Neural Networks (LTAT.02.001)

University of Tartu
Institute of Computer Science

Homework 1: Lecture Materials

- These theoretical questions are about the material covered in the two lectures about "Probability and Information theory" and "Basics of Machine Learning". You can get 10 pts for correctly answering these questions. These points are added to the 20 pts you can get from the coding task.
- All questions must be answered in your own words, do not copy-paste text from the internet.
- Submit the answers to these questions through courses.cs.ut.ee. Under the practice sessions you need to select "Homework 1:theory" and upload the file. The answers must be in a .pdf file. Points can be deducted for terrible formatting or incomprehensible English.

Probability theory

Q_1 : (3pt) Some guys made a study to find out if intake of vitamin C helps you recover faster from a common cold. They asked 1000 people about their recent illness and if they consumed additional vitamin C. The results are summarized in table below:

	Recovered in <3 days	Recovered in 3-5 days	Recovered in >5 days	Total probability of dose
No vitamin C	200 participants	150 participants	150 participants	
Low dose of vitamin C	80 participants	60 participants	60 participants	
High dose of vitamin C	120 participants	90 participants	90 participants	
Total probability of recovery time				

Answer the following questions, also include explanation or calculations that led to this answer. (It is probably best if you turn the counts into probabilities first and calculate marginal probabilities)

1. What proportion of the participants took more than 5 days to recover from illness?
2. What is the probability to recover in less than 3 days given that you took no vitamin C? (HINT: it is a conditional probability)
3. What is the probability to recover in less than 3 days given that you took high dose of vitamin C?
4. Are the two variables (dose and recovery time) independent? Explain your answer.

Q_2 : (3pt) Imagine we have a classifier that classifies an image into one of 4 categories: cars, bicycles, motorbikes and pedestrians. Given an image of a bicycle, it produces the following probabilities for the 4 classes: car 0.05, bicycle 0.70, motorbike 0.20, pedestrian 0.05. Considering that the true probability distribution is [0;1,0,0], find:

- The entropy of the true distribution, the predicted distribution and a uniform distribution ([0.25,0.25,0.25,0.25]). Which one is the highest?
- Kullback-Leibler divergence between the true and the predicted distribution. (notice that $KL(p||q)$ is not equal to $KL(q||p)$)
- Sum up the entropy of the true distribution and the KL divergence calculated above to get the cross-entropy value. Calculate the cross-entropy also using the formula $CE(p,q) = -\sum_i (p_i \cdot log(q_i))$.

Machine Learning

Q_3 : (1pt) Suppose you have measured a series of pairs of values y_i and x_i, for $i = 1...N$, and you would like to establish some relation between the X and Y variables. Assume also that each sample is independent and identically distributed. Show that the maximum likelihood estimator for the Gaussian model of linear regression corresponds to minimizing the mean squared error between the prediction and the true output. Gaussian model assumes that the probability of some target value y given x is the Gaussian distribution $\mathcal{N}(w \cdot x, \sigma^2)$, where w are the weights.

In particular demonstrate that maximizing $log\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\hat{y})^2}{2\sigma^2}}$ is equivalent to minimizing $(y-\hat{y})^2$, where y is the target value, $\hat{y} = w \cdot x$ is the predicted value and variance σ^2 is considered constant.

Q_4 : (1pt) Imagine you have trained a very powerful neural network to classify if a given image has a dog or a cat on it. The network achieves 99.9% accuracy on the images that it was trained on, but has only 65% accuracy on cat/dog images it has never seen before (test images). How is this problem called? How to make the difference between training and testing error smaller (name at least one solution)?

Q_5 : (2pt) **Curse of Dimensionality:** Images are are very common data type in deep learning. Each colored image consists of pixels, each pixel is a combination of three color values - red,green,blue - that have 256 possible values each.

- How many different colors can one pixel take?
- How many different 2x2 pixel images are there?
- In the ImageNet dataset the pictures are of size 256x256 pixels. Compare the number of possible images with the number of images in the dataset (1.2 million).