Machine Learning: from Theory to Practice Exam

F. d'Alché-Buc and E. Le Pennec

December 2014

1 Course matter

1. Describe the logistic model used in supervised classification and explain how to optimize it.

2. Describe the maximum margin hyperplane model used in supervised classification and the optimization problem in the primal space for the case of soft-margin maximization. (Bonus: retrieve the dual form).

2 Theoretical matter

Let us consider a supervised classification task involving two random variables $X \in \mathbb{R}^p$ and $Y \in \{0, 1\}$ whose joint distribution is denoted $P(X, Y)$. Let $h : \mathbb{R}^p \to \{0, 1\}$ be a classification function and assume we consider the $0 - 1$ loss $\ell(X, Y, h(X)) = 1$ if $h(X) = Y$, 0 otherwise.

1. Prove that the best classifier, e.g. the one that minimizes the true risk $R(h) = \mathbb{E}_{X,Y}[\ell_{0-1}(X, Y, h(X))]$, is the Bayes classifier defined as follows:

$$h(x) = \arg\max_{i=0,1} P(y = i|x) \tag{1}$$

2. Now let us consider a very close task, the least square regression problem. We note: $\ell_2(X, Y, f(X)) = (Y - f(X))^2$. What is the function $f : \mathbb{R}^p \to \mathbb{R}$ that minimizes the least square loss $R(f) = \int (y - f(x))^2 p(x, y) dxdy$ (find it and prove that is the minimizer) ?

3 Practical matter

1. Why do one need to perform variable/feature selection?
2. Explain the methodology of cross-validation and its applications?
3. How does evolve the true risk (or a good estimate of it) of a classifier resulting from a training phase when increasing the complexity of the family of candidate classifiers ?
4. Is it always possible to achieve a null training error ?

4 Articles

4.1 AnyBoost

1. Explain how this is a generalization of the original AdaBoost?
2. Why is the convexity of the loss function important?
3. What could be the gain of the removing this convexity requirement?

4.2 MKL

1. What is the goal of multiple kernel learning ?
2. In terms of optimization, why is it interesting to choose the complexity term as $\sum_m \frac{||f_m||^2_{\mathcal{H}_m}}{d_m}$ in (2)? What does this choice give in the dual problem ?