STOR 565 Homework

1. Let $(X, Y) \in \mathbb{R}^p \times \mathbb{R}$ be a jointly distributed pair following the signal plus noise model $Y = f(X) + \varepsilon$ where ε is independent of X, $\mathbb{E}\varepsilon = 0$, and $\text{Var}(\varepsilon) = \sigma^2$.

   a. Find simple expressions for $\mathbb{E}Y$ and $\text{Var}(Y)$.

   b. Argue that $\mathbb{E}(Y|X) = f(X)$. Thus f is the regression function of Y based on X.

   c. Show that $\varphi = f$ minimizes the risk $R(\varphi) = \mathbb{E}(\varphi(X) - Y)^2$ over prediction rules $\varphi : \mathbb{R}^p \to \mathbb{R}$. What is the minimum value of $R(\varphi)$?

2. Let $(X_1, Y_1), \ldots, (X_n, Y_n) \in \mathcal{X} \times \mathbb{R}$ be iid observations from the signal plus noise model $Y = f(X) + \varepsilon$ you considered above.

   a. Define the empirical risk $\hat{R}_n(\varphi)$ of a rule $\varphi : \mathbb{R}^p \to \mathbb{R}$.

   b. Assuming that $\text{Var}(\varphi(X)) < \infty$, find the expectation and variance of $\hat{R}_n(\varphi)$.

   c. What does Chebyshev's inequality tell you in this setting?

3. Let $x_1, \ldots, x_n \in \mathbb{R}^{p+1}$ be fixed vectors with initial component equal to one 1. Suppose that we observe responses $y_1, \ldots, y_n \in \mathbb{R}$ generated from the linear model $y_i = \beta^t x_i + \varepsilon_i$, where β is an unknown coefficient vector and $\varepsilon_1, \ldots, \varepsilon_n$ are iid $\sim \mathcal{N}(0, \sigma^2)$.

   a. Argue that $y_1, \ldots, y_n$ are independent and that $y_i \sim \mathcal{N}(x_i^t \beta, \sigma^2)$.

   b. Find the joint likelihood $L(\beta)$ of $y_1, \ldots, y_n$.

   c. Find the log likelihood $\ell(\beta)$ of $y_1, \ldots, y_n$ and show that maximizing $\ell(\beta)$ over β is equivalent to minimizing the empirical risk $\hat{R}_n(\beta) = n^{-1} \sum_{i=1}^n (Y_i - X_i^t \beta)^2$ over β.

   d. Define the response vector y and design matrix X associated with the data above, giving the dimensions of each. Show carefully that $\hat{R}_n(\beta) = n^{-1} ||y - X\beta||^2$.

4. Let y and X be the response vector and design matrix, respectively, associated with observations (x_i, y_i) of the previous problem. Recall from class that the OLS coefficient $\hat{\beta} = (X^t X)^{-1} X^t y$

   a. Show that $y = X\beta + \varepsilon$ with $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I)$. Conclude that $y \sim \mathcal{N}_n(X\beta, \sigma^2 I)$.

b. Show that $\hat{\beta} = \beta + (X^t X)^{-1} X^t \varepsilon$.

c. Find $\mathbb{E}\hat{\beta}$ and $\text{Var}(\hat{\beta})$.

d. Argue that $\hat{\beta} \sim \mathcal{N}_p(\beta, \sigma^2 (X^t X)^{-1})$, and conclude that $\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2 (X^t X)^{-1}_{jj})$.

e. Use the distribution of $\hat{\beta}_j$ to find a 95% confidence interval for β_j.

5. Chi-squared distribution. A random variable X has a chi-squared distribution with $k \geq 1$ degrees of freedom, written $X \sim \chi^2_k$, if X has the same distribution as $Z_1^2 + \cdots + Z_k^2$ where $Z_1, \ldots, Z_k$ are iid $\sim \mathcal{N}(0, 1)$.

a. If $X \sim \chi^2_k$ find $\mathbb{E}X$ and $\text{Var}(X)$. You may use the fact that $\mathbb{E}Z^4 = 3$ if $Z \sim \mathcal{N}(0, 1)$.

b. If $X \sim \chi^2_k$ and $Y \sim \chi^2_l$ are independent, what is the distribution of $X + Y$?

6. Let $f_1, \ldots, f_k : \mathbb{R}^p \to \mathbb{R}$ be convex functions.

a. Show that for each number t the set $L_r = \{x : \sum_{j=1}^k f_j(x) \leq t\}$ is convex. Hint: Use results from the previous HW concerning sums and level sets of convex functions.

b. Show that for each t the sets $\{\beta \in \mathbb{R}^p : \sum_{j=1}^p \beta_j^2 \leq t\}$ and $\{\beta \in \mathbb{R}^p : \sum_{j=1}^p |\beta_j| \leq t\}$ are convex.

7. Let y and X be the response vector and design matrix, respectively, associated with observations $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^{p+1} \times \mathbb{R}$.

a. Show that $X^t X$ is symmetric and non-negative definite.

b. Find a simple relation between the eigenvalues of $X^t X + \lambda I_p$ and the eigenvalues of $X^t X$.

c. Show that $X^t X + \lambda I_p$ is invertible if $\lambda > 0$.

Now let $\hat{R}_{n,\lambda}(\beta) = ||y - X\beta||^2 + \lambda \, ||\beta||^2$ be the penalized sum of squares employed in ridge regression.

d. By following the argument used for the OLS estimator, show that if $\lambda > 0$ then $\hat{R}_{n,\lambda}(\beta)$ is strictly convex and has unique minimizer $\hat{\beta}_\lambda = (X^t X + \lambda I_p)^{-1} X^t y$.

8. Let $\hat{\beta}_\lambda$ be the minimizer of $\hat{R}_{n,\lambda}(\beta) = ||y - X\beta||^2 + \lambda\,||\beta||^2$.

   a. Show that $\hat{\beta}_0$ is the usual OLS estimator (when the rank of X is equal to p).

   b. Show that $||y - X\hat{\beta}_\lambda||^2 \leq ||y - X\beta||^2$ for every β such that $||\beta|| \leq ||\hat{\beta}_\lambda||$. Hint: Assume the stated inequality fails to hold and show that this implies that $\hat{\beta}_\lambda$ is not the minimizer of $\hat{R}_{n,\lambda}(\beta)$.