Stat 462/862 Assignment 3

(Due on Nov 20, 2015)

1. Consider the data in Problem 3 in the 2nd assignment. Randomly partition the data into two parts (with equal sample size), one as a training set, and the other as the test set. Compare the the performance of logistic regression, LDA, and QDA in terms of the test error. Interpret your results.

2. Let X be a random variable normally distributed with mean $\mu = 2$ and variance $\sigma^2 = 5$. The follow simulation studies the property of certain confidence interval of μ.

   (a) Let n be the sample size and $nsim$ be the number of simulations. Let α be significance level. Write a function in which the inputs are $\mu, \sigma, n, nsim, \alpha$, the output is $(1-\alpha)100$ percent confidence interval for μ at in each simulation.

   (b) Consider the function in part (a). Let w be the coverage of the $(1-\alpha)100$ percent confidence interval. Define

   $w =$ (the frequency that the confidence interval contains the true mean μ)$/nsim$.

   Compute the coverage for the following setting (i) $n = 10, nsim = 1000, \alpha = 0.05$; (ii) $n = 10, nsim = 1000, \alpha = 0.025$; (iii) $n = 100, nsim = 1000, \alpha = 0.05$; (iv) $n = 100, nsim = 1000, \alpha = 0.025$. Compare these four coverage.

   (c) Suppose that σ^2 is unknown but can be estimated based on the sample. Repeat parts (a) and (b).

3. Let $X_1, \ldots, X_n$ be n i.i.d. random variables with mean zero and unknown variance σ^2. Suppose the prior distribution of σ^2 is inverse gamma distribution with parameters α and β, that is, $\pi(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)}\exp\{-\frac{\beta}{\sigma^2}\}$. Derive the posterior distribution of σ^2.

4. Consider the problem of generating sample from a Beta distribution $Be(\alpha, \beta)$.

   (a) One result is, if two Gamma random variables are $X_1 \sim Ga(\alpha, 1)$ and $X_2 \sim Ga(\beta, 1)$, then
   $$X = \frac{X_1}{X_1 + X_2} \sim Be(\alpha, \beta).$$
   Use this result to construct an algorithm to generate a Beta random sample. Provide a density histogram to emulate the performance.

(b) Compare the algorithm in (a) with the rejection method based on (i) the uniform distribution; (ii) the truncated normal distribution.

5. Consider estimating the integral

$$\theta = \int_0^\infty \exp(-(\sqrt{x} + 0.5x))sin^2(x)dx$$

where the pdf of x is $f(x) = \exp(-0.5x)$.

(a) Conduct the Monte Carlo (MC) integration for estimating θ.

(b) Conduct MC integration using importance sampling with the following proposal functions

$$\begin{aligned} g_1(x) &= \frac{1}{2}\exp(-|x|), \\ g_2(x) &= \frac{1}{2\pi}\frac{1}{1+x^2/4}, \\ g_3(x) &= \frac{1}{\sqrt{2\pi}}\exp(-x^2/2). \end{aligned}$$

For sample size $M = 100, 500, 1000, 2000$, compare the mean and standard deviations of the estimates.

(c) (For graduate students 862 only) Implement MC integration using self-normalized importance sampling with $g(x)$ from a mixture normal density. Explain the procedure and integrate your results clearly.