Kernel Methods for Statistical Learning
Homework 1

Posted: October 16, 2014
Due: November 6, 2014

Hand your work in at the November 4 course, or send it by email to: jakob.verbeek@inria.fr

1 Logistic discriminant classification for two classes

The logistic discriminant classifier is given by:

$$p(y = +1|x) = \sigma(w^T x), \tag{1}$$

where the sigmoid function is given by

$$\sigma(z) = (1 + \exp(-z))^{-1} . \tag{2}$$

The logistic loss for a training sample x_i with class label y_i is given by:

$$L(y_i, w^T x_i) = -\log p(y_i|x_i). \tag{3}$$

Exercise 1.a : Show that $p(y = -1|x) = \sigma(-w^T x)$.

Exercise 1.b : Derive that the gradient of the logistic loss has the form

$$\nabla_w L(y_i, w^T x_i) = -y_i\big(1 - p(y_i|x_i)\big)x_i. \tag{4}$$

Exercise 1.c : Show that the logistic loss function is convex.

2 Manipulation with kernels

The following exercises require basic manipulations with kernels.

Exercise 2.a : Given the kernel $k(x, y) = \sum_{i=1}^{p} \min(x(i), y(i))$ for p-dimensional vectors of non-negative integers $x, y \in \mathbb{N}_+^p$. Show that $k(x, y)$ is a positive definite kernel.

Exercise 2.b : Let $\mathcal{H}$ be the set of second-order polynomials over real numbers, *i.e.* for $x \in \mathbb{R}$, we have $\mathcal{H} = \{f_{a,b,c}(x) : f(x) = ax^2 + bx + c\}$, forming a Hilbert space with inner product $\langle f_1, f_2\rangle_H = a_1a_2 + b_1b_2/2 + c_1c_2$. Show that $\mathcal{H}$ is a reproducing kernel Hilbert space.

3 Bias-variance decomposition for linear regression

Consider a uni-dimensional linear regression problem where we try to estimate a function $f(x) = \theta x$ from a set of training data (x_i, y_i) for $i = 1, \ldots, n$, and $\theta, x_i, y_i \in \mathbb{R}$. Suppose that the data is i.i.d. sampled from the model $y = \bar{\theta}x + \epsilon$, where ϵ is drawn from a zero-mean distribution with variance σ^2. We estimate the unknown parameter $\bar{\theta}$ by means of a penalized empirical risk minimisation as:

$$\hat{\theta} = \arg\min_{\theta} \left\{ \lambda \frac{1}{2}\theta^2 + \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta x_i)^2 \right\}. \tag{5}$$

Exercise 3.a : Show that the estimator can be obtained in closed form as:

$$\hat{\theta} = \left(\sum_{i=1}^{n} x_i^2 + \lambda\right)^{-1} \sum_{i=1}^{n} x_i y_i. \tag{6}$$

Exercise 3.b : Show that the bias of the estimator is given by

$$\mathbb{E}_{p(\epsilon)}\left[\hat{\theta} - \bar{\theta}\right] = -\lambda \left(\sum_{i=1}^{n} x_i^2 + \lambda\right)^{-1} \bar{\theta} \tag{7}$$

Exercise 3.c : Show that the variance of the estimator is given by

$$\mathbb{E}_{p(\epsilon)}\left[\left(\hat{\theta} - \mathbb{E}_{p(\epsilon)}\left[\hat{\theta}\right]\right)^2\right] = \sigma^2 \left(\sum_{i=1}^{n} x_i^2 + \lambda\right)^{-2} \left(\sum_{i=1}^{n} x_i^2\right) \tag{8}$$

Exercise 3.d : Derive similar results as in exercises 3.a, 3.b, and 3.c, for the case where $\theta, x_i \in \mathbb{R}^p$ and $y \in \mathbb{R}$.