T-61.3050 MACHINE LEARNING: BASIC PRINCIPLES, EXAMINATION

28 October 2011.

To pass the course you must also pass the term project. Results of this examination are valid for one year after the examination date.

This examination has five problems in two pages. Each problem is worth 6 points. Please write clearly and leave a wide left or right margin. You can have a calculator, with memory erased. No other extra material is allowed.

You can keep this paper.

1. Write about the terms below in the context of the course, e.g. what is in common and what are the differences. Use full sentences and give examples.

   (a) generative learning–discriminative learning (2 points)

   (b) parametric methods–nonparametric methods (2 points)

   (c) classification–clustering (2 points)

2. Consider a Bayesian network that has three binary variables M (trip to Mexico), S (swine flu), and F (fever). The joint distribution is $P(M, S, F) = P(M)P(S \mid M)P(F \mid S)$ and the parameters are: $P(M = 1) = 0.05$, $P(S = 1 \mid M = 0) = 0.01$, $P(S = 1 \mid M = 1) = 0.05$, $P(F = 1 \mid S = 0) = 0.01$, and $P(F = 1 \mid S = 1) = 0.9$.

   (a) Draw the graphical representation of the Bayesian network. (3 points)

   (b) Compute $P(M = 1 \mid F = 1)$, that is, the probability that one has been to Mexico if we know that she have fever. (3 points)

3. Consider a parametric regression problem

   (a) Write a pseudocode function to choose a regression model among $M_1, M_2, \ldots, M_8$ and its parameters θ given a data set of 1000 samples $\{\mathbf{x}^t, \mathbf{r}^t\}_{t=1}^{1000}$. You should implement 10-fold cross validation for model selection in your function. You can use abstract auxiliary functions such as one for estimating parameters, but you should describe each with one sentence and carefully list each function's inputs and outputs. (5 points)

   (b) Mention one advantage and one disadvantage of 10-fold cross validation when compared to basic validation. (1 point)

4. Assume that your data $\mathcal{X}$ is N d-dimensional real vectors, that is, $\mathcal{X} = \{\mathbf{x}^t\}_{t=1}^N$, $\mathbf{x}^t \in \mathbb{R}^d$. Consider the problem of reducing the dimensionality of your data to k dimensions, where $k < d$, using principal component analysis (PCA).

   (a) Write down in pseudocode how you could find the PCA representation of the data in k dimensions. (Hint: it is probably easiest to use matrix representation here. You can assume that you have access to a function that gives eigenvectors and eigenvalues of a matrix.) (4 points)

(b) What is the objective of PCA? (1 point)

(c) What is the relationship between the objective and the eigenvalues? (1 point)

5. Do three iterations of the Lloyd's algorithm for K-means clustering on the 2-dimensional data below. Use $K = 2$ clusters and the initial prototype vectors (=mean vectors) $\mathbf{m}_1 = (0.0, 2.0)$ and $\mathbf{m}_2 = (2.0, 0.0)$. Write down calculation procedure and the cluster memberships as well as mean vectors after each iteration. Draw the data points, cluster means and cluster boundary after each iteration. (6 points)

t	$\mathbf{x}^t$
1	(0.0,1.0)
2	(1.0,2.0)
3	(4.0,5.0)
4	(5.0,3.0)
5	(5.0,4.0)