CS446: Machine Learning, Fall 2017, Homework 2

Name: Lanxiao Bai (lbai5)

Worked individually

Problem 1

Solution: By definition we can write p-norm of vector $\mathbf{x}$ as

$$L_p(\mathbf{x}) = ||\mathbf{x}||_p = (\sum_{i=1}^{n} x_i^p)^{\frac{1}{p}}$$

in which n is the length of $\mathbf{x}$, so we have

$$\begin{aligned} L_p(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) &\leq L_p(t\mathbf{x}_1) + L_p((1-t)\mathbf{x}_2) \\ &= |t|L_p(\mathbf{x}_1) + |1-t|L_p(\mathbf{x}_2) \\ &= tL_p(\mathbf{x}_1) + (1-t)L_p(\mathbf{x}_2) \qquad \text{Since } t \in [0,1] \end{aligned}$$

Hence, we conclude that L_p is convex when $p > 1$.

Problem 2

Solution: Let $g(\mathbf{x}) = ||\mathbf{A}\mathbf{x} + \mathbf{b}||_2$, let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ then

$$\begin{aligned} g(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) &= f(t\mathbf{y}_1 + (1-t)\mathbf{y}_2) \\ &\leq tf(\mathbf{y}_1) + (1-t)f(\mathbf{y}_2) \\ &= tg(\mathbf{x}_1) + (1-t)g(\mathbf{x}_2) \end{aligned}$$

which givens that $||\mathbf{A}\mathbf{x} + \mathbf{b}||_2$ is convex for any $\mathbf{y} \in \mathbb{R}^n$ is convex.

Problem 3

Solution:

$$\begin{aligned} f(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) &= \sum_{i=1}^{m} \alpha_i f_i(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) \\ &\leq \sum_{i=1}^{m} \alpha_i [tf_i(\mathbf{x}_1) + (1-t)f_i(\mathbf{x}_2)] \\ &= \sum_{i=1}^{m} \alpha_i t f_i(\mathbf{x}_1) + \sum_{i=1}^{m} \alpha_i (1-t) f_i(\mathbf{x}_2) \end{aligned}$$

$$
\begin{aligned}
&= t\sum_{i=1}^{m}\alpha_i f_i(\mathbf{x}_1) + (1-t)\sum_{i=1}^{m}\alpha_i f_i(\mathbf{x}_2) \\
&= tf(\mathbf{x}_1) + (1-t)f(\mathbf{x}_2)
\end{aligned}
$$

So we see that f is convex.

Problem 4

Solution: Suppose f is strictly convex, then

$$f(tx_1 + (1-t)x_2) < tf(x_1) + (1-t)f(x_2)$$

suppose there is x_1, x_2 that makes $f(x_1) = f(x_2) = \min f$, then because it's strictly convex function, for any $t \in [0,1]$, we plug x_1, x_2 into the inequality and get

$$
\begin{aligned}
f(tx_1 + (1-t)x_2) &< tf(x_1) + (1-t)f(x_2) \\
&= tf(x_1) + (1-t)f(x_1) \\
&= f(x_1) = \min f
\end{aligned}
$$

which violates our definition of minimum. Hence, we say that the minimum of f is unique.

Problem 5

Solution:

$$
\begin{aligned}
L(\lambda_1, \lambda_2, \mathbf{w}) &= ||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2 + \lambda_1||\mathbf{w}||_1 + \lambda_2||\mathbf{w}||_2^2 \\
&= ||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2 + \lambda_1\sum_{i=1}^{n} w_i + \lambda_2\sum_{i=1}^{n} w_i^2 \\
&= ||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2 + \sum_{i=1}^{n}(\lambda_1 w_i + \lambda_2 w_i^2) \\
&= ||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2 + \frac{\lambda_1}{\sqrt{1+\lambda_2}}\sum_{i=1}^{n}(\sqrt{1+\lambda_2}w_i + \frac{\lambda_2\sqrt{1+\lambda_2}}{\lambda_1}w_i^2) \\
&= ||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2 + \frac{\lambda_1}{\sqrt{1+\lambda_2}}\sum_{i=1}^{n}(w_i^* + \frac{\lambda_2\sqrt{1+\lambda_2}}{\lambda_1}w_i^2) \\
&= ||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2 + \gamma\mathbf{w}^* + \sum_{i=1}^{n}\frac{\lambda_2}{\gamma}w_i^2
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{j=1}^{n}(y_j - \mathbf{x}_j\mathbf{w})^2 + \gamma\mathbf{w}^* + \sum_{i=1}^{n}\frac{\lambda_2}{\gamma}w_i^2 \\
&= \sum_{i=1}^{n}\left[y_i^2 - 2\mathbf{x}_i\mathbf{w}y_i + (\mathbf{x}_i\mathbf{w})^2 + \frac{\lambda_2\sqrt{1+\lambda_2}}{\lambda_1}w_i^2\right] + \gamma||\mathbf{w}^*||_1 && \text{(with 0 paddings)} \\
&= \sum_{i=1}^{n}\left[y_i^2 - 2y_i\sum_{j=1}^{n}x_{ij}w_j + (\sum_{j=1}^{n}x_{ij}w_j)^2 + \frac{\lambda_2\sqrt{1+\lambda_2}}{\lambda_1}w_i^2\right] + \gamma||\mathbf{w}^*||_1
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}(\mathbf{y} - \mathbf{x}_i\frac{1}{\sqrt{1+\lambda_2}}\sqrt{1+\lambda_2}\mathbf{w})^2 + \gamma||\mathbf{w}^*||_1 \\
&= \sum_{i=1}^{n}(\mathbf{y}^* - \mathbf{x}_i^*\mathbf{w}^*)^2 + \gamma||\mathbf{w}^*||_1 \\
&= \sum_{i=1}^{n}||\mathbf{Y}^* - \mathbf{X}^*\mathbf{w}^*||_2^2 + \gamma||\mathbf{w}^*||_1 \\
&= L(\gamma, \mathbf{w}^*)
\end{aligned}
$$

