4F13 Probabilistic Machine Learning: Coursework #1: Gaussian Processes

Carl Edward Rasmussen and Hong Ge

Due (for non-MLMI students): 12:00 noon, Friday Nov 8th, 2024, online via moodle

Your answers should contain an explanation of what you do, and max 2-4 central commands to achieve it (but complete listings are unnecessary and discouraged). You must give an *interpretation* of what the numerical values and graphs you provide *mean* – why are the results the way they are, and what are the consequences? Explain your reasoning. **Each question should be labelled and answered separately and distinctly.** Total combined length of answers must not exceed 5 sides of a4 (plus cover page), minimum 11pt font, 1 inch margins.

You need the Gaussian Processes for Machine Learning (GPML) toolbox (version 4.2) for matlab and octave. Get the toolbox and walk through the documentation concerning regression from the Gaussian Process Web site at `www.gaussianprocess.org/gpml/code` Note, that sometimes hyperparameters are encoded using their logarithms (to avoid having to deal with *constrained* optimization for positive parameters), but you will want to report them in their natural domain. All logs are natural (ie, base e). All questions carry approximately equal weight.

a) Load data from `cw1a.mat`. Train a GP with a squared exponential covariance function, `covSEiso`. Start the log hyper-parameters at `hyp.cov = [-1 0]; hyp.lik = 0;` and minimize the negative log marginal likelihood. Show the 95% predictive error bars. Explain the values of the optimized hyperparameters and the shape of the predictive error bars.

b) How can you find out whether the hyper parameter optimum is unique, or whether there maybe other local optima? If there are local optima, find some, and explain what the model isdoing in each case. Which fit is best, and why? Quantify how confident are you about thisand why?

c) Train instead a GP with a periodic covariance function, `covPeriodic`. Show the fit. Compare the behaviour of the error-bars, with a). Do you think the data generating mechanism (apart from the noise) was strictly periodic[1]? Carefully discuss the evidence for or against periodicity.

d) Generate random (essentially) noise free functions evaluated at `x = linspace(-5,5,200)';` from a GP with the following covariance function: `{@covProd, {@covPeriodic, @covSEiso}}`, with covariance log hyperparameters `hyp.cov = [-0.5 0 0 2 0]`. In order to apply the Cholesky decomposition to the covariance matrix, you may have to add a small diagonal matrix, for example `1e-6*eye(200)`, why? Plot some sample functions. Carefully explain the relationship between the properties of those random functions and the form of the covariance function.

e) Load `cw1e.mat`. This data has 2-D input and scalar output. Visualise the data, for example using `mesh(reshape(x(:,1),11,11),reshape(x(:,2),11,11),reshape(y,11,11));` Rotate the data, to get a good feel for it. Compare two GP models of the data, one with `covSEard` covariance and the other with `{@covSum, {@covSEard, @covSEard}}` covariance. For the second model be sure to break symmetry with the initial hyperparameters (eg by using `hyp.cov = 0.1*randn(6,1);`).

   Compare the models: give a careful quantitative interpretation of the relationship between data fit, model complexity and marginal likelihood for each of the two models; which model is best and why, explain your reasoning.

[1]By strictly periodic, is meant a function where there exists a p such that $f(x) = f(x + np)$ for integer n and all x, not just a function which "goes up and down".