ \begin{array}{cccccc}
 x & y & x - \bar{x} & y - \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
 1 & 2 & -0.5 & -1 & 0.25 & 0.5 \\
 2 & 4 & 0.5 & 1 & 0.25 & 0.5 \\
\end{array} \]

Validation

\[ \bar{x} = 1.5 \quad \bar{y} = 3 \quad \sum = 0.5 \quad \sum = 1 \]

Error (2\times x) = 4

Model #2

Error (1.5\times x) = 1.25 + (w_1)^2
Linear Regression (with **L2** regularisation)

**Train:**

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x - (\bar{x})</th>
<th>y - (\bar{y})</th>
<th>((x - \bar{x})^2)</th>
<th>((x - \bar{x})(y - \bar{y}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\(\bar{x} = 1.5\) \(\bar{y} = 3\)

**Error**

\((2 \times x) = 4\)

\((1.5 \times x) = 1.25 + (1.5)^2\)
Linear Regression (with L2 regularisation)

Train:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
<th>$x - \bar{x}$</th>
<th>$y - \bar{y}$</th>
<th>$(x - \bar{x})^2$</th>
<th>$(x - \bar{x})(y - \bar{y})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.5</td>
<td>-1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>

$\bar{x} = 1.5$, $\bar{y} = 3$

$\sum = 0.5$  $\sum = 1$

Error $(2^*x) = 4$

Error $(1.5^*x) = 1.25 + 2.25$
Linear Regression (with \textbf{L2} regularisation)

\begin{align*}
\hat{y} &= 2x \\
\hat{y} &= 1.5x
\end{align*}

Train:

\begin{tabular}{|c|c|c|c|c|c|}
\hline
\textbf{x} & \textbf{y} & \textbf{x - } \bar{x} & \textbf{y - } \bar{y} & (x - \bar{x})^2 & (x - \bar{x})(y - \bar{y}) \\
\hline
1 & 2 & -0.5 & -1 & 0.25 & 0.5 \\
2 & 4 & 0.5 & 1 & 0.25 & 0.5 \\
\hline
\end{tabular}

\begin{align*}
\bar{x} &= 1.5 \\
\bar{y} &= 3 \\
\sum &= 0.5 \\
\sum &= 1
\end{align*}

\textbf{Error} (2x) = 4

\textbf{Error} (1.5x) = 3.5
Linear Regression (with \textbf{L2} regularisation)

\hspace{1cm}

\begin{itemize}
  \item [model #1] \hspace{1cm} \hat{y} = 2 \times x
  \item [model #2] \hspace{1cm} \hat{y} = 1.5 \times x
\end{itemize}

Train:

\begin{tabular}{|c|c|c|c|c|c|}
\hline
$x$ & $y$ & $x - \bar{x}$ & $y - \bar{y}$ & $(x - \bar{x})^2$ & $(x - \bar{x})(y - \bar{y})$ \\
\hline
1 & 2 & -0.5 & -1 & 0.25 & 0.5 \\
2 & 4 & 0.5 & 1 & 0.25 & 0.5 \\
\hline
\end{tabular}

\hspace{1cm}

\begin{itemize}
  \item [model #1] \hspace{1cm} \textbf{Error} \hspace{0.5cm} (2 \times x) = 4
  \item [model #2] \hspace{1cm} \textbf{Error} \hspace{0.5cm} (1.5 \times x) = 3.5
\end{itemize}
Linear Regression (with \textbf{L2} regularisation)

Train:

This time we would prefer the \textbf{second model} due to lower total train error.

- \textbf{Model #1:} \( \hat{y} = 2 \times x \)  
  \( \text{Error (2\,x)} = 4 \)

- \textbf{Model #2:} \( \hat{y} = 1.5 \times x \)  
  \( \text{Error (1.5\,x)} = 3.5 \)
Linear Regression (with \textbf{L2} regularisation)

$$\hat{y} = 2 \cdot x$$  
$$\hat{y} = 1.5 \cdot x$$

These models do not have \textbf{intercept} \((w_0)\) but even if they would it \textbf{would not be} used as part of \textbf{L2 penalty}

$$\lambda \sum_{i=1}^{n} (w_i)^2 = \lambda (w_1^2 + w_2^2 + \ldots + w_n^2)$$

Error \((1.5 \cdot x) = 3.5\)
Linear Regression (with \textbf{L2 regularisation})

model #1
\( \hat{y} = 2 \times x \)

model #2
\( \hat{y} = 1.5 \times x \)

Train:

This time we would prefer the \textbf{second model} due to lower total train error

\[
\begin{array}{c|c|c}
\text{model #1} & \text{model #2} \\
\hline
\text{Error } (2 \times x) = 4 & \text{Error } (1.5 \times x) = 3.5 \\
\end{array}
\]
Linear Regression (with L2 regularisation)

model #1
\[ \hat{y} = 2 \times x \]
model #2
\[ \hat{y} = 1.5 \times x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Error (2*x) = 4
Error (1.5*x) = 3.5
Linear Regression (with $L_2$ regularisation)

Let's now recover the validation set.

