T-61.3050 MACHINE LEARNING: BASIC PRINCIPLES, EXAMINATION

6 August 2012.

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 in your answer.
   (a) supervised learning–unsupervised learning (2 points)
   (b) classification–clustering (2 points)
   (c) empirical error–generalization error (2 points)
2. Consider the problem of linear regression using least squares estimates, given a data set of $\mathcal{X} = \{(r^t, x^t)\}_{t=1}^N$, where $r^t \in \mathbb{R}$ is the output (variate) to be predicted and $x^t \in \mathbb{R}$ is the input (covariate).
   (a) Write the model equation $r^t \approx g(x^t \mid \boldsymbol{\theta}) = \ldots$ and the error function $E(\boldsymbol{\theta} \mid \mathcal{X})$ to be minimized. (2 points)
   (b) Give the solution of the parameters $\boldsymbol{\theta}$ either as mathematical equations or as pseudocode. (If you have memorized the solution, explain with a few words how you could have derived it.) (2 points)
   (c) How could you estimate the prediction error for yet unseen data? (2 points)
3. Principal Component Analysis (PCA)
   (a) Do the PCA learning using the 2-dimensional data set in the table below. Describe the steps of your solution. (4 points)
   (b) Compute the proportion of variance (PoV) explained by the first principal component. (1 point)
   (c) Find the reconstruction $\hat{\mathbf{x}}$ of point $\mathbf{x} = [4.0 \;\; 7.0]^T$ with the first principal component. (1 point)

t	x_1^t	x_2^t
1	2.0	2.0
2	3.0	4.0
3	5.0	6.0

4. Consider the problem of clustering N real valued data vectors into k clusters using the Lloyd's algorithm, also known as the k-means algorithm.

   (a) Write down the Lloyd's algorithm in pseudocode. Pay attention to clearly marking the inputs and outputs of each function. Include an initialization in your algorithm. (4.5 points)

   (b) What can you say about the convergence and solutions found by the Lloyd's algorithm? How could you take this into account in practical data analysis? (1.5 point)

5. Classification tree

   (a) What is classification tree? Define it. (1 point)

   (b) Sketch the running of the vanilla ID3 algorithm with a toy data set in the figure below (binary classification task in $\mathbb{R}^2$). (4 points)

   (c) How to avoid overfitting in the vanilla ID3 algorithm? (1 point)