MathDNN HW 1

2018-13260 Jaehyeon Cha

1 Problem 1

(a) Write $\theta = (\theta_1, \theta_2, \cdots \theta_p)^\top$ and $X_i = (x_{1i}, x_{2i}, \cdots x_{pi})^\top$. We can rewrite $l_i(\theta)$ as

$$l_i(\theta) = \frac{1}{2}(X_i^\top \theta - Y_i)^2 = \frac{1}{2}(\sum_{k=1}^{p} x_{ki}\theta_p - Y_i)^2.$$

Computing j-th partial derivative gives

$$\frac{\partial}{\partial \theta_j} l_i(\theta) = (\sum_{k=1}^{p} x_{ki}\theta_k - Y_i)x_{ji} = (X_i^\top \theta - Y_i)x_{ji}.$$

Gathering up, we get

$$\begin{aligned}\nabla_\theta l_i(\theta) &= \left(\frac{\partial}{\partial \theta_1} l_i(\theta), \frac{\partial}{\partial \theta_2} l_i(\theta), \cdots, \frac{\partial}{\partial \theta_N} l_i(\theta)\right)^\top \\ &= (X_i^\top \theta - Y_i)(x_{1i}, x_{2i}, \cdots, x_{pi})^\top = (X_i^\top \theta - Y_i)X_i.\end{aligned}$$

(b) Since we are working in Hilbert space $\mathbb{R}^N$ with its inner product as usual dot product,

$$\|X\theta - Y\|^2 = \langle X\theta - Y, X\theta - Y\rangle = \sum_{i=1}^{N}(X_i\theta - Y_i)^2.$$

Thus, from (a) we conclude

$$\begin{aligned}\nabla_\theta \mathcal{L}(\theta) &= \sum_{i=1}^{N} \nabla_\theta \frac{1}{2}(X_i\theta - Y_i)^2 = \sum_{i=1}^{N} \nabla_\theta l_i(\theta) \\ &= \sum_{i=1}^{N} (X_i^\top \theta - Y_i)X_i \\ &= \begin{pmatrix} x_{11} & \cdots & x_{1N} \\ & \vdots & \\ x_{p1} & \cdots & x_{pN} \end{pmatrix} \begin{pmatrix} X_1\theta - Y_1 \\ \vdots \\ X_N\theta - Y_N \end{pmatrix} \\ &= X^\top (X\theta - Y).\end{aligned}$$

2 Problem 2

Since $f'(\theta) = \theta)$, $\theta^{k+1} = (1-\alpha)\theta^k$. Induction gives $\theta^{k+1} = (1-\alpha)^k\theta^0$. Now, since we assume $\alpha > 2$, $|1-\alpha| > 1$. Therefore θ^k diverges unless $\theta^0 = 0$, which is our hypothesis.

3 Problem 3

From Problem 1 we have $\nabla_\theta f(\theta) = X^\top(X\theta - Y)$, so

$$\theta^{k+1} = \theta^k - \alpha X^\top(X\theta - Y) = (I - \alpha X^\top X)\theta^k + \alpha X^\top Y,$$

where I is the identity matrix on $\mathcal{M}_{p\times p}(\mathbb{R})$. Subtracting $(X^\top X)^{-1}X^\top Y$ on both side, we get

$$\theta^{k+1} - (X^\top X)^{-1}X^\top Y = (I - \alpha X^\top X)(\theta^k - (X^\top X)^{-1}X^\top Y)).$$

Thus, if we let $\phi^k = \theta^k - (X^\top X)^{-1}X^\top Y$, $\phi^{k+1} = (1 - \alpha X^\top X)\phi^k$.
There exists an eigenvalue λ of $X^\top X$ satisfying $|\lambda| = \rho(X^\top X)$ because the number of eigenvalues of $X^\top X$ is finite. In this case $\alpha X^\top X$ has eigenvalue $\alpha\lambda$, $|\alpha\lambda| > 2$ by the condition of α. Let v be its corresponding eigenvector. We have $(I - \alpha X^\top X)v = (1-\alpha\lambda)v$, so $\phi^{k+1}v = (1-\alpha\lambda)^0\phi(v)$. Unless $\phi^0 = 0$, there is nonzero component ϕ of ϕ^0. Condition $|\alpha\lambda| > 2$ and applying Problem 2 on ϕ implies $\phi^k v$ diverges, so ϕ^k diverges, which implies θ^k diverges. Now, if we give a topology on $\mathbb{R}^N$ as product topology of $\mathbb{R}$ and its measure as usual Lebesgue measure, the set $\{\theta^0 \mid \phi^0 = 0\} = \{(X^\top X)^{-1}X^\top Y\}$ has measure zero. Therefore θ^0 diverges almost everywhere.