Recall from Lecture 3: Customer segmentation

Intuition-based

Historical/behavioral-based
- RFM
- Value tier
- Lifecycle stage
  - new customer
  - regular
  - loyal
  - returning

Data-driven
- K-means
- Hierarchical clustering
- DB Scan
Recall from Lecture 3: Customer segmentation

Intuition-based

Historical/behavioral-based

Data-driven

RFM

Value tier

Lifecycle stage

new customer

regular

loyal

returning
Marketing and Sales

Customer Lifecycle Management: Regression Problems
Customer lifecycle
Customer lifecycle
Moving companies grow not because they force people to move more often, but by attracting new customers.
Relationships based on commitment

event-based  subscription-based
Relationships based on commitment

Event-based
- Packers and Movers
- Wedding Planners

Subscription-based
- Telco
- Banks
- Retail (Walmart, Konsume, etc)
- Hairdressers
Customer lifecycle
Customer lifecycle
Customer Lifecycle
(Techniques/Approaches)

**Problems**

- Start with understanding of your existing customers (Segment the customers)
- Acquire profitable customers
- Understanding future behavior with Propensity Model
- Convince/Influencing your customers to Spend more
- Causality or Just by chance?

**Solutions**

- Clustering (Lecture 3)
- Regression techniques (Present, Lecture 4)
- Classification (Lecture 5)
- Cross-selling/Up-selling (Lectures 6)
- AB Testing (Lectures 7)
Customer lifecycle (Returns)

Customer lifetime value (CLV)

- or CLTV, lifetime customer value (LCV), or life-time value (LTV).
- Describes the amount of profit a customer generates over his or her entire lifetime*
- We attempt to minimize “cost per acquisition” (CAC) and keeping any given customer.

*: Very much depends on the domain. For example, 1 to 20 years.
CLV is often referred to two forms of lifetime value analysis:

**Historical lifetime value**: simply sums up revenue or profit per customer.

**Predictive lifetime value**: projects what new customers will spend over their entire lifetime.
Historical Life time value

RFM

- F: Frequency
  - How often does the customer purchase?

- R: Recency
  - How recently did the customer purchase?

- M: Monetary Value
  - How much does the customer spend?

(time)
CLV is often referred to two forms of lifetime value analysis:

**Historical lifetime value:** simply sums up revenue or profit per customer.

**Predictive lifetime value:** projects what new customers will spend over their entire lifetime.
“Algorithms predict purchase frequency, average order value, and propensity to churn to create an estimate of the value of the customer to the business. Predictive CLV is extremely useful for evaluating acquisition channel performance, using modeling to target high value customers, and identifying and cultivating VIP customers early in their brand journey.”
Predicting Life Time Value

Predictive models
Mapping CLV

- 2M
- 3M
- 4M
- 5M
- 6M

# Products Bought

- 2
- 3
- 5
- 6
Predicting CLV

- CLV vs. # Products Bought
- Chart shows CLV values of 2M, 3M, 4M, 5M, 6M
- Users: 2, 3, 4, 5, 6

Question mark indicates uncertainty or missing data in the 5M CLV category.
Predicting CLV

![Graph showing the relationship between the number of products bought (# Products Bought) and Customer Lifetime Value (CLV). The graph has a linear trend line with points marked at 2M, 3M, 4M, 5M, and 6M for CLV, and 2, 3, 4, 5, and 6 for the number of products bought.](image)
Predicting CLV

# Products Bought

CLV:

2M 3M 4M 5M 6M
Predicting CLV

We just did a linear regression

**NOTE:** But the data is not always spread like this (that is in this nice way)
Technical Terms

Unsupervised learning

Supervised learning
Supervised vs. Unsupervised Learning

The goal of the **supervised approach** is to learn function that maps input $x$ to output $y$, given a labeled set of pairs $D = \{(x_i, y_i)\}_{i=1}^{N}$

How many will leave? (regression)
If a particular customer will leave or not? (classification)

The goal of the **unsupervised approach** is to learn “interesting patterns”
given only an input $D = \{x_i\}_{i=1}^{N}$

Categorize different types of customers?
Regression vs. Classification

If a particular customer will leave or not?

