CMSE 820 HW2

This HW is due on Sep 21st at 11:59 pm.

Question 1: Prove that Norm ball defined as $\{x : \|x\| \leq r, \quad x \in \mathbb{R}^p\}$, for any given norm $\|\cdot\|$ and radius $r > 0$ is a convex set. You need to use the properties for a norm $\|\cdot\|$.

Question 2: Consider an optimization problem

$$\min f(x) \quad \text{s.t. } x \in \Omega,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex on a convex set Ω. Then the optimal solution (assuming it exists) must be unique.

Question 3: Prove that the subdifferential of function $f(x) = |x|$ at $x = 0$ is the set $[-1, 1]$ for $x \in \mathbb{R}$.

Question 4: Prove that The set $A = \{x \in \mathbb{R}^n : g(x) = 0\}$ is convex if $g(x)$ is an affine function.

Question 5: Given a response vector $\mathbf{y} \in R^n$, predictor mvatrix $\mathbf{X} \in \mathbb{R}^{p\times n}$, and tuning parameter $\lambda \geq 0$, recall the ridge regression estimate (assume we have already center $\mathbf{y}$ and $\mathbf{X}$)

$$\hat{\beta}^{\text{ridge}} = \arg\min_{\beta} \|\mathbf{y} - \mathbf{X}^T\beta\|^2 + \lambda\|\beta\|^2,$$

a. Show that $\hat{\beta}^{\text{ridge}}$ is simply the vector of ordinary regression coefficients from regressing the response $\tilde{\mathbf{y}} = (\mathbf{y}^T, 0^T)^T \in \mathbb{R}^{n+p}$ onto the predictor matrix $\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} & \sqrt{\lambda} I \end{bmatrix} \in \mathbb{R}^{(n+p)\times p}$, where $0 \in \mathbb{R}^p$, and $I \in \mathbb{R}^{p\times p}$ is the identity matrix.

b. Show that the matrix $\tilde{\mathbf{X}}$ always has full row-rank, i.e, its rows are always linearly independent, regardless of the row of $\mathbf{X}$. Hence argue that the ridge regression estimate is always unique, for any matrix of predictors $\mathbf{X}$.

c. Write out an explicit formula for $\hat{\beta}^{\text{ridge}}$ involving $\mathbf{X}, \mathbf{y}, \lambda$. Conclude that for any $a^T \in \mathbb{R}^p$, the estimate $a^T\hat{\beta}^{\text{ridge}}$ is a linear function of $\mathbf{y}$.

d. Now consider the estimation of $a^T\hat{\beta}^{\text{ridge}}$, with β^* being the true coefficient vector. Based on what weve seen in lecture, ridge regression can have a lower MSE than linear regression. But on the last homework we proved that the linear regression estimate is the BLUE. Given that it is indeed linear (from part (c)), what does this imply about the ridge regression estimate, $a^T\hat{\beta}^{\text{ridge}}$?

e. Let $\mathbf{X}$ have singular value decomposition $\mathbf{X} = UDV^T$, where $U \in \mathbb{R}^{p\times r}, D \in \mathbb{R}^{r\times r}, V \in \mathbb{R}^{n\times r}, U, V$ have orthonormal columns, and D is diagonal with elements $d_1 \geq \ldots \geq dr \geq 0$. Rewrite your formula for the ridge regression solution $\hat{\beta}^{\text{ridge}}$ from (c) by replacing $\mathbf{X}$ with UDV^T , and simplifying the expression as much as possible.

f. Assume that

$$\mathbf{y} = \mathbf{X}^T \beta^* + \epsilon, \quad \text{with } \mathbb{E}[\epsilon] = 0, \quad \mathrm{Cov}(\epsilon) = \sigma^2 I,$$

and let $a \in \mathbb{R}^p$. Prove that $a^T \hat{\beta}^{\text{ridge}}$ is indeed a biased estimate of $a^T \beta^*$, for any $\lambda > 0$.