582631 Introduction to Machine Learning (Fall 2016)
Course examination, December 20th 2016
Examiner: Teemu Roos

Answer all the four (4) problems. The maximum score for the exam is 60 points.

You are allowed to have a calculator and a "cheat sheet" with you at the exam. The cheat sheet is a two-sided, handwritten, A4 where you can write any information whatsoever.

Please write in c l e a r handwriting. You may answer in English, Finnish or Swedish. If you use Finnish or Swedish, it will be helpful to include the English translations to any technical terms that may be ambiguous.

1. [*10 points*] Explain briefly the following terms and concepts. Your explanation should include, when appropriate, both a precise definition and a brief description of how the concept is useful in machine learning. Your answer to each subproblem should fit to roughly one third of a page of normal handwriting or less.

   (a) *Gini index* and *entropy*

   (b) *squared error* and *logarithmic loss* (*log-loss*)

   (c) *dimension reduction*

   (d) *linear regression*

   (e) *kernel trick*

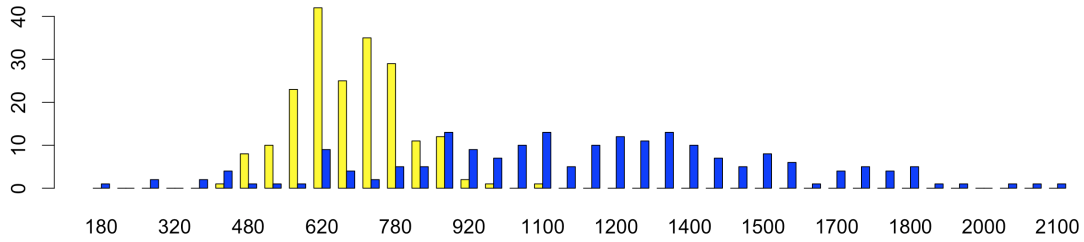
2. [*15 points*] Consider a data set with $n = 100$ observations. Imagine you learn a classification model and find that it classifies all the training examples correctly.

   (a) What can you say about the performance of your classifier on new test data? Explain what makes generalization hard. What properties of the classification method are most relevant?

   (b) Explain cross-validation.

   (c) Now suppose that instead of classification, the task would have been to estimate, for example, the median of an unknown distribution from which we have $n = 100$ data points. How would you apply resampling to measure the accuracy of an estimate computed from the given n points?

3. [*20 points*] Consider a classification task with one real-valued feature (e.g., some medical test result). Below are two histograms of the feature, X, showing $n = 200$ data point from two classes $Y = 0$ (yellow) and $Y = 1$ (blue).



(a) Just by looking at the data distributions, how would you classify three test data points with $X = 250$, $X = 500$, and $X = 1000$? What additional information about the two classes, not contained in the above histograms, would be useful?

(b) Explain how you would apply Quadratic Discriminant Analysis (QDA) to this task. Draw a diagram to explain the learned model. How would the resulting classifier classify the three test data points in item (a)?

(c) Compare the QDA classifier in this task (one-dimensional feature X) to the naive Bayes classifier. Also, compare QDA and naive Bayes in the multidimensional case where there are multiple features $X_1, \ldots, X_p$.

4. [*15 points*]

(a) For what kind of tasks can we use the *K-means algorithm*? Explain carefully what the inputs and outputs of the algorithm are, and give a very brief intuitive explanation of how the results are to be interpreted.

(b) Describe the actual K-means algorithm (Lloyd's algorithm). The description should be brief and on a high level.

(c) Define formally the objective (or cost) function that the K-means algorithm tries to minimise. Comment on how the objective function changes in the two stages of each iteration of Lloyd's algorithm.

(d) Consider the following set of data points.

Let $K = 3$ and take the three right-most points as initial cluster means (exemplars/prototypes). Simulate the algorithm for a couple of iterations. Draw the cluster assignments and the cluster means after each iteration.