Problem 6

Solution: From Question 1, we have

$$
\begin{aligned}
\hat{\mathbf{w_{LS}}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\
&= \mathbf{X}^T\mathbf{y}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\mathbf{w}_{\text{Lasso}}} &= \arg\min_{\mathbf{w}} ||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2 + \lambda_1||\mathbf{w}||_1 \\
&= \arg\min_{\mathbf{w}}(\mathbf{y} - \mathbf{X}\mathbf{w})^T(\mathbf{y} - \mathbf{X}\mathbf{w}) + \lambda_1\sum_{i=1}^{n}w_i \\
&= \arg\min_{\mathbf{w}}(\mathbf{y}^T - (\mathbf{X}\mathbf{w})^T)(\mathbf{y} - \mathbf{X}\mathbf{w}) + \lambda_1\sum_{i=1}^{n}w_i \\
&= \arg\min_{\mathbf{w}} \mathbf{y}^T(\mathbf{y} - \mathbf{X}\mathbf{w}) - (\mathbf{X}\mathbf{w})^T(\mathbf{y} - \mathbf{X}\mathbf{w}) + \lambda_1\sum_{i=1}^{n}w_i \\
&= \arg\min_{\mathbf{w}} \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\mathbf{w} - (\mathbf{X}\mathbf{w})^T\mathbf{y} + (\mathbf{X}\mathbf{w})^T\mathbf{X}\mathbf{w} + \lambda_1\sum_{i=1}^{n}w_i \\
&= \arg\min_{\mathbf{w}} -\mathbf{y}^T\mathbf{X}\mathbf{w} - \mathbf{w}^T\mathbf{X}^T\mathbf{y} + \mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w} + \lambda_1\sum_{i=1}^{n}w_i \\
&= \arg\min_{\mathbf{w}} -\mathbf{y}^T\mathbf{X}\mathbf{w} - \mathbf{w}^T\mathbf{X}^T\mathbf{y} + \mathbf{w}^T\mathbf{w} + \lambda_1\sum_{i=1}^{n}w_i
\end{aligned}
$$

$$= \text{Prox}_{n\lambda||\cdot||_1}\hat{\mathbf{w_{LS}}}$$

In which,

$$\hat{\mathbf{w}_{\text{Lasso}}}_i = \begin{cases} \hat{\mathbf{w_{LS}}}_i - \lambda & w_i > \lambda \\ 0 & |w_i| \leq \lambda \\ \hat{\mathbf{w_{LS}}}_i + \lambda & w_i < -\lambda \end{cases}$$

Problem 7

Solution: By the result of Question 5, we have

$$\begin{aligned} \hat{\mathbf{w}}_{\text{L}} &= \arg\min_{\mathbf{w}} ||\mathbf{y} - \mathbf{X}\mathbf{w}||_2^2 + \lambda_1||\mathbf{w}||_1 + \lambda_2||\mathbf{w}||_2^2 \\ &= \arg\min_{\mathbf{w}} ||\mathbf{y}^* - \mathbf{X}^*\mathbf{w}^*||_2^2 + \gamma||\mathbf{w}^*||_1 \end{aligned}$$

So by the result of Question 6,

$$\hat{\mathbf{w}}_{\text{L}i} = \begin{cases} \hat{\mathbf{w_{LS}}}_i^* - \gamma = \sqrt{1+\lambda_2}\hat{\mathbf{w_{LS}}}_i - \frac{\lambda_1}{\sqrt{1+\lambda_2}} & w_i^* > \gamma \\ 0 & |w_i| \leq \gamma \\ \hat{\mathbf{w_{LS}}}_i^* + \gamma = \sqrt{1+\lambda_2}\hat{\mathbf{w_{LS}}}_i + \frac{\lambda_1}{\sqrt{1+\lambda_2}} & w_i^* < -\gamma \end{cases}$$