CMSE820 Homework 2. Due 22 Sept, 2020.

I. For a vector $x = (x_1, \dots, x_p) \in \mathbb{R}^p$, its l^∞-norm is defined as

$$\|x\|_\infty := \max_k |x_k|,$$

that is, the maximal absolute value of its components. The norm can be viewed as a function $\| \bullet \|_\infty : \mathbb{R}^p \to \mathbb{R}$ whose value at x is $\|x\|_\infty$. Prove that $\| \bullet \|_\infty$ is a convex function.

Solution. Let $x, y \in \mathbb{R}^p$ be two arbitrary vectors and $t \in [0,1]$ be a real number. Consider $\|tx + (1-t)y\|_\infty$. From the definition of a norm on a vector space (i.e., the triangle inequality), $\|tx + (1-t)y\|_\infty \le \|tx\|_\infty + \|(1-t)y\|_\infty$. Additionally, for any norm, $\|\alpha v\| = |\alpha|\|v\|$ for real value α and vector v. Thus, $\|tx + (1-t)y\|_\infty \le \|tx\|_\infty + \|(1-t)y\|_\infty = t\|x\|_\infty + (1-t)\|y\|_\infty$. Since $\|tx + (1-t)y\|_\infty \le t\|x\|_\infty + (1-t)\|y\|_\infty$, it follows that the l^∞-norm is a convex function. ■

II. Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function with single variable. Let $t_1, \dots, t_n \in [0,1]$ be n numbers such that $t_1 + \cdots + t_n = 1$ and let $x_1, \dots, x_n \in \mathbb{R}$ be n real number.

(1) Consider the case $n = 3$. Prove that

$$f(t_1x_1 + t_2x_2 + t_3x_3) \le t_1 f(x_1) + (1-t_1) f\left(\frac{t_2}{t_2+t_3}x_2 + \frac{t_3}{t_2+t_3}x_3\right).$$

Solution. From the definition of $t_1, \dots, t_n$, we know that for $n = 3$, $t_1 + t_2 + t_3 = 1 \Rightarrow 1 - t_1 = t_2 + t_3$. Therefore,

$$\begin{aligned} f(t_1x_1 + t_2x_2 + t_3x_3) &= f\left(t_1x_1 + (1-t_1)\left(\frac{t_2}{1-t_1}x_2 + \frac{t_3}{1-t_1}x_3\right)\right) \\ &= f\left(t_1x_1 + (1-t_1)\left(\frac{t_2}{t_2+t_3}x_2 + \frac{t_3}{t_2+t_3}x_3\right)\right). \end{aligned}$$

Let $\tilde{x}_2 = \frac{t_2}{t_2+t_3}x_2 + \frac{t_3}{t_2+t_3}x_3$. By the convexity of f, it follows that

$$\begin{aligned} f(t_1x_1 + t_2x_2 + t_3x_3) = f(t_1x_1 + (1-t_1)\tilde{x}_2) &\le t_1 f(x_1) + (1-t_1) f(\tilde{x}_2) \\ &= t_1 f(x_1) + (1-t_1) f\left(\frac{t_2}{t_2+t_3}x_2 + \frac{t_3}{t_2+t_3}x_3\right). \blacksquare \end{aligned}$$

(2) Consider the case $n = 3$. Use the result in (1) to prove that

$$f(t_1x_1 + t_2x_2 + t_3x_3) \le t_1 f(x_1) + t_2 f(x_2) + t_3 f(x_3).$$

Solution. From (1), we know

$$f(t_1x_1 + t_2x_2 + t_3x_3) \le t_1 f(x_1) + (1-t_1) f\left(\frac{t_2}{t_2+t_3}x_2 + \frac{t_3}{t_2+t_3}x_3\right).$$

Consider now $f\left(\frac{t_2}{t_2+t_3}x_2 + \frac{t_3}{t_2+t_3}x_3\right)$. Let $\lambda = \frac{t_2}{t_2+t_3}$. From this, $1-\lambda = \frac{t_2+t_3}{t_2+t_3} - \frac{t_2}{t_2+t_3} = \frac{t_3}{t_2+t_3}$. By the convexity of f, it thus follows that

$$\begin{aligned} f(t_1x_1 + t_2x_2 + t_3x_3) &\leq t_1 f(x_1) + (1-t_1) f\left(\frac{t_2}{t_2+t_3}x_2 + \frac{t_3}{t_2+t_3}x_3\right) \\ &\leq t_1 f(x_1) + (1-t_1)\left[\frac{t_2}{t_2+t_3}f(x_2) + \frac{t_3}{t_2+t_3}f(x_3)\right]. \end{aligned}$$