Train:

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Error $(2 \times x) = 4$

Model #1

$\hat{y} = 2 \times x$

Model #2

$\hat{y} = 1.5 \times x$

Error $(1.5 \times x) = 3.5$
Linear Regression (with $\text{L2}$ regularisation)

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Error $(2x) = 4$

Validation:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Error $(1.5x) = 3.5$
Linear Regression (with L2 regularisation)

model #1
\[ \hat{y} = 2 \times x \]

model #2
\[ \hat{y} = 1.5 \times x \]

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Error \((2 \times x) = 4\)

Error \((1.5 \times x) = 3.5\)

Validation:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
Linear Regression (with **L2** regularisation)

Model #1
\[
\hat{y} = 2 \times x
\]

Model #2
\[
\hat{y} = 1.5 \times x
\]

**Train:**

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

- Model #1
  - Error \((2 \times x) = 4\)
- Model #2
  - Error \((1.5 \times x) = 3.5\)

**Validation:**

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

- Model #1
  - Error \((2 \times x) = ?\)
- Model #2
  - Error \((1.5 \times x) = ?\)
At this point I really got tired of arrows….

So I tried vector format for a change.
Linear Regression (with \textbf{L2} regularisation)

\[ \text{Error} \ (2x) = \sum \left( -\left(2 \times \begin{array}{c} x \\ 3 \\ 4 \\ 5 \end{array} \right) \right)^2 + w_1^2 \]
Linear Regression (with L2 regularisation)

Error \((2x)\) = \[\sum (y - 2x)^2 + w_1^2\]
Linear Regression (with \textbf{L2} regularisation)

Error \((2^x) = \sum (y - 2^x)^2 + w_1^2\)

Validation
Linear Regression (with \textbf{L2} regularisation)

Error \( (2^*x) = \sum (y - 2^*x)^2 \) + \( w^2 \)
Linear Regression (with \textbf{L2} regularisation)

\textbf{Error} \ (2^*x) = 42 + w_1^2
Linear Regression (with L2 regularisation)

model #1

**Error** \((2^*x) = 42 + w_1^2\)
Linear Regression (with L2 regularisation)

Error \((2^x) = 42 + 2^2\)
Linear Regression (with $L_2$ regularisation)

**Error** $(2^*x) = 42 + 4$
Linear Regression (with $L_2$ regularisation)

Error $(2^x) = 46$
Linear Regression (with L2 regularisation)

model #1
Error \( (2\times x) = 46 \)

model #2
Error \( (1.5\times x) = \sum (y - (1.5 \times x))^2 + w_1^2 \)
Linear Regression (with L2 regularisation)

model #1
Error \((2^*x) = 46\)

model #2
Error \((1.5^*x) = \sum (y - (1.5^*x))^2 + w_1^2\)
Linear Regression (with $L_2$ regularisation)

model #1

Error $(2^x) = 46$

model #2

Error $(1.5^x) = \sum (y - 1.5^x)^2 + w_1^2$

<table>
<thead>
<tr>
<th>$y - 1.5^x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
</tr>
<tr>
<td>-2</td>
</tr>
<tr>
<td>-2.5</td>
</tr>
</tbody>
</table>
Linear Regression (with L2 regularisation)

Model 1: Error $(2^*x) = 46$

Model 2: Error $(1.5^*x) = \sum (y - 1.5^*x)^2 + w_1^2$

<table>
<thead>
<tr>
<th>$(y - 1.5^*x)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>6.25</td>
</tr>
</tbody>
</table>
Linear Regression (with L2 regularisation)

Model #1

Error \((2 \times x) = 46\)

Model #2

Error \((1.5 \times x) = 10.5 + w_1^2\)
Linear Regression (with $L2$ regularisation)

**model #1**

\[ \text{Error } (2^*x) = 46 \]

**model #2**

\[ \text{Error } (1.5^*x) = 10.5 + 1.5^2 \]
Linear Regression (with L2 regularisation)

Error $(2 \times x) = 46$

Error $(1.5 \times x) = 10.5 + 2.25$
Linear Regression (with L2 regularisation)

**model #1**
**Error** \( (2^*x) = 46 \)

**model #2**
**Error** \( (1.5^*x) = 12.75 \)
Linear Regression (with L2 regularisation)

Train:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Error (2*x) = 4
Error (1.5*x) = 3.5

Validation:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Error (2*x) = 46
Error (1.5*x) = 12.75

model #1
\[ \hat{y} = 2 \times x \]
model #2
\[ \hat{y} = 1.5 \times x \]
Linear Regression (with L2 regularisation)

\[ \hat{y} = 2 \times x \]

\[ \hat{y} = 1.5 \times x \]

L2 regularisation forces models to prefer smaller weights \((w)\)
Linear Regression (with **L2** regularisation)

L2 regularisation forces models to prefer **smaller weights** ($w$).

