

582669 Supervised Machine Learning (Spring 2011)

Course examination, Tuesday 1 March, 16:00–19:00 A111

Examiner: Jyrki Kivinen

Answer all the problems. The maximum score for the exam is 48 points. You may answer in Finnish, Swedish or English.

When you are asked to give an algorithm, bound etc., make sure that you **explain** briefly but precisely all the relevant terms, underlying assumptions etc.

1. [8 points] Consider the Weighted Majority algorithm on a class of hypotheses $H = \{h_1, \dots, h_n\}$. Prove that if for some $M \geq 0$ there are k different hypotheses $h_i \in H$ such that for a given sequence S they all satisfy $L_{0-1}(h_i, S) \leq M$, then the loss of the Weighted Majority algorithm satisfies

$$L_{0-1}(S, \text{WM}) \leq cM \ln \frac{1}{\beta} + c \ln \frac{n}{k}.$$

Here β is the learning rate of WM and c is the constant for which we proved the inequality

$$L_{0-1}(\hat{y}_t, y_t) \leq P_t - P_{t+1}$$

where $P_t = c \ln \sum_{i=1}^n w_{t,i}$ and $w_{t,i}$ is the weight of hypothesis i at time t (with initial weights $w_{1,i} = 1$). You may assume this second inequality as known, and you do not need to care about the actual value of the constant c .

2. [4+8 points]
 - (a) Give the pseudocode for the Perceptron algorithm (the most basic online algorithm we had for linear classification, without any additions we introduced later). State also the Perceptron Convergence theorem (the most basic mistake bound for the above algorithm).
 - (b) Prove the Perceptron Convergence theorem.

Continues on the reverse side!

3. [10+4 points]

- (a) We are given m points $\mathbf{x}_i \in \mathbb{R}^d$, $i = 1, \dots, m$. The problem is to find the smallest enclosing ball in \mathbb{R}^d for the points. In other words, we need to find a centre $\mathbf{w} \in \mathbb{R}^d$ and radius r such that $\|\mathbf{w} - \mathbf{x}_i\|_2 \leq r$ holds for all i , and r is as small as possible. Formulate this as a convex optimisation problem. Obtain the dual function for the problem. Using the dual show that the problem can be kernelised; in other words, if we replace \mathbf{x}_i by some feature vectors $\psi(\mathbf{x}_i)$, the problem can be written in terms of the corresponding kernel function $k(\cdot, \cdot)$.
- (b) In the “soft” version of the smallest enclosing ball, we do not require that all the points must be inside the ball, but charge a loss for the points that are outside, with the loss being larger for points farther away from the ball. Formulate this as a convex optimisation problem.

Hint: The intention is that parts (a) and (b) relate to each other like the hard-margin and soft-margin SVM. However in part (b) you are only asked to write out a suitable formulation of the original problem (there are more than one reasonable ways of doing this), not find the dual etc.

4. [8+6 points]

- (a) Consider $X = \{-1, 1\}^n$ and recall that for $I \subseteq \{1, \dots, n\}$, we define the *monotone conjunction* $f_I = \bigwedge_{i \in I} v_i$ as a function from X to $\{-1, 1\}$ such that $f_I(\mathbf{x}) = 1$ if $x_i = 1$ for all $i \in I$, and $f_I(\mathbf{x}) = -1$ otherwise. (In the special case $I = \emptyset$ we accordingly interpreted $f_I(\mathbf{x}) = 1$ for all \mathbf{x} .) Let $H = \{f_I \mid I \subseteq \{1, \dots, n\}\}$ be the class of such monotone conjunctions. Prove that the Vapnik-Chervonenkis dimension of H is n .
- (b) Assuming you know that the Vapnik-Chervonenkis dimension of your hypothesis class is d , give a bound that relates the empirical risks $\hat{R}(h)$ and true risks $R(h)$ of hypotheses h to each other. Don't worry about the values of any constant factors etc., just try to get the form of the bound right.