STOR 565 Homework

1. Describe the difference between a fixed classification rule and a classification procedure. Define and discuss the conditional and expected risk of a classification procedure.

2. Let $f : \mathbb{R}^d \to \mathbb{R}$ be defined by $f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x}$ where $\mathbf{A} \in \mathbb{R}^{d \times d}$ is symmetric.

   a. Show that the gradient of f is given by $\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x}$.

   b. Show that the Hessian of f is given by $\nabla^2 f(\mathbf{x}) = 2\mathbf{A}$.

3. Let $a_1, \ldots, a_n$ be positive numbers. Use the Cauchy-Schwartz inequality for inner products to show that $n^2 \leq (\sum_{k=1}^n a_k)(\sum_{k=1}^n a_k^{-1})$. Hint: Begin with the identity $1 = a_k^{1/2} a_k^{-1/2}$ which holds for $k = 1, \ldots, n$.

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. (Hoeffding's MGF Bound) Let X be a discrete random variable with probability mass function $p(\cdot)$. Assume that $a \leq X \leq b$ and that $\mathbb{E}X = 0$. Let $M_X(s) = \mathbb{E}e^{sX}$ be the moment generating function of X and define $\varphi(s) := \log M_X(s)$.

   a. Show that

$$\varphi'(s) = \frac{\mathbb{E}[Xe^{sX}]}{\mathbb{E}e^{sX}} \quad \text{and} \quad \varphi''(s) = \frac{\mathbb{E}[X^2 e^{sX}]}{\mathbb{E}e^{sX}} - (\varphi'(s))^2$$

   b. Verify that $\varphi(0) = \varphi'(0) = 0$

Now fix $t > 0$ and let U be a new random variable having the "exponentially tilted" probability mass function

$$q(x) \;=\; \frac{p(x)e^{tx}}{\mathbb{E}e^{tX}}$$

   c. Verify that $q(\cdot)$ is a probability mass function, that is, $q(x) \geq 0$ and $\sum_x q(x) = 1$.

   d. Argue that $a \leq U \leq b$. This follows from the fact that U has the same possible values as X, only with different probabilities.

e. Show that $\mathbb{E}(U) = \varphi'(t)$ and that $\mathrm{Var}(U) = \varphi''(t)$.

f. Using the variance bound for bounded random variables, conclude from (c) and (d) that $\varphi''(t) \leq (b-a)^2/4$.

g. Use the second order Taylor series expansion of φ around $s = 0$ and what you've shown above to establish that $\varphi(s) \leq s^2(b-a)^2/8$ for $s > 0$.

h. Exponentiating the inequality in (g) gives Hoeffding's MGF bound.

6. Let $D_n = (X_1, Y_1), \ldots, (X_n, Y_n) \in \mathcal{X} \times \{0, 1\}$ be iid observations for a classification problem. Recall that the empirical risk of a fixed classification rule $\phi : \mathcal{X} \to \{0, 1\}$ is defined by

$$\hat{R}_n(\phi) \; = \; \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(\phi(X_i) \neq Y_i)$$

and that the risk of ϕ is $R(\phi) = \mathbb{P}(\phi(X) \neq Y)$.

a. Show that $\mathbb{E}[\hat{R}_n(\phi)] = R(\phi)$

b. Show that $\mathrm{Var}(\hat{R}_n(\phi)) = n^{-1} \, R(\phi)(1 - R(\phi)) \leq 1/(4n)$

c. Argue carefully that $n\hat{R}_n(\phi)$ has a $\mathrm{Bin}(n, R(\phi))$ distribution

d. Use Chebyshev's inequality to show that for $t \geq 0$

$$\mathbb{P}(|\hat{R}_n(\phi) - R(\phi)| \geq t) \; \leq \; \frac{R(\phi)(1 - R(\phi))}{n\, t^2} \; \leq \; \frac{1}{4\, n\, t^2}$$

e. Use Hoeffding's inequality to show that for $t \geq 0$

$$\mathbb{P}(|\hat{R}_n(\phi) - R(\phi)| \geq t) \; \leq \; 2 \exp\{-2nt^2\}$$

7. Consider a classification problem in which you have access to a test set containing $m = 120$ iid observations $(X_i, Y_i) \in \mathcal{X} \times \{0, 1\}$. You would like to use the test set to assess the risk of a given rule ϕ using the empirical risk $\hat{R}_m(\phi)$. As shown in the previous problem, Chebyshev's inequality and Hoeffding's inequality provide bounds on $\mathbb{P}(|\hat{R}_m(\phi) - R(\phi)| \geq t)$ for $t \geq 0$. Compute and compare these probability bounds, with $m = 120$, at the following values of t: $1/20$, $1/11$, $1/9$, and $1/5$.

8. Consider a classification problem in which you would like to assess the risk of a given rule ϕ using its empirical risk $\hat{R}_m(\phi)$ on a test data set D_m. In particular, you wish to determine the size m of the test set necessary to conclude that

$$\mathbb{P}(|\hat{R}_m(\phi) - R(\phi)| \geq \delta) \leq \epsilon$$

Use Chebyshev's and Hoeffding's inequalities to find suitable values for m as a function of δ and ϵ. How do the resulting quantities depend on δ and ϵ? Generally speaking, which inequality permits you to use a smaller test set?

9. Let $X_1, \ldots, X_n$ be iid Uniform$(-\theta, \theta)$ random variables.

a. Use Chebyshev's inequality to find a bound on $\mathbb{P}(\sum_{i=1}^n X_i \geq t)$ for $t \geq 0$.

b. Use Hoeffding's inequality to find a bound on $\mathbb{P}(\sum_{i=1}^n X_i \geq t)$ for $t \geq 0$.