How many will leave?
Sleeping habits

4 hours of sleep

8 hours of sleep

exam performance
Simple Linear regression

![Graph showing the relationship between exam_scores and hours_slept.](image)
Simple Linear regression

How to fit the line?
Simple Linear regression

- Come close
- Come close
- Come close
Linear regression
Simple Linear regression

Task: given a list of observations $D = \{(x_i, y_i)\}_{i=1}^{N}$ find a line

$$\hat{y} = ax + b$$

that **approximates** the correspondence in the data
Simple Linear regression

Error
Simple Linear regression
Simple Linear regression

Sum of Squares: to find the optimal line

*Sum of Squares: is a type of loss function.
Simple Linear regression

Given a hours of sleep of a student predict the exam score?
Simple linear regression

Task: given a list of observations $D = \{(x_i, y_i)\}_{i=1}^{N}$ find a line that approximates the correspondence in the data

$\hat{y} = ax + b$
Simple linear regression

\[ y = \alpha + \beta x + \epsilon \]
Simple linear regression

\[ y = \alpha + \beta x + \epsilon \]

- **Intercept (bias)**
  - Shows how increases output if input increases by one unit.
  - Shows what we are not able to predict with x.

- **Coefficient (slope, or weight w)**

- **Noise (error term, residual)**
  - Shows what we are not able to predict with x.
Simple linear regression

\[ y = \alpha + \beta x + \epsilon \]
Simple linear regression: example in python

A collection of observations of the Old Faithful geyser in the USA Yellowstone National Park

df = pd.read_csv("faithful")

df.head(6)

<table>
<thead>
<tr>
<th>df</th>
<th>shape</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>eruptions</th>
<th>waiting</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3.600</td>
</tr>
<tr>
<td>1</td>
<td>1.800</td>
</tr>
<tr>
<td>2</td>
<td>3.333</td>
</tr>
<tr>
<td>3</td>
<td>2.283</td>
</tr>
<tr>
<td>4</td>
<td>4.533</td>
</tr>
<tr>
<td>5</td>
<td>2.883</td>
</tr>
</tbody>
</table>

X: length of the waiting period until the next one (in mins)

Y: the duration of the geyser eruptions (in mins)

model = LinearRegression().fit(x, y)
y_pred = model.predict(x)

What do we want to model here? i.e. What is input and output?
Simple linear regression

The fitted model is: eruptions = -1.87 + 0.08 x waiting

What is the eruption time if waiting was 70?
Simple linear regression

The fitted model is: eruptions = -1.87 + 0.08 x waiting

What is the eruption time if waiting was 70?

> -1.874016 + 70*0.075628
[1] 3.419944
> coef(model)[[1]] + coef(model)[[2]]*70
Prediction using linear regression

- Doesn’t not apply in every case
  example: not in event based scenarios
Machine learning “secret sauce”
Prediction Problem

Step 1

Training Data (with labeled information) For Learning the model

\[ X_1 \rightarrow Y_1 \]
\[ X_2 \rightarrow Y_2 \]
\[ \vdots \]
\[ X_{100} \rightarrow Y_{200} \]
Prediction Problem

Training Data
(with labeled information)
For Learning the model

X1 -> Y1
X2 -> Y2
...
X100 -> Y200

Step 1

Test Data
(hide labeled information)
For prediction

X101 -> ?
X102 -> ?
...
X110 -> ?

Step 2

Hide Y (predictive info) from the model
Prediction Problem

**Step 1**
- Training Data (with labeled information):
  - For Learning the model

  
  
  **X1 -> Y1**
  
  **X2 -> Y2**
  
  ...
  
  **X100 -> Y200**

**Step 2**
- Test Data (hide labeled information):
  - For prediction

  
  
  **X101 -> ?**
  
  **X102 -> ?**
  
  ...
  