Linear Regression with **L2** regularisation will have smaller slope ($w_1$).
Linear Regression (with \textbf{L2} regularisation)

\[ \hat{y} = 2 \times x \]
\[ \hat{y} = 1.5 \times x \]

\[ \text{Error} + \lambda \sum_{i}^{n} (w_i)^2 \]

\textbf{L2} regularisation can be made \textbf{stronger} or \textbf{weaker} by tweaking \( \lambda \).
Linear Regression (with \textbf{L2} regularisation)

\[ \hat{y} = 2 \times x \]
\[ \hat{y} = 1.5 \times x \]

We initialised \( \lambda \) equal to 1

\[ Error + \lambda \sum_{i}^{n} (w_i)^2 \]

\textbf{L2} regularisation can be made \textbf{stronger} or \textbf{weaker} by tweaking \( \lambda \)
Validation

\[
\hat{y} = 2 \times x \\
\hat{y} = 1.5 \times x
\]

We initialised \( \lambda \) equal to 1

\[
Error + 1 \times \sum_{i}^{n} (w_i)^2
\]

\textbf{L2} regularisation can be made \textit{stronger} or \textit{weaker} by tweaking \( \lambda \)
**Linear Regression (with L2 regularisation)**

Model #1
\[ \hat{y} = 2 \times x \]

Model #2
\[ \hat{y} = 1.5 \times x \]

We initialised \( \lambda \) equal to 1

\[
\text{Error} + \sum_{i=1}^{n} (w_i)^2
\]

**L2** regularisation can be made **stronger** or **weaker** by tweaking \( \lambda \).
Linear Regression (with L2 regularisation)

model #1
\[ \hat{y} = 2 \times x \]
model #2
\[ \hat{y} = 1.5 \times x \]

We initialised \( \lambda \) equal to 1
This resulted in model #2

\[
\text{Error} + \sum_{i}^{n} (w_i)^2
\]

L2 regularisation can be made stronger or weaker by tweaking \( \lambda \)
Linear Regression (with L2 regularisation)

\[ \hat{y} = 2 \times x \]

We initialised \( \lambda \) equal to 1
This resulted in model #2

\[ \text{Error} + \sum_{i}^{n} (w_i)^2 \]

L2 regularisation can be made stronger or weaker by tweaking \( \lambda \)
Linear Regression (with \textbf{L2} regularisation)

\[ \hat{y} = 2 \times x \]

\[ \lambda = 1 \]

\textbf{Error} + \lambda \sum_{i}^{n} (w_i)^2

\textbf{L2} regularisation can be made \textbf{stronger} or \textbf{weaker} by tweaking \( \lambda \)
Linear Regression (with L2 regularisation)

What if we use $\lambda = 0$ instead?

$$Error + \lambda \sum_{i} (w_i)^2$$

L2 regularisation can be made stronger or weaker by tweaking $\lambda$.
Linear Regression (with \textbf{L2} regularisation)

\[
\hat{y} = 2 \times x
\]

\( \lambda = 1 \)

What if we use \( \lambda = 0 \) instead?

\[\text{Error} + 0 \times \sum_{i}^{n} (w_i)^2\]

\textbf{L2} regularisation can be made \textbf{stronger} or \textbf{weaker} by tweaking \( \lambda \)
Linear Regression (with L2 regularisation)

model #1
\[ \hat{y} = 2 \times x \]

model #2
\[ \lambda = 1 \]

What if we use \( \lambda = 0 \) instead?
This results in standard Linear Regression model #1

L2 regularisation can be made stronger or weaker by tweaking \( \lambda \)
Linear Regression (with L2 regularisation)

model #1
\( \lambda = 0 \)

model #2
\( \lambda = 1 \)

What if we use \( \lambda = 0 \) instead?

