10-702 Statistical Machine Learning: Assignment 1

Due Friday, January 25

Hand in to Diane Stidle, WH 4609 by 3:00. Use R for all numerical computations.

1. Let Θ be a finite set. Let $L(\theta, \widehat{\theta}) = 0$ if $\theta = \widehat{\theta}$ and $L(\theta, \widehat{\theta}) = 1$ otherwise. Show the that posterior mode is the Bayes estimator.

2. Let $X \sim N(\theta, 1)$. Suppose that $\theta \in \Theta = [-C, C]$ where $C = 1/2$. Assume squared error loss.

   (a) Verify that $\widehat{\theta} = C \tanh(CX)$ is minimax. Hint: Show that $\widehat{\theta}$ is the Bayes estimator under the prior $\pi = (1/2)\delta_{-C} + (1/2)\delta_C$ where δ_a denotes a distribution that puts probability 1 at a. You may assume that $R(\theta, \widehat{\theta})$ has the following properties: it is continuous, symmetric about 0 and increasing on $[0, c]$.

   (b) Find the mle (maximum likelihood estimator) $\widehat{\theta}$.

   (c) Find the risk of the mle.

   (d) Plot the risk functions of these two estimators.

3. Let $X \sim \text{Binomial}(n, \theta)$.

   (a) Find a minimax estimator. Hint: Consider a Bayes estimator based on a beta prior.

   (b) Plot the risk of the the minimax estimator, the mle and the Bayes estimator using a flat prior, for $n = 5, 50, 100$.

4. In class, we outlined the proof that X is minimax when $X \sim N(\theta, 1)$ and $\theta \in \mathbb{R}$. Fill in the details.

5. 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 $\widehat{\theta} = ||\overline{X}_n||^2 - p/\sqrt{n}$. Show that, for any n,

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

as $p \to \infty$.

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

(e) Set $n = 10$, $p = 1000$, $\theta = (0, \ldots, 0)^T$. Simulate (in R) data N times, with $N = 1000$. Draw a histogram of the Bayes estimator (with flat prior) and the frequentist estimator.

(e) Interpret your findings.

6. The following is a list of some loss functions commonly used for large-margin classification algorithms. For each loss function $\phi(x)$ determine whether ϕ is a convex function, then calculate and plot its conjugate ϕ^* (together with ϕ).

(a) Exponential loss: $\phi(x) = \exp(-x)$

(b) Truncated quadratic loss: $\phi(x) = [\max(1 - x, 0)]^2$

(c) Hinge loss: $\phi(x) = \max(1 - x, 0)$

(d) Sigmoid loss: $\phi(x) = 1 - \tanh(\kappa x)$, for fixed $\kappa > 0$