N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
Exercise Sheet 6
Due on Mon, December 12, 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. *'Completing the square': suppose you encounter an expression*

$$\frac{1}{2}\mathbf{w}C\mathbf{w} + \mathbf{b}^T\mathbf{w} + a \tag{1}$$

with a symmetric square matrix C, vectors $\mathbf{w}$ and $\mathbf{b}$, and constant a. Show that you can bring this into the form

$$\frac{1}{2}(\mathbf{w} - \mathbf{m})^T C(\mathbf{w} - \mathbf{m}) + u\,, \tag{2}$$

where $\mathbf{m} = -C^{-1}\mathbf{b}$ and $u = a - \frac{1}{2}\mathbf{b}^T C^{-1}\mathbf{b}$. Hint: insert $\mathbf{m}$ and u into the expression (2) above. **1 point + 2 bonus points**

Exercise 2. *Now use the method of 'completing the square' in the exponential to derive the N-dimensional posterior distribution for Bayesian regression. Assume a prior of the form*

$$p(\mathbf{w}) = \mathcal{N}(\boldsymbol{w}|\mathbf{0}, \alpha^{-1}\boldsymbol{I})$$

and the likelihood

$$p(\boldsymbol{t}|\mathbf{X}, \boldsymbol{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|\boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_n), \beta^{-1})\,,$$

and show that the posterior is given by

$$p(\boldsymbol{w}|\boldsymbol{t}) = \mathcal{N}(\boldsymbol{w}|\boldsymbol{m}_N, \mathbf{S}_N)$$

with

$$\boldsymbol{m}_N = \beta\mathbf{S}_N\mathbf{\Phi}^T\boldsymbol{t}$$

and

$$\mathbf{S}_N^{-1} = \alpha\boldsymbol{I} + \beta\mathbf{\Phi}^T\mathbf{\Phi}\,,$$

where $\mathbf{\Phi}$ is the design matrix. **1 point + 3 bonus points**

N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
Exercise Sheet 6
Due on Mon, December 12, 12:15

Exercise 3. *Reproduce the plots shown in Figure 1 (from the book by CM Bishop (Springer Series), p.155), using a slightly more complicated model. Generate your own synthetic data from the function*

$$f(x, \mathbf{a}) = a_0 + a_1 x + a_2 x^2$$

with parameter values $a_0 = -0.3$, $a_1 = 0.5$, $a_2 = 0.4$ *by first choosing values of* x_n *from the uniform distribution* $U(x|-1,1)$, *then evaluating* $f(x_n, \mathbf{a})$, *and finally adding Gaussian noise with standard deviation of s=0.2 to obtain the target values* t_n. *The goal is to recover the values of* a_0, a_1 *and* a_2 *from such data, and to explore the dependence on the size of the data set. To achieve this, assume a model in which individual data points are generated by*

$$p(t|x, \mathbf{w}) = \mathcal{N}(t|y(x, \mathbf{w}), s^2) \ ,$$

where $y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2$ *with weights* $\mathbf{w}$ *to be estimated and a fixed standard deviation* $s = 0.2$, *i.e. assumed to be known. The likelihood is then given by*

$$p(\mathbf{t}|X, \mathbf{w}) = \prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n, \mathbf{w}), s^2) \ .$$

Finally, assume a Gaussian distributed prior $p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha)$ *with* $\alpha = 2$. *Generate two plots analog to those shown in Figure 1, one for* (w_0,w_1) *and one for* (w_1,w_2). *Describe and interpret these plots thoroughly.*

8 points

N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
Exercise Sheet 6
Due on Mon, December 12, 12:15

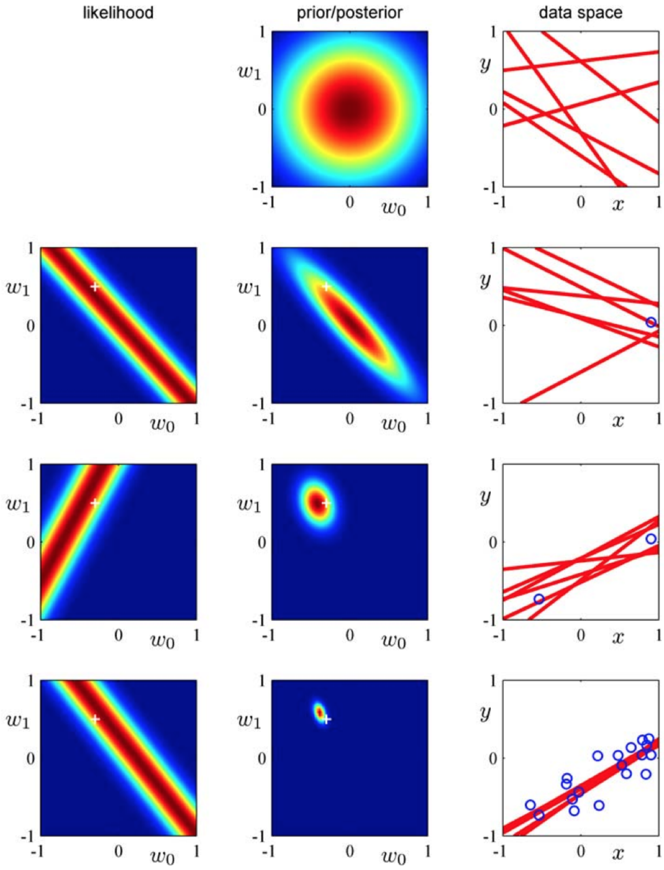

Figure 1: Illustration of sequential Bayesian learning for a linear model.