This results in standard Linear Regression model #1

Error

L2 regularisation can be made stronger or weaker by tweaking \( \lambda \)
Linear Regression (with \textbf{L2 regularisation})

\[ Error + \lambda \sum_{i}^{n} (w_i)^2 \]

\textbf{L2} regularisation can be made \textbf{stronger} or \textbf{weaker} by tweaking \( \lambda \)
Linear Regression (with L2 regularisation)

model #1
\( \lambda = 0 \)

model #2
\( \lambda = 1 \)

Large the value of \( \lambda \) smaller the \( w \) that will be used in the final model

\[
\text{Error} + \lambda \sum_{i}^{n} (w_i)^2
\]

L2 regularisation can be made stronger or weaker by tweaking \( \lambda \)
Linear Regression (with $\textbf{L2}$ regularisation)

When $\lambda$ is \textbf{extremely large},
model prefers slope ($w_1$) to be very close to 0

$$\text{Error} + \lambda \sum_{i}^{n} (w_i)^2$$

$L2$ regularisation can be made \textbf{stronger} or \textbf{weaker} by tweaking $\lambda$
Linear Regression (with $L2$ regularisation)

looks like a lot of fun…
Linear Regression (with L2 regularisation)

model #1
\( \lambda = 0 \)

model #2
\( \lambda = 1 \)

Looks like a lot of fun…
… but however large or small value of \( \lambda \) gets…
Linear Regression (with L2 regularisation)

- \( \lambda = 0 \)  
- \( \lambda = 1 \)  
- \( \lambda = 1000 \)

Looks like a lot of fun…  
… but however large or small value of \( \lambda \) gets...

Linear Regression trained on one variable will not be powerful enough to overfit
Linear Regression (with $\textbf{L2}$ regularisation)

Looks like a lot of fun…

… but however large or small value of $\lambda$ gets…

Linear Regression trained on one variable will not be powerful enough to overfit

Let’s add a twist :)
So far we used linear function:

\[ \hat{y} = w_0 + w_1 \cdot x \]
So far we used linear function:

\[ \hat{y} = w_0 + w_1 x \]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

We can make our function more powerful adding degrees of polynomial

\[
\begin{array}{c|ccccc|c}
  x & x^2 & x^3 & x^4 & y \\
  \hline
  1 & 1 & 1 & 1 & 2 \\
  2 & 4 & 8 & 16 & 4 \\
  3 & 9 & 27 & 81 & 5 \\
  4 & 16 & 64 & 256 & 4 \\
  5 & 25 & 125 & 625 & 5 \\
\end{array}
\]
So far we used linear function:

\[ \hat{y} = w_0 + w_1 x \]

We can make our function more powerful adding degrees of polynomial

\[ \hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 \]

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x^2</th>
<th>x^3</th>
<th>x^4</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>9</td>
<td>27</td>
<td>81</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>25</td>
<td>125</td>
<td>625</td>
<td>5</td>
</tr>
</tbody>
</table>
*here I show only $x^1$, the others degrees are hidden behind the curtains*
Polynomial model

\[ \hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 \]

*here I show only \( x^1 \), the others degrees are hidden behind the curtains*
Polynomial model
trained with
conventional error function (RSS)
without L2 regularisation ($\lambda = 0$)

*here I show only $x^1$, the others degrees are hidden behind the curtains*
Polynomial **model**

trained with

conventional **error function** (RSS)

**without L2** regularisation ($\lambda = 0$)

$\hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$

$$\text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2$$

*here I show only $x^1$, the others degrees are hidden behind the curtains*
Polynomial model trained with conventional error function (RSS) without L2 regularisation ($\lambda = 0$) results in

$$\hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$$

$$\text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2$$

*here I show only $x^1$, the others degrees are hidden behind the curtains*
Polynomial model
trained with
conventional error function (RSS)
without L2 regularisation ($\lambda = 0$)
results in

$$\hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$$

$$Error = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2$$

*here I show only $x^1$, the others degrees are hidden behind the curtains
Polynomial model trained with conventional error function (RSS) without L2 regularisation ($\lambda = 0$) results in flexible predictions (i.e. overfitting)

$$\hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$$

$$\text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2$$

*here I show only $x^1$, the others degrees are hidden behind the curtains*
Polynomial model

\[ \hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 \]

trained with

conventional error function (RSS)
without L2 regularisation (\( \lambda = 0 \))

results in

\[ \text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 \]

flexible predictions (i.e., overfitting)

*here I show only \( x^1 \), the others degrees are hidden behind the curtains

Resulting model has large weights (coefficients)

<table>
<thead>
<tr>
<th>( x )</th>
<th>( x^3 )</th>
<th>( x^4 )</th>
<th>( y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>27</td>
<td>81</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>125</td>
<td>625</td>
</tr>
</tbody>
</table>
Polynomial model

trained with conventional error function (RSS) without L2 regularisation ($\lambda = 0$)

results in flexible predictions (i.e. overfitting)

