Machine Learning 2
Exercise Sheet 2

Prof. Dr. Dr. Lars Schmidt-Thieme, Brad Baker
Information Systems and Machine Learning Lab
University of Hildesheim

April 19th, 2017

Submission until April 25th, 8:00 AM by learnweb.
Please put your name in all filenames and somewhere visible in the margins of the pdf.
Non-pdf submissions for non-programming exercises will not be graded.

It is highly recommended you read Murphy chapter 15, especially the starting sections. This exercise sheet will focus on topics discussed in this chapter. If you are having trouble understanding Kernel functions, see Murphy chapter 14.

For further reading on Gaussian Processes specifically, see Rasmussen & Williams' 2006 book "Gaussian Processes for Machine Learning" (`http://www.gaussianprocess.org/gpml/chapters/RW.pdf`). A link to this pdf has been provided on the Moodle.

Exercise 3: Gaussian Processes (10 Points)

a) (3 points) Briefly describe the role that **Kernel functions** play in modelling Gaussian processes. Why are the Kernel functions used in Gaussian Processes required to be positive definite?

b) (7 points)
For $x \in [0, 5]$, compute the Covariance function of a Gaussian Process using three different Kernels:

$$k_1(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2}\right)$$

$$k_2(x_1, x_2) = (x_1^\top x_2 + 2)^2$$

$$k_3(x_1, x_2) = \exp(-|x_1 - x_2|)$$

Then, sample a set of 10 different latent functions f according to the Gaussian process prior:

$$f \sim \mathcal{N}(0, K)$$

for all three different Kernel/Covariance Matrices and plot them. What are the differences?

Exercise 4: Gaussian Process Regression (10 Points)

a) (2 points) The general form of the squared exponential kernel is given as

$$k(x_1, x_2) = \sigma_f^2 \exp\left(-\frac{\|x_p - x_q\|^2}{2l^2}\right) + \sigma_y^2 \delta_{pq}$$

.

Identify each of the following variables, and specify whether or not the variable is a kernel hyper-parameter: $\sigma_f, x_p, x_q, l, \sigma_y, \delta_{pq}$.

b) (5 points)

Given are two training examples

$$X = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

with ground truth given as:

$$y = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Learn a Gaussian Process on the model and predict the mean and average for the instance $x = 0$. Use the squared exponential kernel given above with $l = 1$, $\sigma_f = 1$, $\sigma_y = 0$. Plot the model and add the standard deviation.

The underlying data generation function is $f(x) = x^2$ and therefore the true label for $x = 0$ is 0. Relearn the model including the new data instance and plot it again.

c) (3 points) Identify a potential computational bottleneck in the Gaussian Process regression model used above. What steps could one take to avoid this bottleneck?

Bonus 2: Gaussian Processes in R (10 points)

a) (5 points) For the training data provided in "NoisyObs_train.csv", plot the results of a 1d Gaussian Process using the squared exponential kernel. Provide plots for at a small grid-search with least 2 different choices for each hyper-parameters (l, σ_f, σ_y). Discuss the changes you see within your search-space, and with respect to predicting the test data in "NoisyObs_test.csv". (For some coding help see `https://github.com/probml/pmtk3/blob/master/demos/gprDemoChangeHparams.m`).

b) (5 points) (from "Gaussian Processes for Machine Learning") "Replicate the generation of random functions from Figure 2.2 in "Gaussian Processes for Machine Learning". Use a regular (or random) grid of scalar inputs and the covariance function from eq. (2.16). Hints on how to generate random samples from multi-variate Gaussian distributions are given in section A.2. Invent some training data points, and make random draws from the resulting GP posterior using eq. (2.19)."

(For a big hint, see `https://github.com/probml/pmtk3/blob/master/demos/gprDemoNoiseFree.m`)

A pdf of "Gaussian Processes for Machine Learning" has been provided on the Moodle for reference.