Recall that $1-t_1 = t_2+t_3$. Therefore,

$$\begin{aligned} f(t_1x_1 + t_2x_2 + t_3x_3) &\leq t_1 f(x_1) + (1-t_1)\left[\frac{t_2}{t_2+t_3}f(x_2) + \frac{t_3}{t_2+t_3}f(x_3)\right] \\ &= t_1 f(x_1) + t_2 f(x_2) + t_3 f(x_3). \blacksquare \end{aligned}$$

(3) Use mathematical induction and your proofs in (1) and (2) to show the following inequality (known as Jensen's inequality) for general n:

$$f(t_1x_1 + \cdots + t_nx_n) \leq t_1 f(x_1) + \cdots + t_n f(x_n).$$

Solution. We have already proven the base case for $n=3$ in (1) and (2). The base case must be $n=3$ because for $n=2$ and the definition of $t_1 + \cdots + t_n = 1$, this would exactly recover the definition of a convex function, $f(t_1x_1 + t_2x_2) \leq t_1 f(x_1) + t_2 f(x_2)$. Suppose that $f(t_1x_1 + \cdots + t_nx_n) \leq t_1 f(x_1) + \cdots + t_n f(x_n)$ holds true. Rewriting this using summation notation, we see

$$f\left(\sum_{j=1}^{n} t_j x_j\right) \leq \sum_{j=1}^{n} t_j f(x_j)..$$

Let $k = n+1$ and t_k be defined such that $t_1 + \cdots + t_n + t_k = 1$ holds true. Thus,

$$\begin{aligned} f\left(\sum_{j=1}^{k} t_j x_j\right) &= f\left(t_k x_k + \sum_{j=1}^{n} t_j x_j\right) \\ &= f\left(t_k x_k + (1-t_k)\sum_{j=1}^{n} \frac{1}{1-t_k} x_j\right). \end{aligned}$$

Let $\tilde{x} = \sum_{j=1}^{n} \frac{1}{1-t_k} x_j$. From the definition of convexity and the proof of (1), it follows that

$$\begin{aligned} f\left(t_k x_k + (1-t_k)\tilde{x}\right) &\leq t_k f(x_k) + (1-t_k) f(\tilde{x}) \\ &= t_k f(x_k) + (1-t_k) f\left(\sum_{j=1}^{n} \frac{t_j}{1-t_k} x_j\right) \\ &= t_k f(x_k) + f\left(\sum_{j=1}^{n} t_j x_j\right). \end{aligned}$$

From the induction hypothesis and proof of (2),

$$
\begin{aligned}
f\left(\sum_{j=1}^{k} t_j x_j\right) &= t_k f(x_k) + f\left(\sum_{j=1}^{n} t_j x_j\right) \\
&\le t_k f(x_k) + \sum_{j=1}^{n} t_j f(x_j) \\
&= \sum_{j=1}^{k} t_j f(x_j).
\end{aligned}
$$

By mathematical induction, $f\left(\sum_{j=1}^{n} t_j x_j\right) \le \sum_{j=1}^{n} t_j f(x_j)$.■

III. A function f is said to be strictly convex if $D(f)$ is a convex, and for any $x, y \in D(f)$ with $x \neq y$ and for any t:

$$f[tx + (1-t)y] < tf(x) + (1-t)f(y).$$

Consider the optimization problem

$$\min_{x \in D(f)} f(x)$$

where f is strictly convex. Prove that the optimal solution (assuming it exists) is unique.