$\hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$

$Error = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2$

*here I show only $x^1$, the others degrees are hidden behind the curtains

<table>
<thead>
<tr>
<th>x</th>
<th>x²</th>
<th>x³</th>
<th>x⁴</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>27</td>
<td>81</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>125</td>
<td>625</td>
<td>5</td>
</tr>
</tbody>
</table>
Polynomial model trained with conventional error function (RSS) with L2 regularisation results in

$$\hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$$

where

$$\text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i}^{5} (w_i)^2$$

Let's see what happens if we add L2 regularisation

*here I show only $x^1$, the others degrees are hidden behind the curtains
Polynomial model trained with conventional error function (RSS) with L2 regularisation results in:

\[ \hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 \]

\[ \text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i} (w_i)^2 \]

Validation

<table>
<thead>
<tr>
<th>x</th>
<th>x^2</th>
<th>x^3</th>
<th>x^4</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>27</td>
<td>81</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>125</td>
<td>625</td>
<td>5</td>
</tr>
</tbody>
</table>

Let's see what happens if we add L2 regularisation: \[ \lambda = [0.1, 0.5, 0.8, 1.0] \]

*here I show only \( x^1 \), the others degrees are hidden behind the curtains*
Here I show only $x^1$, the others degrees are hidden behind the curtains.
Polynomial model

trained with

classical error function (RSS)

with L2 regularisation

results in

\[
\hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4
\]

Error function:

\[
Error = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i}^{5} (w_i)^2
\]

\(\lambda = [0.1, \ldots, \ldots, \ldots]\)

*here I show only \(x^1\), the others degrees are hidden behind the curtains

<table>
<thead>
<tr>
<th>x</th>
<th>x^2</th>
<th>x^3</th>
<th>x^4</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>27</td>
<td>81</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>125</td>
<td>625</td>
<td>5</td>
</tr>
</tbody>
</table>
Polynomial model trained with conventional error function (RSS) with \textit{L2} regularisation results in

\[
\hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4
\]

\[
\text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i} (w_i)^2
\]

\[
\lambda = 0.1 \quad \lambda = 0.5
\]

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
x & x^2 & x^3 & x^4 & y \\
\hline
1 & 1 & 1 & 1 & 2 \\
\hline
2 & 4 & 8 & 16 & 4 \\
\hline
3 & 9 & 27 & 81 & 5 \\
\hline
4 & 16 & 64 & 256 & 4 \\
\hline
5 & 25 & 125 & 625 & 5 \\
\hline
\end{array}
\]

*here I show only \(x^1\), the others degrees are hidden behind the curtains*
Polynomial model trained with conventional error function (RSS) with L2 regularisation results in

\[ \hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 \]

\[ \text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i} (w_i)^2 \]

Here I show only \( x^1 \), the others degrees are hidden behind the curtains

<table>
<thead>
<tr>
<th>( x )</th>
<th>( x^2 )</th>
<th>( x^3 )</th>
<th>( x^4 )</th>
<th>( y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>27</td>
<td>81</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>125</td>
<td>625</td>
<td>5</td>
</tr>
</tbody>
</table>
Polynomial model

trained with

conventional error function (RSS)

with L2 regularisation

results in

\[ \hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 \]

\[ Error = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i} (w_i)^2 \]

\[ \lambda = [0.1, 0.5, 0.8, 1.0] \]

*here I show only \( x^1 \), the others degrees are hidden behind the curtains
Polynomial model trained with conventional error function (RSS) with L2 regularisation results in trained with conventional error function (RSS) with L2 regularisation results in

Final weights \( (\lambda = 1) \)

\[
1.6 + 0.4x + 0.6x^2 - 0.2x^3 + 0.02x^4
\]

\( \lambda = 1.0 \)

<table>
<thead>
<tr>
<th>( x )</th>
<th>( x^2 )</th>
<th>( x^3 )</th>
<th>( x^4 )</th>
<th>( y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>27</td>
<td>81</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>125</td>
<td>625</td>
<td>5</td>
</tr>
</tbody>
</table>

*here I show only \( x^1 \), the others degrees are hidden behind the curtains*
Polynomial model

$\hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$

Initial weights ($\lambda = 0$)

$5 - 8.8^*x + 7.8^*x^2 - 2.3^*x^3 + 0.2^*x^4$

Final weights ($\lambda = 1$)

$1.6 + 0.4^*x + 0.6^*x^2 - 0.2^*x^3 + 0.02^*x^4$

Error

$\sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i}^{5} (w_i)^2$

Results in

$\lambda = [0.1, 0.5, 0.8, 1.0]$

Table:

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>x^2</th>
<th>x^3</th>
<th>x^4</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>27</td>
<td>81</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>125</td>
<td>625</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

