CSE 515T (Spring 2018) Assignment 2

Due Monday, 26 February 2018

1. Suppose you have a standard normal belief about an unknown parameter θ, $p(\theta) = \mathcal{N}(\theta; 0, 1^2)$. You are asked to give a point estimate $\hat{\theta}$ of θ, and are told the penalty for overestimation is more lenient than for underestimation.

$$\ell(\hat{\theta}, \theta) = \begin{cases} (\theta - \hat{\theta})^2 & \hat{\theta} < \theta; \\ \hat{\theta} - \theta & \hat{\theta} \geq \theta \end{cases},$$

   What is the Bayesian estimator?

   Consider the following generalization of the above loss, with a constant multiplicative term $c \geq 0$ on the second term:

$$\ell(\hat{\theta}, \theta; c) = \begin{cases} (\theta - \hat{\theta})^2 & \hat{\theta} < \theta; \\ c(\hat{\theta} - \theta) & \hat{\theta} \geq \theta \end{cases}.$$

   Plot the Bayesian estimator as a function of c; $0 < c < 10$. Interpret the results.

   What should you do if $c = 0$?

2. (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."

3. (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?