N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
Exercise Sheet 7
Due on Mon, December 19, 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. *Reproduce the plots shown in Figure 1 (from the book by CM Bishop (Springer Series), p.157-158). Consider a target variable* t *given by a deterministic function* $y(\mathbf{x}, \mathbf{w})$ *depending on input* $\mathbf{x}$ *and parameters* $\mathbf{w}$ *with additive Gaussian noise so that*

$$t = y(\mathbf{x}, \mathbf{w}) + \epsilon \ ,$$

where ϵ *is a zero mean Gaussian random variable with precision parameter* β*. This can also be written as*

$$p(t|\mathbf{x}, \mathbf{w}, \beta) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \beta^{-1}) \ .$$

Choosing a Gaussian prior

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

the predictive distribution is also Gaussian and given by

$$p(t|\mathbf{x}, \mathbf{t}, \mathbf{X}, \alpha, \beta) = \mathcal{N}(t|\mathbf{m}_N^T \phi(\mathbf{x}), \sigma_N^2(\mathbf{x}))$$

with mean

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

and variance

$$\sigma_N^2(\mathbf{x}) = \frac{1}{\beta} + \phi(\mathbf{x})^T \mathbf{S}_N \phi(\mathbf{x}) \, ,$$

where the matrix $\mathbf{S}_N$ *is defined as*

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

the vector of basis functions given by

$$\phi(\mathbf{x}_n) = (\phi_0(\mathbf{x}_n), \phi_1(\mathbf{x}_n), \cdots, \phi_{M-1}(\mathbf{x}_n))^T \ ,$$

the matrix of basis functions given by

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

with the vectors of input training data $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ *and corresponding output training values* $\mathbf{t} = \{t_1, \ldots, t_n\}$*, and the value* t *to be predicted for a new input* $\mathbf{x}$*.*

Generate synthetic data from

$$f(x) = \sin(2\pi x) - \cos(\pi x)$$

where $x \in [0, 1]$ *(i.e. the inputs* x_n *are one-dimensional and randomly drawn from a uniform distribution over the interval* $[0, 1]$*) and add Gaussian noise with some standard deviation* $\beta^{-1/2}$ *(try out different values) to the data points generated. Explore data sets of various size, e.g.* $N = 2$*,* $N = 4$*,* $N = 10$ *and* $N = 100$*. Plot an analog to Figure 1 by computing the predictive distribution for different data samples using the recipe above and then describe and interpret it thoroughly. Consider a model consisting of* 1 *constant function* ϕ_0 *and 8 Gaussian basis functions (thus* M *is 9-dimensional) of the form*

$$\phi_j(x) = \exp\left\{-\frac{(x - \mu_j)^2}{2s^2}\right\}$$

with identical width s *and means* μ_j *equally distributed between 0 and 1.*

10 points

Exercise 2. *Draw samples from the posterior distribution over* $\mathbf{w}$ *and then plot the corresponding functions* $y(x, \mathbf{w})$ *analog to those shown in Figure 2 for your solutions obtained in the previous exercise. The posterior distribution is given by*

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

To sample from this multivariate Gaussian distribution, apply the technique learned in one of the previous problem sheets. Describe and interpret these plots thoroughly.

5 bonus points

N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
Exercise Sheet 7
Due on Mon, December 19, 12:15

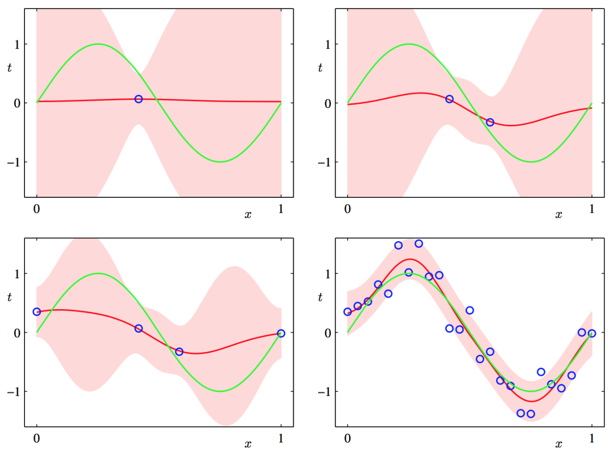

Figure 1: Predictive distribution for a model comprising a linear combinationq of Gaussian basis functions using data generated by a sinusoid. For each plot, the red curve shows the mean of the corresponding Gaussian predictive distribution, and the red shaded region spans one standard deviation either side of the mean.

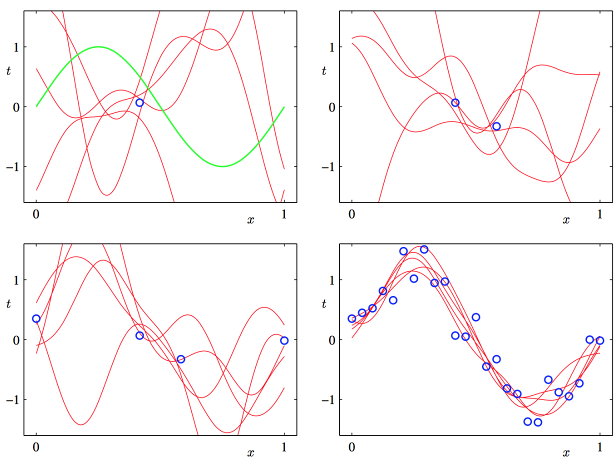

Figure 2: Plots of the function $y(x, \mathbf{w})$ using samples from the posterior distributions over $\mathbf{w}$ corresponding to the plots in Figure 1.