*here I show only $x^1$, the others degrees are hidden behind the curtains
Polynomial model

\[ \hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 \]

Initial weights (\( \lambda = 0 \))

\[ 5 - 8.8^*x + 7.8^*x^2 - 2.3^*x^3 + 0.2^*x^4 \]

Final weights (\( \lambda = 1 \))

\[ 1.6 + 0.4^*x + 0.6^*x^2 - 0.2^*x^3 + 0.02^*x^4 \]

*here I show only \( x^1 \), the others degrees are hidden behind the curtains

Error

\[ \text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i} (w_i)^2 \]

\( \lambda = [0.1, 0.5, 0.8, 1.0] \)

Most optimal \( \lambda \) must be chosen using cross-validation

<table>
<thead>
<tr>
<th>( \lambda = 0.1 )</th>
<th>( \lambda = 0.5 )</th>
<th>( \lambda = 0.8 )</th>
<th>( \lambda = 1.0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>27</td>
<td>81</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>125</td>
<td>625</td>
</tr>
</tbody>
</table>

Results in

Trained with

Conventional error function (RSS)
Polynomial model

trained with

conventional error function (RSS)

results in

\[ \hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 \]

\[ Error = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i} (w_i)^2 \]

No regularisation \((\lambda = 0)\)

\[ 5 - 8.8^*x + 7.8^*x^2 - 2.3^*x^3 + 0.2^*x^4 \]

L2 regularisation \((\lambda = 1)\)

\[ 1.6 + 0.4^*x + 0.6^*x^2 - 0.2^*x^3 + 0.02^*x^4 \]

\[ \lambda = 0 \]

\[ \lambda = 1.0 \]

*here I show only \(x^1\), the others degrees are hidden behind the curtains

<table>
<thead>
<tr>
<th>(x)</th>
<th>(x^2)</th>
<th>(x^3)</th>
<th>(x^4)</th>
<th>(y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>27</td>
<td>81</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>125</td>
<td>625</td>
<td>5</td>
</tr>
</tbody>
</table>
Polynomial model

![Diagram showing polynomial model with different regularisation strengths.](image_url)

**No regularisation** ($\lambda = 0$)

$$5 - 8.8^*x + 7.8^*x^2 - 2.3^*x^3 + 0.2^*x^4$$

**L2 regularisation** ($\lambda = 1$)

$$1.6 + 0.4^*x + 0.6^*x^2 - 0.2^*x^3 + 0.02^*x^4$$

**L1 regularisation** ($\lambda = 1$)

![Table showing data points and their corresponding y values](image_url)

*here I show only $x^1$, the others degrees are hidden behind the curtains.*
Two regularisation methods are very similar

\[ \text{Error} + \lambda \sum_{i}^{n} (w_i)^2 \quad \text{L2 regularisation} \]

\[ \text{Error} + \lambda \sum_{i}^{n} |w_i| \quad \text{L1 regularisation} \]
Two regularisation methods are very similar

\[ \text{Error} + \lambda \sum_{i}^{n} (w_i)^2 \quad \text{L2 regularisation} \]

\[ \text{Error} + \lambda \sum_{i}^{n} |w_i| \quad \text{L1 regularisation} \]

the only difference
Two regularisation methods are very similar

$\text{Error} + \lambda \sum_{i}^{n} (w_i)^2$ \text{L2 regularisation}

$\text{Error} + \lambda \sum_{i}^{n} |w_i|$ \text{L1 regularisation}

Yet, this difference is very important
Polynomial model trained with conventional error function (RSS) results in trained with L2 regularisation (λ = 1)

No regularisation (λ = 0)

$L^2$ regularisation (λ = 1)

$\hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$

$5 - 8.8 * x + 7.8 * x^2 - 2.3 * x^3 + 0.2 * x^4$

$\text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i} (w_i)^2$

$\lambda = [0.1, 0.5, 0.8, 1.0]$

L1 regularisation (λ = 1)

$x$ | $x^2$ | $x^3$ | $x^4$ | $y$
---|---|---|---|---
1 | 1 | 1 | 1 | 2
2 | 4 | 4 | 4 | 4
3 | 9 | 27 | 81 | 5
4 | 16 | 64 | 256 | 4
5 | 25 | 125 | 625 | 5

Red and blue curves look similar

*here I show only $x^1$, the others degrees are hidden behind the curtains
Polynomial model trained with conventional error function (RSS) results in

\[ \hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 \]

No regularisation (\( \lambda = 0 \))

Conventional error function (RSS)

\[ 5 - 8.8 \times x + 7.8 \times x^2 - 2.3 \times x^3 + 0.2 \times x^4 \]

\[ \text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i} (w_i)^2 \]

