T-61.3050 PROBLEMS 5/2007

In T1 on 12 October 2007 at 10 o'clock.

You should solve the problems before the problem session and give the solved problems to the assistant. Please write clearly and leave a wide (left or right) margin. The solutions should be stapled together **with a cover sheet** containing your name, student number and the numbers of problems you have solved.

For the problems where a "correct" solution exists (math and algorithm questions) the assistant will present one possible solution during the session. In some cases the questions do not have a single correct answer, but the idea is that you think about the problem and are prepared to discuss it with the assistant and other students during the session.

See `http://www.cis.hut.fi/Opinnot/T-61.3050/2007/problems` for up-to-date information of the problem session.

1. (Alpaydin (2004) Ch 4, Exercise 4) Given two normal distributions $p(x \mid C_1) \sim N(\mu_1, \sigma_1^2)$ and $p(x \mid C_2) \sim N(\mu_2, \sigma_2^2)$ and $P(C_1)$ and $P(C_2)$, calculate the Bayes' discriminant points analytically.

2. (Alpaydin (2004) Ch 4, Exercise 6) For a two-class problem, generate normal samples from two classes with different variances, then use parametric classification to estimate the discriminant points. Compare these with the theoretical values.

3. Consider a binary classification problem, with N samples. Assume that the data is split in random into training and test sets of sizes $N - n$ and n, respectively. The test set can be used to approximate the generalization error (fraction of samples classified incorrectly) of a model trained using the training set. There is a tradeoff: the bigger the test set, the more accurate is the approximation of the generalization error, but the classifier is trained with less data. Approximate the confidence of the approximation of the generalization error as a function of the test set size n and the error on test set E_{TEST}. Test your approximation with the classifier of problem 2. (Hint: Binomial distribution may be relevant.)

4. (Alpaydin (2004) Ch 4, Exercise 8) When the training set is small, the contribution of variance to error may be more than that of bias and in such a case, we may prefer a simple model even though we know that it is too simple for the task. Can you give an example?