  **X110 -> ?**

**Step 3**
- Bring back Hidden information:
  - For evaluation

  
  
  **Y101**
  
  **Y102**
  
  ...
  
  **Y110**
Simple Vs. Multiple Regression

Simple Regression
1 to 1

Multiple Regression
Many to 1

**D.V:** Dependent Variable or Predictive variable

**I.V:** Independent Variable or Input variable/features
Multiple regression (Multivariate linear regression)

Use case: Regional Delivery service

Problem: Estimate the delivery time based on

1) Total distance of the trip
2) # of deliveries that have to be made during the trip

<table>
<thead>
<tr>
<th>Total Distance</th>
<th># Deliveries</th>
<th>Delivery time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Source: Brandon Foltz learning material.
Multiple regression

Use case: Regional Delivery service

Problem: Estimate the delivery time based on

1) Total distance of the trip
2) # of deliveries that have to be made during the trip

<table>
<thead>
<tr>
<th>Total Distance</th>
<th># Deliveries</th>
<th>Delivery time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Source: Brandon Foltz learning material.
Some Considerations

- Distance
- # Deliveries

**Estimated time**
Some Considerations

- Distance
- # Deliveries
- Estimated time

Diagram showing the relationship between distance, number of deliveries, and estimated time.
Some Considerations

Independent variables could not only be related (in some proportion) with dependent variable but they could be related with each other (called as multicollinearity)

Ideally, all the independent variables to be correlated with the dependent variable but not with each other.

Some ways to avoid multicollinearity:
• Correlations
• Scatter plots
• Simple regressions.

Distance

Estimated time

# Deliveries
Some Considerations

Adding more variables does not mean will make things better: It can lead to problem of overfitting.

Independent variables could not only be related (in some proportion) with dependent variable but they could be related with each other (called as multicollinearity).

Ideally, all the independent variables to be correlated with the dependent variable but not with each other.

Some ways to avoid multicollinearity:
- Correlations
- Scatter plots
- Simple regressions.

Distance

Estimated time

# Deliveries
Multiple regression

all the same, but instead of one feature, \( x \) is a \( k \)-dimensional vector

\[
\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})
\]

the model is the linear combination of all features:

\[
\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k
\]

**NOTE:** error term is assumed to be zero
Multiple regression
Interpreting coefficients

\[ \hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k \]

Each coefficient is interpreted as the estimated change in \( \hat{y} \) corresponding to a one unit change in a variable, when all other variables are held constant.

\[ \hat{y} = 20 + 9x_1 + 10x_2 \]

\( \hat{y} \) = Estimated time
\( x_1 \) = Distance
\( x_2 \) = # deliveries

9 times is an estimate of the expected increase in estimated time in delivery time corresponding to a unit increase in distance when # deliveries are held constant.
Multivariate linear regression

Linear model requires parameters to be linear, not features!

This is linear model

\[ y = \beta_0 + \beta_1 x_1^2 + \beta_2 x_1 + \beta_3 x_2 \]

This is linear model

\[ y = \beta_0 + \beta_1 x_1^7 + \beta_2 x_1^3 + \beta_3 x_1 + \beta_4 x_2^2 \]

This is not linear model

\[ y = \beta_0 + \beta_1 x_1 + \beta_2^2 x_2 \]
Quality Assessment

- **MAE**: Mean Absolute Error
- **MSE**: Mean Square Error
- **RMSE**: Root Mean Square Error
- **MAPE**: Mean Absolute Percentage Error
- **R²**
- **RAE** (Relative Absolute Error)
- **RSE** (Relative Square Error)
MAE = \frac{1}{n} \sum \text{absolute difference between the data and the model's predictions.}

small MAE suggests the model is great at prediction, while a large MAE suggests that your model may have trouble in certain areas.

Does not indicate underperformance or overperformance of the model (whether or not the model under or overshoots actual data).
MSE/RMSE

What about outliers?