\( \lambda = [0.1, 0.5, 0.8, 1.0] \)

L2 regularisation (\( \lambda = 1 \))

\[ 1.6 + 0.4 \times x + 0.6 \times x^2 - 0.2 \times x^3 + 0.02 \times x^4 \]

L1 regularisation (\( \lambda = 1 \))

```
\begin{tabular}{|c|c|c|c|c|c|c|c|}
\hline
x  & x^2 & x^3 & x^4 & y  \\
\hline
1  & 1  & 1   & 1   & 1  & 2  \\
2  & 4  & 8   & 16  & 5  \\
3  & 9  & 27  & 81  & 4  \\
4  & 16 & 64  & 256 & 5  \\
5  & 25 & 125 & 625 & 5  \\
\hline
\end{tabular}
```

Red and blue curves look similar

What is the difference?

*here I show only \( x^1 \), the others degrees are hidden behind the curtains
Polynomial model
trained with
conventional error function (RSS)
5 - 8.8*x + 7.8*x^2 - 2.3*x^3 + 0.2*x^4
and
Error = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i} (w_i)^2
\lambda = [0.1, 0.5, 0.8, 1.0]

**L2 regularisation (λ = 1)**

1.6 + 0.4*x + 0.6*x^2 - 0.2*x^3 + 0.02*x^4

Red and blue curves look similar
What is the difference?

**L1 regularisation (λ = 1)**

3.1 + 0*x + 0*x^2 - 0.06*x^3 + 0.008*x^4
Polynomial model
trained with conventional error function (RSS)

\[ \text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i} (w_i)^2 \]

No regularisation ($\lambda = 0$)

L2 regularisation ($\lambda = 1$)

error function (RSS)

\[ \text{Error} = \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i} (w_i)^2 \]

Polynomial model

\[ \hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 \]

L2 regularisation ($\lambda = 1$)

\[ \hat{y} = 1.6 + 0.4x + 0.6x^2 - 0.2x^3 + 0.02x^4 \]

L1 regularisation ($\lambda = 1$)

\[ \hat{y} = 3.1 + 0x + 0x^2 - 0.06x^3 + 0.008x^4 \]

No regularisation ($\lambda = 0$)

\[ \hat{y} = 5 - 8.8x + 7.8x^2 - 2.3x^3 + 0.2x^4 \]

Red and blue curves look similar

What is the difference?

While L2 regularisation forces some weights to be close to 0

\[ \hat{y} = 1.6 + 0.4x + 0.6x^2 - 0.2x^3 + 0.02x^4 \]

\[ \hat{y} = 3.1 + 0x + 0x^2 - 0.06x^3 + 0.008x^4 \]

Here I show only $x^1$, the others degrees are hidden behind the curtains.
Polynomial model

\[ \hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 \]

Error = \[ \sum_{i=1}^{5} (y_i - \hat{y}_i)^2 + \lambda \sum_{i} (w_i)^2 \]

No regularisation (\( \lambda = 0 \))

5 - 8.8 * x + 7.8 * x^2 + 2.3 * x^3 + 0.2 * x^4

L2 regularisation (\( \lambda = 1 \))

1.6 + 0.4 * x + 0.6 * x^2 - 0.2 * x^3 + 0.02 * x^4

L1 regularisation (\( \lambda = 1 \))

0 * x + 0 * x^2 - 0.06 * x^3 + 0.008 * x^4

*here I show only \( x^1 \), the others degrees are hidden behind the curtains

Red and blue curves look similar

What is the difference?

While L2 regularisation forces some weights to be close to 0

L1 regularisation can drive some weights all the way down to 0

No conventional error function (RSS)

Error function (RSS) with L2 regularisation results in

\( \lambda = [0.1, 0.5, 0.8, 1.0] \)
**L1 regularisation (λ = 1)**

\[3.1 + 0^*x + 0^*x^2 - 0.06^*x^3 + 0.008^*x^4\]
\textbf{L1 regularisation ($\lambda = 1$)}

\[ 3.1 + 0^*x + 0^*x^2 - 0.06^*x^3 + 0.008^*x^4 \]

\begin{table}[h]
\begin{tabular}{|c|c|c|c|c|c|}
\hline
x & x^2 & x^3 & x^4 & y  \\
\hline
1 & 1 & 1 & 1 & 2  \\
\hline
2 & 4 & 8 & 16 & 4  \\
\hline
3 & 9 & 27 & 81 & 5  \\
\hline
4 & 16 & 64 & 256 & 4  \\
\hline
5 & 25 & 125 & 625 & 5  \\
\hline
\end{tabular}
\end{table}
**L1 regularisation (λ = 1)**

\[ 3.1 + 0^*x + 0^*x^2 - 0.06^*x^3 + 0.008^*x^4 \]

