10-702 Statistical Machine Learning: Assignment 1

Due Friday, February 8

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

1. (One more question on minimaxity.)

   (a) Let π_n be a sequence of priors and $\widetilde{\theta}_n$ the corresponding Bayes estimators. Suppose that
   $$\int R(\theta, \widetilde{\theta}_n)\pi_n(\theta)d\theta \to c$$
   for some finite c. Suppose that $\widehat{\theta}$ is an estimator such that
   $$\sup_\theta R(\theta, \widehat{\theta}) \leq c.$$
   Show that $\widehat{\theta}$ is minimax.

   (b) Let $X \sim N(\theta, 1)$. Show that $\widehat{\theta} = X$ is minimax.
   Hint: Let π_n be $N(0, n)$. Check that
   $$\int R(\theta, \widetilde{\theta}_n)\pi_n(\theta)d\theta \to 1.$$
   Next show that $R(\theta, X) = 1$. Conclude from part (a) that X is minimax.

2. 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 its conjugate ϕ^*. Plot ϕ and 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 $\kappa > 0$

3. If $f(x, y) = f_1(x) + f_2(y)$, with f_1 and f_2 convex, show that
   $$f^*(x, y) = f_1^*(x) + f_2^*(y)$$
   Does this hold if f_1 and f_2 are not convex?

4. The following is called the *probit regression model*. Suppose $Y \in \{0, 1\}$ is a random variable given by

$$Y = \begin{cases} 1 & a^\top X + b + V \leq 0 \\ 0 & a^\top X + b + V > 0 \end{cases}$$

where $X \in \mathbf{R}^p$ is a vector of explanatory variables and $V \sim N(0, 1)$ is a latent (unobserved) random variable. Formulate the maximum likelihood estimation problem of estimating a and b, given data consisting of pairs (X_i, Y_i), $i = 1, \ldots, n$, as a convex optimization problem.

5. For $x \in \mathbb{R}^n$ define the L_p norm

$$||x||_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}$$

for $p > 0$. Let

$$C = \left\{ x : \; ||x||_p \leq 1 \right\}.$$

Show that C is convex if and only if $p \geq 1$.

6. Linear regression in R. Add brief comments to this code, and to the output, to explain what the code does and what the output means.

```
par(mfrow=c(2,2),bg="cornsilk")
n     = 100
sigma = 1
x     = rnorm(n)
x     = sort(x)
y     = 5 + 3*x + rnorm(n,0,sigma)
plot(x,y,col="blue",lwd=3)
out = lm(y ~ x)
summary(out)
abline(out,col="red",lwd=3)
abline(a=5,b=3,col="green",lwd=2)


y     = 5 + 3*x + rcauchy(n,0,sigma)
plot(x,y,col="blue",lwd=3)
out = lm(y ~ x)
summary(out)
abline(out,col="red",lwd=3)
abline(a=5,b=3,col="green",lwd=2)
```

```
nsim = 100
b = rep(0,nsim)
for(i in 1:nsim){
      x     = rnorm(n)
      x     = sort(x)
      y     = 5 + 3*x + rnorm(n,0,sigma)
      out   = lm(y ~ x)
      b[i]  = out$coef[2]
      }
summary(b)
hist(b)
abline(v=3,lwd=3,col="red")
print(mean((b-3)^2))
```

7. Prove the leave-one-out cross-validation identity:

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_{(-i)})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \widehat{Y}_i}{1 - H_{ii}}\right)^2.$$