STOR 565 Homework

1. Give the formal definition of a classification rule and a classification procedure. Describe the difference between them. What is the training error $\hat{R}_n(\phi)$ of a fixed rule? Find $\mathbb{E}\hat{R}_n(\phi)$ and $\text{Var}(\hat{R}_n(\phi))$.

2. 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.

3. As discussed in class, let $\hat{R}^{\text{k-CV}}(\phi)$ be the k-fold cross-validated risk of a procedure ϕ for data sets of size $(k-1)m$. Find the expected value of $\hat{R}^{\text{k-CV}}(\phi)$. One way to think of the k-fold cross-validated risk is the average of the test errors of k classification rules computed from overlapping data sets. Does this give you any insight into its expectation?

4. Let $A, B \in \mathbb{R}^{m \times n}$ be a matrices.

   a. Show that $A = B$ iff $Ax = Bx$ for all $x \in \mathbb{R}^n$.

   b. Let $v_1, \ldots, v_n$ be a basis for $\mathbb{R}^n$. Show that if $Av_i = Bv_i$ for $1 \leq i \leq n$ then $Ax = Bx$ for all $x \in \mathbb{R}^n$.

5. Let $X \geq 0$ be a random variable with $\mathbb{E}X = 10$ and $\mathbb{E}X^2 = 140$.

   a. Find an upper bound on $\mathbb{P}(X > 14)$ involving $\mathbb{E}X$ using Markov's inequality.

   b. Modify the proof of Markov's inequality to find an upper bound on $\mathbb{P}(X > 14)$ involving $\mathbb{E}X^2$.

   c. Compare the results in (a) and (b) above to what you find from Chebyshev's inequality.

6. Let X be a random variable with $\text{Var}(X) = 3$. Use Chebyshev's inequality to find upper bounds on $\mathbb{P}(|X - \mathbb{E}X| > 1)$ and $\mathbb{P}(|X - \mathbb{E}X| > 2)$. Comment on the potential usefulness of these bounds.

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

8. 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 $\alpha = 1/2, 3/4, 1, 2, 3$. For which values of α is $h_\alpha(x)$ convex? Justify your answer.

9. Show that if $f_1, \ldots, f_m : C \to \mathbb{R}$ are convex then so is $\sum_{j=1}^m f_j$.

10. 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.