Wissenschaftliches Rechnen II/Scientific Computing II

Sommersemester 2016

Prof. Dr. Jochen Garcke
Dipl.-Math. Sebastian Mayer

universitätbonn

Exercise sheet 6

To be handed in on **Thursday, 02.06.2016**

1 Group exercises

G 1. (Bayesian analysis of linear regression)

Consider the standard linear regression model

$$y_i = x_i^T w + \varepsilon_i, \quad i = 1, \ldots, n.$$

where $X = (x_1, \ldots, x_n) \in \mathbb{R}^{d \times n}$ is the matrix of given input vectors, $w \in \mathbb{R}^d$ the unknown weight vector, and the ε_i are i.i.d with $\varepsilon_i \sim \mathcal{N}(0, \sigma_n^2)$.

a) Determine the probability density of $y = (Y_1, \ldots, Y_n)$.

b) The Bayesian approach is to specify a *prior* distribution over w, which expresses the belief about the value of w *before* observing the data. Assume $w \sim \mathcal{N}(0, \Sigma_p)$ with covariance matrix $\Sigma_p \in \mathbb{R}^{d \times d}$. Derive via Bayes' rule the *posterior* density of W, which expresses our beliefs about the value of w after observing the concrete data $y = (y_1, \ldots, y_n)$. Determine also the posterior density $p(y_*|y)$ of the predicted value $y_* = x_*^T w$ given a new data point x_*.

c) Show that $E[y_*] = x_*^T \Sigma_p X(K + \sigma_n^2)^{-1} y$, where $K = X^T \Sigma_p X$. Make a connection between the Bayesian approach and regularization.

Solution. See Rasmussen/Williams, Section 2.1. □

G 2. You are given a random vector $U \sim \mathcal{N}(0, I_d)$, that is, U is standard normally distributed and takes values in $\mathbb{R}^d$. For given mean $m \in \mathbb{R}^d$ and covariance $K \in R^{d \times d}$, find a transformation $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ such that $\varphi(U) \sim \mathcal{N}(m, K)$.

Solution. See Rasmussen/Williams, Appendix A.2. □

G 3. (Lemma 46 revisited)

Assume to be given data $(x_1, y_1), \ldots, (x_n, y_n)$ and a Hilbert space $\mathcal{H}$ with kernel k. Let $f_{(x_n, y_n)}$ be the solution of

$$\min_{f \in \mathcal{H}} \sum_{i=1}^{n-1} (f(x_i) - y_i)^2 + \lambda \|f\|_k.$$

Let $\tilde{y}_n = f_{(x_n, y_n)}(x_n)$. Give an alternative proof of Lemma 46 based on the representer theorem. To this end, consider the system of linear equations $(K + \lambda I_n)\tilde{\alpha} = \tilde{y}$, where $\tilde{y}_i = y_i$ for $i < n$, and show that $\tilde{\alpha}_n = 0$.

Solution. Let $X = \{x_1, \ldots, x_n\}$ and $\tilde{X} = X \setminus \{x_n\}$. Let $K(X,Y) = (k(x_i, x_j))_{x_i \in X, x_j \in Y}$ and write $G(X,X) = K(X,X) + \lambda I$. By the representer theorem the solution can be written as $f_{(x_n,y_n)}(x) = \sum_{i=1}^n \tilde{\alpha}_i k(x, x_i)$ and the coefficient vector $\tilde{\alpha}$ is determined by

$$G(X,X)\tilde{\alpha} = \tilde{y}. \tag{1}$$

At the same time, we have by the representer theorem $\tilde{y}_n = f_{(x_n,y_n)}(x_n) = K(x_n, \tilde{X})\alpha$, where $\alpha \in \mathbb{R}^{n-1}$ is the solution of $G(\tilde{X}, \tilde{X})\alpha = (y_1, \ldots, y_{n-1})^T$. Further, we can write

$$G(X,X) = \begin{pmatrix} G(\tilde{X}, \tilde{X}) & K(x_n, \tilde{X})^T \\ K(x_n, \tilde{X}) & k(x_n, x_n) + \lambda \end{pmatrix}.$$

Hence, (1) can be written as

$$\begin{pmatrix} G(\tilde{X}, \tilde{X}) & K(x_n, \tilde{X})^T \\ K(x_n, \tilde{X}) & k(x_n, x_n) + \lambda \end{pmatrix} \tilde{\alpha} = \begin{pmatrix} G(\tilde{X}, \tilde{X})\alpha \\ K(x_n, \tilde{X})\alpha \end{pmatrix}.$$

Thus, $\tilde{\alpha} = (\alpha, 0)$ solves the problem. $\square$