MFDNN HW1

Shin mingyu

March 11, 2024

Problem 1 : *Least-squares derivatives.*

(a)

$$\frac{\partial}{\partial \theta_j} l_i(\theta) = \frac{\partial}{\partial \theta_j}[\frac{1}{2}(X_{i1}\theta_1 + X_{i2}\theta_2 + \cdots + X_{ip}\theta_p - Y_i)^2] = (X_i^\intercal \theta - Y_i)X_{ij}$$

$$\Rightarrow \nabla_\theta l_i(\theta) = \begin{bmatrix} \frac{\partial}{\partial \theta_1} l_i(\theta) \\ \dots \\ \frac{\partial}{\partial \theta_p} l_i(\theta) \end{bmatrix} = \begin{bmatrix} (X_i^\intercal \theta - Y_i)X_{i1} \\ \dots \\ (X_i^\intercal \theta - Y_i)X_{ip} \end{bmatrix} = (X_i^\intercal \theta - Y_i) \cdot X_i$$

(b)

$$\mathcal{L}(\theta) = \frac{1}{2}\sum_{i=1}^{N}(X_i^\intercal \theta - Y_i)^2 = \sum_{i=1}^{N} l_i(\theta)$$

$$\nabla_\theta \mathcal{L}(\theta) = \begin{bmatrix} \frac{\partial}{\partial \theta_1} \mathcal{L}(\theta) \\ \dots \\ \frac{\partial}{\partial \theta_p} \mathcal{L}(\theta) \end{bmatrix} = \sum_{i=1}^{N} \begin{bmatrix} \frac{\partial}{\partial \theta_1} l_i(\theta) \\ \dots \\ \frac{\partial}{\partial \theta_p} l_i(\theta) \end{bmatrix} = \sum_{i=1}^{N} \nabla_\theta l_i(\theta) = \sum_{i=1}^{N} X_i(X_i^\intercal \theta - Y_i)$$

$$= X^\intercal (X\theta - Y) \quad \textit{(by the given fact, } M \cdot v = \sum_{i=1}^{N} M_{:,i} \cdot v_i\textit{)}$$

Problem 2 : *Diverging univariate GD*

$$f(\theta) = \frac{\theta^2}{2}, f'(\theta) = \theta$$

$$\Rightarrow \theta^{k+1} = \theta^k - \alpha\theta^k = (1-\alpha)\theta^k$$

$$\Rightarrow \theta^n = (1-\alpha)^n \theta^0 \rightarrow \text{diverge as n} \rightarrow \infty$$

$$\text{(since } |1-\alpha| > 1)$$

Problem 3 : *Diverging multivariate GD.*

$$
\begin{aligned}
\text{For optimal } \theta^*, \quad \nabla f'(\theta^*) = 0 \\
&\Leftrightarrow X^\intercal (X\theta^* - Y) = 0 \\
&\Leftrightarrow X^\intercal X\theta^* = X^\intercal Y \\
&\Leftrightarrow \theta^* = (X^\intercal X)^{-1} X^\intercal Y
\end{aligned}
$$

$$
\begin{aligned}
\theta^{k+1} - \theta^* &= \theta^k - \alpha X^\intercal (X\theta^k - Y) - \theta^* \\
&= \theta^k - \alpha X^\intercal X\theta^k + \alpha X^\intercal Y - \theta^* \\
&= \theta^k - \alpha X^\intercal X\theta^k + \alpha X^\intercal X\theta^* - \theta^* \\
&= (I - \alpha X^\intercal X)(\theta^k - \theta^*)
\end{aligned}
$$

Since $X^\intercal X$ is symmetric and invertible, there exists p orthornomal eigenvectors with corresponding positive eigenvalues. ($\exists (v_i, \lambda_i)_{i=1,\dots,p}$) Moreover, for each pair (v_i, λ_i),

$$
\begin{aligned}
(I - \alpha X^\intercal X)v_i &= v_i - \alpha X^\intercal X v_i \\
&= v_i - \alpha \lambda_i v_i \\
&= (1 - \alpha \lambda_i) v_i
\end{aligned}
$$

that is, $(v_i, 1 - \alpha\lambda_i)_{i=1,\dots,p}$ are eigenpair of $I - \alpha X^\intercal X$. WLOG, assume that the indices of pair are arranged in diagonal component order.
Note that there exists nonempty subset of indices J_α such that $1 - \alpha\lambda_i < -1$, for $i \in J_\alpha$ since

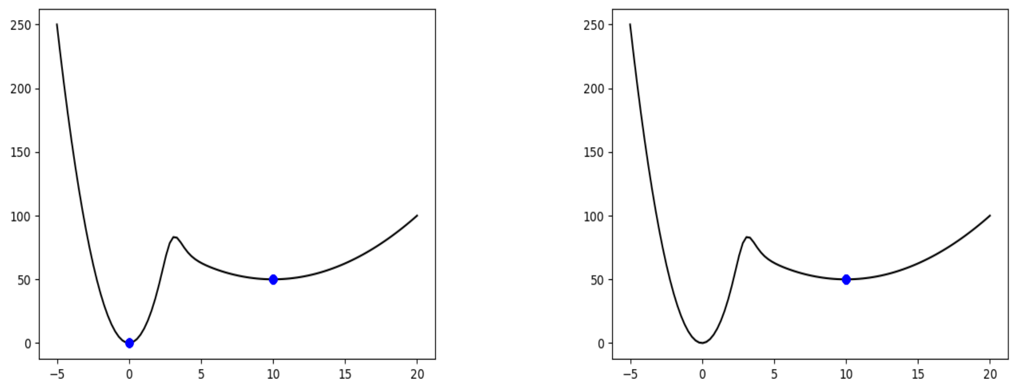
$$\alpha \rho(X^\intercal X) > 2$$

$$\to \exists\ J_\alpha \neq \varnothing \text{ s.t. } \alpha\lambda_i > 2 \quad i \in J_\alpha$$

Thus, for $\theta^0 \in \mathbb{R}^p$ except $a_i = 0 \quad \forall i \in J_\alpha$, with $\theta^0 - \theta^* = a_1 v_1 + \cdots + a_p v_p$, $\theta^n - \theta^* = (I - \alpha X^\intercal X)^n (\theta^0 - \theta^*)$ diverges since

$$
\begin{aligned}
&(I - \alpha X^\intercal X)^n (a_1 v_1 + \cdots + a_p v_p) \\
=& a_1 (1 - \alpha\lambda_1)^n v_1 + \cdots + a_p (1 - \alpha\lambda_p)^n v_p \to \infty \text{ as } n \to \infty.
\end{aligned}
$$

Problem 4 : *GD converging to wide local minima.*



(a) $\alpha = 0.01$

(b) $\alpha = 0.3$

Figure 1: 100 samples with 10000 steps

```
[3951506666864.5513, -1142421625081.9802, -545796775867.0929, 2404255799815.7935, 4392988097827.8774, 4094255043826.4023, 21
```

Figure 2: 6+ samples with only 10 steps, $\alpha = 4$

Problem 5 : *Implementing GD with duck typing.*

```
C:\Users\a0105\anaconda3\python.exe "C:\Users\a0105\iCloudDrive\2024-Spring\Mathematical Foundations of Deep Neural Networks\HW1\conv1D.py"
9.036073614024657

Process finished with exit code 0
```

Figure 3: Result