Solution. Suppose there exists two optimal solutions $x, y \in D(f)$ that minimize $f(x)$ such that $x \neq y$. It follows from the definition of minimization of a function that $\forall w \in D(f), f(x) = f(y) < f(w)$. Let $f_{min} = f(x) = f(y)$ be the minimized value of the function at x and y and let $\lambda \in [0,1]$ be a real number. Suppose $\lambda = 0.5$ Let $w = 1/2x + 1/2y$. Since $D(f)$ is a convex set, we know for any $t \in [0,1], x_1, x_2 \in D(f) \Rightarrow tx_1 + (1-t)x_2 \in D(f)$. Thus, $w \in D(f)$ for $t = \lambda$. Then,

$$f(w) = f(1/2x + 1/2y) < 1/2f(x) + 1/2f(y) = f_{min}.$$

The result $f(w) < f_{min}$ is a contradiction. Therefore, there does not exist two optimal solutions $x, y \in D(f)$ such that $x \neq y$, and thus the optimal solution must be unique. ■

IV. Let $y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{p \times n}$. Recall that ridge regression gives the following regularized solution:

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

where λ is the regularization parameter. We choose $\lambda > 0$ in this problem. Define

$$\tilde{y} := \begin{bmatrix} y \\ 0_p \end{bmatrix} \in \mathbb{R}^{n+p}, \tilde{X} := \begin{bmatrix} X & \sqrt{\lambda} I_p \end{bmatrix} \in \mathbb{R}^{p \times (n+p)}$$

where 0_p is the zero vector in $\mathbb{R}^p$ and I_p is the $p \times p$ identity matrix.

(1) Prove that $\tilde{X}\tilde{X}^T$ is invertible.

Solution. From the definition of $\tilde{X}$,

$$\tilde{X}\tilde{X}^T = \begin{bmatrix} X & \sqrt{\lambda} I_p \end{bmatrix} \begin{bmatrix} X^T \\ \sqrt{\lambda} I_p \end{bmatrix} = XX^T + \lambda I_p.$$

The matrix XX^T and λI_p are of size $p \times p$. Let $z \in \mathbb{R}^p$ be an arbitrary non-zero vector. Consider

$$\begin{aligned} z^T(XX^T + \lambda I_p)z &= z^T(XX^Tz + \lambda I_p z) \\ &= z^T XX^T z + \lambda z^T I_p z \\ &= (X^Tz)^T(X^Tz) + \lambda z^T z \\ &= \|X^Tz\|_2^2 + \lambda \|z\|_2^2. \end{aligned}$$

Since $\|X^Tz\|_2^2 + \lambda\|z\|_2^2 > 0$ for any $z \neq 0$ by the definition of the norm, it follows that $XX^T + \lambda I_p$ is positive definite. Let $M = \tilde{X}\tilde{X}^T$. Since M is positive definite, we know $z^TMz > 0 \Rightarrow Mz \neq 0$ for nonzero z. Therefore, M has full rank and $M = \tilde{X}\tilde{X}^T$ is invertible. ■

(2) Prove that the least square solution of

$$\tilde{X}^T\beta = \tilde{y}$$

is exactly β^{ridge}.

Solution. Let $f(\beta) = \|y - X^T\beta\|^2 + \lambda\|\beta\|^2$ be the function we wish to minimize over β. Finding the β that minimizes $f(\beta)$ and subsequently represents the ridge regression solution can be done by solving $\frac{\partial f}{\partial \beta} = 0$. Assume that the norm represented in $f(\beta)$ is the 2-norm. Thus,

$$\begin{aligned} \frac{\partial f}{\partial \beta} &= 0 \\ \frac{\partial}{\partial \beta}\left(\|y - X^T\beta\|^2 + \lambda\|\beta\|^2\right) &= 0 \\ \frac{\partial}{\partial \beta}\left((y - X^T\beta)^T(y - X^T\beta) + \lambda\beta^T\beta\right) &= 0 \\ \frac{\partial}{\partial \beta}\left(y^Ty - \beta^TXy - (\beta^TXy)^T + \beta^TXX^T\beta + \lambda\beta^T\beta\right) &= 0 \\ -Xy - Xy + 2XX^T\beta + 2\lambda\beta &= 0 \\ -Xy + XX^T\beta + \lambda\beta &= 0 \\ (XX^T + \lambda I)\beta &= Xy \\ \tilde{X}\tilde{X}^T\beta &= \tilde{X}\tilde{y} \\ \beta &= (\tilde{X}\tilde{X}^T)^{-1}\tilde{X}^T\tilde{y}. \end{aligned}$$

Thus, the least square solution to $\tilde{X}^T\beta = \tilde{y}$ is exactly the β that minimizes $f(\beta)$ and subsequently represents the ridge regression solution.■