STOR 565 Homework

1. Let D_n and D_m be independent training and test sets, respectively. Suppose that the rule $\hat{\phi}_n(x) = \phi_n(x : D_n)$ is derived from the training set.

   a. Define the test set error $\hat{R}_m(\hat{\phi}_n)$.

   b. Show that $\mathbb{E}[\hat{R}_m(\hat{\phi}_n) \,|\, D_n] = R(\hat{\phi}_n)$

   c. What is $\mathbb{E}\hat{R}_m(\hat{\phi}_n)$? Compare this to your answer above.

2. Use calculus to show that for $u \in (0, 1)$

$$\frac{u^2}{2(1-u)} \geq -\log(1-u) - u$$

3. Define what it means for a function to be strictly convex. Define the notion of a global minima. Repeat the argument from class showing that the global minima of a strictly convex function is necessarily unique.

4. Let U_1, U_2 be uncorrelated random variables with mean zero and variance one. Define $U = (U_1, U_2)^t$. Let $X = (X_1, X_2)^t$ be a random vector with components

$$X_1 \;=\; a\,U_1 + b\,U_2 \;\text{ and }\; X_2 \;=\; c\,U_1 + d\,U_2$$

   a. Find $\mathbb{E}[U]$.

   b. What is $\text{Var}(U)$?

   c. Find $\mathbb{E}X$.

   d. Find the matrix $\text{Var}(X)$ by directly calculating each entry using the definitions of X_1 and X_2.

   e. Find $\mathbf{A}$ such that $X = \mathbf{A}U$.

   f. Find $\text{Var}(X)$ using the formula for $\text{Var}(\mathbf{A}U)$.

   g. In terms of a, b, c and d, when is $\mathbf{A}$ invertible?

5. Let $h_\alpha : \mathbb{R} \to [0, \infty)$ be defined by $h_\alpha(x) = |x|^\alpha$ where $\alpha > 0$ is fixed. Sketch $h_\alpha(x)$ for $\alpha = 1/2, 1, 2$. For which values of α is $h_\alpha(x)$ convex? Justify your answer.

6. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. For $\gamma \in \mathbb{R}$ the γ-level set of f is defined to be the set of points x where $f(x)$ is less than or equal to γ. Formally,

$$L_\gamma(f) \;=\; \{x : f(x) \leq \gamma\}$$

a. Draw some level sets for the convex functions $f(x) = x^2$ and $f(x) = e^{-x}$. Note that $L_\gamma(f)$ may be empty.

b. Show that for each γ the level set $L_\gamma(f)$ is convex. Hint: If $L_\gamma(f)$ is empty then it is trivially convex. Otherwise, use the definition of a convex set.

7. Let $(x, y) \in \mathbb{R}^n \times \mathbb{R}$ be a fixed predictor-response pair, and define a function $f : \mathbb{R} \to \mathbb{R}$ by $f(\beta) = (y - x^t\beta)^2$.

a. Show that f is convex.

b. Now let $D_n = (x_1, y_1), \ldots, (x_n, y_n)$ be n predictor-response pairs. What can you say about the convexity of the sum of squares $g(\beta) = \sum_{i=1}^n (y_i - x_i^t\beta)^2$?

c. Fix $\lambda \geq 0$ and define the penalized performance criterion

$$h_\alpha(\beta) \;=\; \sum_{i=1}^n (y_i - x_i^t\beta)^2 \,+\, \lambda \sum_{j=1}^p |\beta_j|^\alpha$$

Argue that h_α is convex if $\alpha \geq 1$. Hint: Recall that a sum of convex functions is convex.

8. Consider a data set with design matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ and response vector $\mathbf{y} \in \mathbb{R}^n$. Fix $\lambda > 0$ and define the penalized loss $\hat{R}_{n,\lambda}(\beta) = ||\mathbf{y} - \mathbf{X}\beta||^2 \,+\, \lambda\, ||\beta||^2$. Following the calculus based arguments for OLS, show that $\hat{R}_{n,\lambda}(\beta)$ has unique minimizer $\hat{\beta}_\lambda = (\mathbf{X}^t\mathbf{X} + \lambda I_p)^{-1}\mathbf{X}^t y$.

9. Let $U_1, \ldots, U_m$ be random variables. Find an inequality relating $\mathbb{E}(\min_{1 \leq j \leq m} U_j)$ and $\min_{1 \leq j \leq m} \mathbb{E}U_j$. Hint: Begin by noting that $\min_{1 \leq j \leq m} U_j \leq U_k$ for each k.