Machine Learning 1DL034 Mock Exam

Question	Points	Score
1	20	
2	10	
3	4	
4	10	
Total:	44	

This is not a complete exam, but I hope that it gives you some idea on the sort of questions that I am planning to ask. Question 1 will be a lot of multiple choice questions (roughly 20 points). In order to get a 3 you will have to get 80% on question 1. If do not get 80% on question 1 then I will not mark the other questions. Your will get a 4 if you pass 50% on the remaining questions 2—?? and a 5 if you get more than 80% on the same remaining questions.

The rational is that first question should contain questions that you really should know the answer. Some of the questions are simple questions about terminology in machine learning and some of the questions will be calculations on algorithms covered in the course.

There is a lot of material in the course, I will not necessarily ask you questions on everything that was covered.

To prepare for the more advance questions, you should go through the algorithms that we have covered in the course, and you should understand what they are doing.

1. General Questions
   (a) (1 point) Which of the following problems are more suitable to regression. There is more than one answer. Please circle your answer.
      A. Predicting if a image scan contains a pedestrian.
      B. Predicting the final sale price of a house.
      C. Predicting the network bandwidth demand.
      D. Classifying a message as spam or non-spam.
   (b) (1 point) Consider the following data-sets and problems. Which data sets would require un-supervised learning. There is more than one answer. Please circle your answer.
      A. An un-labelled set of pictures of cells. Your job is to classify cells into healthy and non-healthy.
      B. A collection of images, there are no labels. Your job is to classify the image into different classes.
      C. A data set containing the closing price of every house sold in Uppsala since 2000. You are to predict the final house price of a house based on its location and other data.
   (c) (1 point) Which of the following features are categorical. There is more than one answer Please circle all the correct answers.

A. The gender of a passenger.

B. The amount the passenger paid for his or her ticker.

C. Given a set of images that are scans of handwritten digits, the digit that is contained in that image.

D. The floor area of a house or flat (apartment) in square metres.

(d) (1 point) The goal of a machine learning algorithm will find the best hypothesis that explains that training data. Please circle the correct answer. A. True B. False

(e) (1 point) You always use the same loss (error) function when training your machine learning algorithm and to evaluate the performance of your algorithm. A. True B. False

(f) (1 point) The training set is always a subset of the validation set. A. True B. False

(g) (1 point) You are developing an algorithm to detect fraud (bedrägeri). You have a legal obligation to capture as many cases as possible. If you miss-classify a case as non-fraud, and it was later found to be a fraud case, then you will be fined a very large amount of money. The output of your classification algorithm should reply **True** if the input data represents a fraudulent case. In this case would your prefer a higher

A. False Negative rate

B. False Positive rate.

(h) (2 points) Given the following data set:

x	y
1	2
2	42
3	43

You are using gradient descent to fit a linear model $h_{\theta_0,\theta_1}(x) = \theta_0 + \theta_1 x$. Which of the following expressions is the correct value of the loss (or error) function J.

A. $J(x, \theta_1, \theta_2) = (2 - (\theta_0 + \theta_1)) + 42 - (\theta_0 + 2\theta_1))$

B. More wrong choices ...

C. $J(x, \theta_1, \theta_2) = \frac{1}{2m}((2-(\theta_0+\theta_1))^2+(42-(\theta_0+2\theta_1))^2+(43-(\theta_0+3\theta_1))^2)$

(i) (2 points) You have the following data concerning the occurrence of words in spam email.

Spam (Y/N)	BitCoin Occurs	Uppsala Occurs
Y	Y	N
Y	Y	Y
N	N	N
N	N	Y
N	Y	Y

You are given an email that contains the word "BitCoin" which of the following is the correct value of $P(\text{Spam} \mid \text{BitCon})$

A.

$$\frac{1 \times \frac{2}{5}}{\frac{3}{5}}$$

B. Lots of wrong answers.

(j) (2 points) Calculating the entropy of sets for the ID3 algorithm. Again multiple choice, no calculator required.

(k) (2 points) Calculating the value $\frac{1}{2}||\overline{w}||$ given some data for a linear SVM

(l) (2 points) Similar things on logistic regression.

(m) (3 points) Hopefully this section will add up to 20. I will calibrate the points so that one point questions are things that you should just know.

2. Deeper questions on classification algorithms:

(a) (2 points) In Naive Bayes how are the joint probabilities $P(X_1, X_2 \mid Y)$ calculated. What key assumptions is required for the calculation.

(b) (2 points) Explain with an example the problem with zero-counts in Naive Bayes. Give at least one solution to avoid zero-counts.

(c) (2 points) With a k-nearest neighbour classifier explain what happens as the number of neighbour parameter increases. Explain both how the classification power improves or decreases, and how the time-complexity of the algorithm improves. You do not need to give numeric examples, but you can draw pictures.

(d) (2 points) For k-means clustering give some examples of data sets where k-means clustering will perform badly. Again you can draw pictures.

(e) (2 points) k-means clustering works by gradient descent. Does the clustering algorithm always converge on the global minimum? If so explain why, if it does not then explain why.

3. Validation and Cross validation.
   (a) (2 points) Why do you need to split your data into a training and a validation set.
   (b) (2 points) Why is it hard to tune the hyper-parameters of a machine learning algorithm. points[2] Grid-search is one way of tuning hyper-parameters. Describe gird-search and explain why it can still lead to over fitting.
4. (10 points) The rest of the questions are not really questions, but ideas for revision.
   (a) Questions on PCA, calculating the variance for a data set, what do the eigenvalues actually tell you? I won't ask you to actually calculate the eigenvalues. What is values will a confusion matrix contain. Why do we want to do PCA.

(b) Logistic regression, what does the cost function actually mean.

(c) Regularisation, what is it? Why is it necessary?

(d) Do not forget we looked at boosting for linear regression.