N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
Exercise Sheet 5
Due on Mon, December 5, 12:15

If you turn in your solutions via email, please send to:
kamyshanska@fias.uni-frankfurt.de
Include [ML1-1617] in the subject line, turn in a single PDF file and include your name in the file name.

Exercise 1. *Starting from the sum-of-squares error function*

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\}^2$$

derive the maximum likelihood solution for the parameters

$$\mathbf{w}_{ML} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{t}$$

where

$$\mathbf{\Phi} = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \cdots & \phi_{M-1}(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \cdots & \phi_{M-1}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \phi_1(\mathbf{x}_N) & \cdots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix},$$

is the design matrix with basis functions $\phi_j(\mathbf{x}_i)$, $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ *the vectors of input training data and* $\mathbf{t} = \{t_1, \ldots, t_n\}$ *corresponding output training values.*

3 points

Exercise 2. *Consider a data set in which each data point* $(\mathbf{x}_n, t_n)$ *is associated with a weighting factor* $r_n > 0$, *so that the sum-of-squares error function becomes*

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n} r_n(t_n - \mathbf{w}^T\phi(\mathbf{x_n}))^2$$

Find an expression for the solution $\mathbf{w}^*$ *that minimizes this error function. Give two alternative interpretations of the weighted sum-of-squares error function in terms of (i) data dependent noise variance and (ii) replicated data points.*

2 points

Exercise 3. *Generate own data sets, e.g. using* $t = f(x) + 0.2\epsilon$ *with* $f(x) = sin(2\pi x)$ *and* $\epsilon \sim \mathcal{N}(0,1)$, *and illustrate the bias-variance decomposition by fitting a polynomial model* $y(x;w) = \sum_{i=0}^{r} w_i x^r$ *to many different data sets* $D_1, \ldots, D_L$, *each of length* N.
Let $w^{*,D}$ *denote the parameters minimizing the mean squared error on data set* D. *Then,*

$$\begin{aligned} \textbf{bias}^2 &\approx \frac{1}{L}\sum_l \frac{1}{N}\sum_n (\bar{y}(x) - f(x))^2 \\ \textbf{variance} &\approx \frac{1}{L}\sum_l \frac{1}{N}\sum_n (y(x;w^{*,D_l}) - \bar{y}(x))^2 \end{aligned}$$

where $\bar{y}(x) = \frac{1}{L}\sum_l y(x;w^{*,D_l})$.

5 points