CSE 515T (Spring 2019) Assignment 2

Due Monday, 25 March 2019

1. (Curse of dimensionality.) Consider a d-dimensional, zero-mean, spherical multivariate Gaussian distribution:
$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{I}_d).$$

   Equivalently, each entry of $\mathbf{x}$ is drawn iid from a univariate standard normal distribution.

   In familiar small dimensions ($d \leq 3$), "most" of the vectors drawn from a multivariate Gaussian distribution will lie near the mean. For example, the famous 68–95–99.7 rule for $d = 1$ indicates that large deviations from the mean are unusual. Here we will consider the behavior in larger dimensions.

   - Draw 10 000 samples from $p(\mathbf{x})$ for each dimension in $d \in \{1, 5, 10, 50, 100\}$, and compute the length of each vector drawn: $y_d = \sqrt{\mathbf{x}^\top \mathbf{x}} = (\sum_i^d x_i^2)^{1/2}$. Estimate the distribution of each y_d using either a histogram or a kernel density estimate (in MATLAB, hist and ksdensity, respectively). Plot your estimates. (Please do not hand in your raw samples!) Summarize the behavior of this distribution as d increases.
   - The true distribution of y_d^2 is a chi-square distribution with d degrees of freedom (the distribution of y_d itself is the less-commonly seen chi distribution). Use this fact to compute the probability that $y_d < 5$ for each of the dimensions in the last part.
   - For $d = 1\,000$, compute the 5th and 95th percentiles of y_d. Is the mean $\mathbf{x} = \mathbf{0}$ a representative summary of the distribution in high dimensions? This behavior has been called "the curse of dimensionality."

2. (Laplace approximation.) Find a Laplace approximation to the gamma distribution:

$$p(\theta \mid \alpha, \beta) = \frac{1}{Z} \theta^{\alpha-1} \exp(-\beta\theta).$$

   Plot the approximation against the true density for $(\alpha, \beta) = (3, 1)$.

   The true value of the normalizing constant is

$$Z = \frac{\Gamma(\alpha)}{\beta^\alpha}.$$

   If we fix $\beta = 1$, then $Z = \Gamma(\alpha)$, so we may use the Laplace approximation to estimate the Gamma function. Analyze the quality of this approximation as a function of α.

   Read the Wikipedia article about Stirling's approximation. Do you see a connection?

3. (Gaussian process regression). Consider the following data:

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

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

   - Examining a scatter plot of the data, guess which values of (λ, ℓ) in the squared exponential 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 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 log model evidence $\log p(\mathbf{y} \mid \mathbf{x}, \lambda, \ell, \sigma^2)$ as a function of (λ, ℓ). You can choose to make, for example, a heatmap or a contour plot of this function.