10-702 Statistical Machine Learning: Assignment 1

Due Friday, January 22

Hand in to Sharon Cavlovich, GHC (Gates Hillman Center) 8215 by 3:00. Use R for all numerical computations.

1. (Review of Maximum Likelihood.) Let $X_1, \ldots, X_n$ be a random sample where $X_i \in \{1, 2, \ldots, k\}$. Let $\theta \in [0, 1]$ and suppose that $\mathbb{P}(X_i = 1) = \theta$ and $\mathbb{P}(X_i = j) = \overline{\theta}$ for $j > 1$ where $\overline{\theta} = (1 - \theta)/(k - 1)$.

   (a) Find the mle $\widehat{\theta}$.

   (b) Find the Fisher information.

   (c) Find an approximate $1 - \alpha$ confidence interval for θ.

   (d) Find the bias and variance of $\widehat{\theta}$.

   (e) Show that $\widehat{\theta}$ is consistent.

2. (Probability.) Let $X_1, \ldots, X_n$ be iid and assume that $-1 \leq X_i \leq 1$. Also assume that X_i has mean 0.

   (a) Use Hoeffding's inequality to show that $\overline{X}_n = n^{-1} \sum_{i=1}^n X_i$ is close to 0 with high probability.

   (b) Show that there exists c such that

   $$\frac{1}{n} \sum_{i=1}^{n} X_i^2 \xrightarrow{P} c$$

   and find c.

   (c) Say whether the following statements are true or false and explain why.

      i. $\overline{X}_n = o(1)$.
      ii. $\overline{X}_n = o_P(1)$.
      iii. $\overline{X}_n = o_P(n)$.
      iv. $\overline{X}_n = o_P(1/n)$.
      v. $\overline{X}_n = O_P(n^{-1/2})$.
      vi. $\overline{X}_n = O_P(n^{-1})$.

3. This question will help you explore the differences between Bayesian and frequentist inference. Let $X_1, \ldots, X_n$ be a sample from a multivariate Normal distribution with mean $\mu = (\mu_1, \ldots, \mu_p)^T$ and covariance matrix equal to the identity matrix I. Note that each X_i is a vector of length p.

The following facts will be helpful. If $Z_1, \ldots, Z_k$ are independent $N(0,1)$ and $a_1, \ldots, a_k$ are constants, then we say that $Y = \sum_{j=1}^{p}(Z_j + a_j)^2$ has a non-central χ^2 distribution with k degrees of freedom and noncentrality parameter $||a||^2$. The mean and variance of Y are $k + ||a||^2$ and $2k + 4||a||^2$.

(a) Find the posterior under the improper prior $\pi(\mu) = 1$.

(b) Let $\theta = \sum_{j=1}^{p} \mu_j^2$. Our goal is to learn θ. Find the posterior for θ. Express your answers in terms of noncentral χ^2 distributions. Find the posterior mean $\widetilde{\theta}$.

(c) The usual frequentist estimator is $\widehat{\theta} = ||\overline{X}_n||^2 - p/n$. Show that, for any n,

$$\frac{\mathbb{E}_\mu|\theta - \widetilde{\theta}|^2}{\mathbb{E}_\mu|\theta - \widehat{\theta}|^2} \to \infty$$

as $p \to \infty$.

(d) Repeat the analysis with a $N(0, \tau^2 I)$ prior.

(e) Set $n = 10$, $p = 1000$, $\mu = (0, \ldots, 0)^T$. Simulate (in R) data N times, with $N = 1000$. Draw a histogram of the Bayes estimator (with flat prior) and the frequentist estimator. (R code for this question can be found on the web site.)

(f) Interpret your findings.

4. (Minimaxity and Bayes.) Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$. In what follows we use squared error loss.

(a) Find the mle $\widehat{p}$. Find the bias, variance and risk (mean squared error) $R(p, \widehat{p})$ of $\widehat{p}$.

(b) Recall that a Beta(α, β) density has the form

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \propto p^{\alpha-1}(1-p)^{\beta-1}.$$

Let p have a Beta(v,v) prior. Find the Bayes estimator $\overline{p}$. Find the bias, variance and risk $R(p, \overline{p})$ of $\overline{p}$.

(c) Show that $R(p, \overline{p})$ is contant (as a function of p) if v is chosen appropriately. Since $\overline{p}$ is a Bayes estimator and has constant risk, it is the minimax estimator.