N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
3. Exercise Sheet

Note: You can use any programming language and machine learning toolbox of your choice.

Exercise 1. *Implement the polynomial regression example from the lecture. Make plots illustrating overfitting, i.e. plot the training and testing error over the degree of the polynomial, using the data sets from the course website* `http://www.ccc.cs.uni-frankfurt.de/teaching/machine-learning-ws-201516/`. *Start by implementing the pseudo-inverse according to the formula from the lecture. Does it work? How does your toolbox solve for the regression weights?*

4 points

Exercise 2. *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 $\bar{y}(x) = \frac{1}{L}\sum_l y(x; w^{*,D_l})$.

3 points

Exercise 3. *Consider a multi-variate normal distribution on* $\mathbf{x} \in \mathbb{R}^K$ *with mean* μ *and covariance matrix* Σ. *Its probability density is given by*

$$p(\mathbf{x}|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^K \det|\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)}$$

Now, assume that $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$ *is split into two parts. Correspondingly, the mean and covariance are split as*

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \textit{and} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Show that the conditional distribution $p(\mathbf{x}_2|\mathbf{x}_1)$ *is also multi-variate Gaussian with*

$$\mu_{\mathbf{x}_2|\mathbf{x}_1} = \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \mu_1) \quad \textit{and} \quad \Sigma_{\mathbf{x}_2|\mathbf{x}_1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

3 points