CSE 515T (Spring 2015) Assignment 3

Due Monday, 16 March 2015

1. (Sampling from a multivariate Gaussian distribution). Assume you can sample from a univariate standard normal distribution $\mathcal{N}(0, 1^2)$. In the last assignment, this allowed you to sample trivially from a multivariate standard normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$. Use the closure of the Gaussian distribution to affine transformations to derive a procedure to sample from an arbitrary multivariate Gaussian distribution

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

   using an oracle that can draw a sample from the univariate standard normal distribution.

2. (Sampling from a Gaussian process). Consider the squared exponential covariance function

$$K(x, x'; \lambda, \ell) = \lambda^2 \exp\left(-\frac{\|x - x'\|^2}{2\ell^2}\right).$$

   Take a dense grid (somewhere around 100-1 000 points) in the domain $x \in [-4, 4]$. For several values of (λ, ℓ) of your choosing, draw five samples from a Gaussian process prior with mean function $\mu(x) = 0$ and this covariance. Plot your samples.

   Note: you may encounter numerical problems here. The most common issues are that the sample covariance $\mathbf{K}$ is not quite symmetric or that it is not quite positive definite. To ensure positive definiteness, you may need to add a small constant (on the order of 10^{-6} perhaps) to the diagonal to the covariance matrix $\mathbf{K}$ in case of numerical instability. If the matrix is not exactly symmetric, a simple trick to symmetrize it is to take $\frac{1}{2}(\mathbf{K} + \mathbf{K}^\top)$.

3. (Gaussian process regression). Consider again the data from question 2 of assignment 2:

$$\mathbf{x} = [-2.26, -1.31, -0.43, 0.32, 0.34, 0.54, 0.86, 1.83, 2.77, 3.58]^\top;$$
$$\mathbf{y} = [1.03, 0.70, -0.68, -1.36, -1.74, -1.01, 0.24, 1.55, 1.68, 1.53]^\top.$$

   Fix the noise variance at $\sigma^2 = 0.5^2$.

   - Examining a scatter plot of the data, guess which values of (λ, ℓ) in the above covariance (if any) might explain this data well.
   - Perform Gaussian process regression for these data on the interval $x_* \in [-4, 4]$ using the squared exponential covariance above for the same set of hyperparameters (λ, ℓ) above. Plot the posterior mean and the pointwise 95% credible interval for each. Which predictions look the best?
   - Visualize the model evidence $p(\mathbf{y} \mid \mathbf{x}, \lambda, \ell, \sigma^2)$ as a function of (λ, ℓ). You can choose to make, for example, a heatmap or a contour plot of this function. Probably it will be a good idea to compute and plot the logarithm of the model evidence rather than the model evidence directly.
   - Are there any settings of the hyperparameters that explain the data better than the best polynomial model from question 2 of assignment 2?

4. (Optimization of mean hyperparameters). Assume the mean function of a Gaussian process has a parameter vector θ, $\mu(\mathbf{x}; \theta)$. Given a fixed covariance function K, noise variance σ^2, and dataset $\mathcal{D} = (\mathbf{X}, \mathbf{y})$, compute the partial derivative of the log marginal likelihood with respect to the ith mean hyperparameter θ_i:

$$\frac{\partial}{\partial \theta_i} \log p(\mathbf{y} \mid \mathbf{X}, \theta, \sigma^2).$$

Consider the data from the last question. Keep the covariance hyperparameters (λ, ℓ) fixed to the best values you found. Replace the mean function μ with a constant, potentially non-zero mean function $\mu(x; \theta) = \theta$. Use the above gradient you computed above to optimize the marginal likelihood of the data as a function of θ. Is the best value the sample mean of $\mathbf{y}$? Should we expect it to be?