While each residual in MAE contributes proportionally to the total error, the error grows quadratically in MSE. What it means:

• outliers in our data will contribute to much higher total error in the MSE than they would in MAE.
• our model will be penalized more for making predictions that differ greatly from the corresponding actual value.

• Reference: https://www.dataquest.io/blog/understanding-regression-error-metrics/

\[ MSE = \frac{1}{n} \sum \left( y - \hat{y} \right)^2 \]

Source: https://www.dataquest.io/blog/understanding-regression-error-metrics/
MAE and (R)MSE

Loss/cost function

MAE is more robust to outliers since it does not make use of square.

With errors, MAE is steady

MSE is more useful if we are concerned about large errors.

With increase in errors, RMSE increases as the variance associated with the frequency distribution of error magnitudes also increases.

Reference: https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d
RMSE Vs. MAE

<table>
<thead>
<tr>
<th>CASE 1: Evenly distributed errors</th>
<th>CASE 2: Small variance in errors</th>
<th>CASE 3: Large error outlier</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Error</td>
<td>Error</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

MAE and RMSE for cases of increasing error variance

RMSE should be more useful when large errors are particularly undesirable.

https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d
RMSE does not necessarily increase with the variance of the errors. RMSE increases with the variance of the frequency distribution of error magnitudes.
MAPE

Mean Absolute Percentage error (MAPE) is the percentage equivalent of MAE.

MAE is the average magnitude of error produced by the model, & the MAPE is how far the model's predictions are off from their corresponding outputs on average.

$$\text{MAPE} = \frac{100\%}{n} \sum \left| \frac{y - \hat{y}}{y} \right|$$

- MAPE is biased towards predictions that are systematically less than the actual values themselves

- MAPE is biased towards predictions that are systematically greater than the actual values themselves

$$\hat{y}$$ is smaller than the actual value

\[ \begin{align*}
  n &= 1 \\
  \hat{y} &= 10 \\ y &= 20 \\
  \text{MAPE} &= 50\%
\end{align*} \]

$$\hat{y}$$ is greater than the actual value

\[ \begin{align*}
  n &= 1 \\
  \hat{y} &= 20 \\ y &= 10 \\
  \text{MAPE} &= 100\%
\end{align*} \]
R^2: What about improvement?

Coefficient of Determination

- **SSE (Sum of Squares error)** = \( \sum (y - \hat{y})^2 \)
- **SST (Sum of Squares total)** = \( \sum (y - \bar{y})^2 \)
- **SSR (Sum of Squares Regression)** = \( \sum (\hat{y} - \bar{y})^2 \)

- \( R^2 \) ranges from 0 to 1 (as mentioned in wikipedia) or from -1 to 1 (in libraries).

Reference: https://ragrawal.wordpress.com/2017/05/06/intuition-behind-r2-and-other-regression-evaluation-metrics/
R2: What about improvement?

- R2 ranges from 0 to 1 (as mentioned in wikipedia) or from -1 to 1 (in libraries).

- Equation 1: made an assumption that our model will be always better than mean model and hence will be in between mean model and the best model.

- Equation 2: in practices its possible that our model is worst than mean model and it falls on right side of the mean model.

Reference: https://ragrawal.wordpress.com/2017/05/06/intuition-behind-r2-and-other-regression-evaluation-metrics/
R2 and more

- R-squared *cannot* determine whether the coefficient estimates and predictions are biased, which is why you must assess the residual plots.

- “Adjusted R-square” penalizes you for adding variables which do not improve your existing model.

- Typically, the more non-significant variables you add into the model, the gap in R-squared and Adjusted R-squared increases.
Customer lifecycle

Where does CLV falls among 4?
Summary

• CLV:
  • Historical
  • Prediction: Techniques -- 1) Linear Regression and 2) Multiple regression.

• How to evaluate (regression) models:
  • MAE
  • MSE/ RMSE
  • R2
Demo time!

https://courses.cs.ut.ee/2019/bda/fall/Main/Practice