Statistical Learning Theory

Exam Spring 2016

The 4 problems below will be graded on a basis of 25 points each.
Use the first answer sheet for P.1 and P.2 and the second one for P.3. and P.4.

P.1 We are given n samples $\{(\mathbf{x}_i, y_i)\}_{i=1\dots n}$, where for each $i \leq n$, $(\mathbf{x}_i, y_i) \in \mathbf{R}^d \times \{0, 1\}$. Our goal is to define a linear classifier, namely a vector $\boldsymbol{\alpha} \in \mathbf{R}^d$ and a function g, such that the true label of an observation y_i and its prediction $g(\boldsymbol{\alpha}^\top \mathbf{x}_i)$ are as close as possible on average, and, more generally, such that this vector $\boldsymbol{\alpha}$ can be used to predict the value y corresponding to a previously unseen instance $\mathbf{x}$.

1. What function g is used in the case of logistic regression (give its formula)? What about the support vector machine?
2. Describe the optimization problem you will need to solve to find the $\boldsymbol{\alpha}$ that makes the least number of mistakes when using logistic regression?

P.2

1. Prove Markov's inequality, which states that for a non-negative real-valued random variable X,
$$P(X \geq t) \leq \frac{E[X]}{t}.$$
2. Introducing notations and using formulas, explain the difference between the true risk and the empirical risk of a binary classifier. Provide an example of a classifier with high true risk but low empirical risk on a given sample, and vice-versa.

P.3

1. Define the active learning problem and the semi-supervised learning problem.
2. Briefly explain a solution for each problem.

P.4

1. Consider an online linear optimization problem with the loss function at the t-th round given as $\ell_t(\mathbf{w}_t) = \mathbf{w}_t^\top \mathbf{x}_t$, where $\mathbf{w}_t$ is the parameter vector and $\mathbf{x}_t$ is the instance at round t. Derive the update formula for the parameter $\mathbf{w}_{t+1}$ using online gradient descent.
2. Describe a theoretical guarantee of online gradient descent for the above problem.