The L1 regularisation is effective in reducing the impact of useless features.
**L1 regularisation (λ = 1)**

\[ y = 3.1 + 0^*x + 0^*x^2 - 0.06^*x^3 + 0.008^*x^4 \]

Useless features
**L1** regularisation ($\lambda = 1$)

$$3.1 + 0^*x + 0^*x^2 - 0.06^*x^3 + 0.008^*x^4$$

**L1** regularisation can be used as feature selection
When should I apply \textbf{L2} and when \textbf{L1}?

\[ \text{Error} + \lambda \sum_{i}^{n} (w_i)^2 \quad \text{L2 regularisation} \]

\[ \text{Error} + \lambda \sum_{i}^{n} |w_i| \quad \text{L1 regularisation} \]
When should I apply \textbf{L2} and when \textbf{L1}?

\[ \text{Error} + \lambda \sum_{i}^{n} (w_i)^{2} \quad \text{L2 regularization} \]

If you want to \textbf{reduce overfitting}, apply \textbf{L2}, always.
When should I apply \textbf{L2} and when \textbf{L1}? 

\[
\text{Error} + \lambda \sum_{i}^{n} (w_i)^2 \quad \text{L2 regularisation}
\]

\[
\text{Error} + \lambda \sum_{i}^{n} |w_i| \quad \text{L1 regularisation}
\]
When should I apply **L2** and when **L1**?

If you want to try exquisite feature selection, use **L1**

\[
\text{Error} + \lambda \sum_{i}^{n} |w_i| \quad \text{L1 regularisation}
\]
ML + L1 or L2 regularisation =
Machine Learning model with L1 or L2 regularisation
Machine Learning model with \textbf{L1} or \textbf{L2} regularisation
\[ \text{ML} + \text{L1 or L2 regularisation} = \text{Machine Learning model with L1 or L2 regularisation} \]
Machine Learning model with L1 or L2 regularisation

ML + L1 or L2 regularisation =

LR + L1 regularisation =

LR + L2 regularisation =
Machine Learning model with L1 or L2 regularisation

Lasso (LASSO) regression
Machine Learning model with L1 or L2 regularisation

Lasso (LASSO) regression

Ridge regression
Machine Learning model with L1 or L2 regularisation = Lasso (LASSO) regression

Machine Learning model with L1 or L2 regularisation = Ridge regression
Machine Learning model with L1 or L2 regularisation

\[ \text{LR} + \text{L1 regularisation} = \text{Lasso (LASSO)} \text{ regression} \]

**LASSO** is an acronym
Machine Learning model with L1 or L2 regularisation

\[ LR + \text{L1 regularisation} = \text{Lasso (LASSO) regression} \]

**LASSO** is an **acronym**

Least Absolute Shrinkage and Selection Operator

\[
RSS + \lambda \sum_{i}^{n} |w_i|
\]
Machine Learning model with L1 or L2 regularisation

Lasso (LASSO) regression

Ridge regression
Machine Learning model with L1 or L2 regularisation

Lasso (LASSO) regression

Ridge regression
**Ridge** regression also known as **Tikhonov** regularisation proposed by Andrey Tikhonov

\[
RSS + \lambda \sum_{i}^{n} (w_i)^2
\]

**Ridge** regression also known as **Tikhonov** regularisation proposed by Andrey Tikhonov

\[
RSS + \lambda \sum_{i}^{n} (w_i)^2
\]
**Ridge** regression also known as **Tikhonov** regularisation proposed by Andrey Tikhonov

\[
\text{RSS} + \lambda \sum_{i}^{n} (w_{i})^2
\]

Arguably one of the most popular regularisation techniques in **ML**
Machine Learning model with L1 or L2 regularisation

Lasso (LASSO) regression

Ridge regression
Where to find more?

Regularization Part 2:
Lasso Regression
Clearly Explained!!!

StatQuest: Lasso Regression
https://youtu.be/NGf0voTMIcs

Lasso Regression
https://youtu.be/jbwSCwoT51M
**Machine Learning model with L1 or L2 regularisation**

- **ML** + L1 or L2 regularisation = Machine Learning model with L1 or L2 regularisation

- **LR** + L1 regularisation = Lasso (LASSO) regression

- **LR** + L2 regularisation = Ridge regression

- **DL** + L1 or L2 regularisation = Machine Learning model with L1 or L2 regularisation
Machine Learning model with L1 or L2 regularisation

Lasso (LASSO) regression

Ridge regression

Deep Learning model with weight decay
Regularisation methods

Way A methods

Way B methods
Regularisation methods

Explicit methods

Way B methods
That's all Folks!