學號：B06901080　系級： 電機四　姓名：楊欣睿

請實做以下兩種不同feature的模型，回答第 (1) ~ (2) 題：

(1)抽全部9小時內的污染源feature當作一次項(加bias)

(2)抽全部9小時內pm2.5的一次項當作feature(加bias)

備註：

a. NR請皆設為0，其他的非數值(特殊字元)可以自己判斷

b. 所有 advanced 的 gradient descent 技術(如: adam, adagrad 等) 都是可以用的

c. 第1-2題請都以題目給訂的兩種model來回答

d. 同學可以先把model訓練好，kaggle死線之後便可以無限上傳。

e. 根據助教時間的公式表示，(1) 代表 p = 9x18+1 而(2) 代表 p = 9*1+1

1. (1%)記錄誤差值 (RMSE)(根據kaggle public+private分數)，討論兩種feature的影響

第一個model我train 500 個epoch後RMSE是**5.4**(kaggle 4.83487)，第二個則是**5.67**(kaggle:5.03893)。第一個model把所有資料都參考進去了，這樣子產生的function值域比較廣，但是因為training data的數量並不是趨近於無限大，所以取樣會有偏誤。因此第一個model訓練過後較容易有overfit的情形。然而第二個model因為只考慮一個feature，所以MSE會壓不下去。但是好處是他較不會overfitting。我認為第二個model考慮的feature太少了，所以她最後在kaggle的成績並不理想。

2. (1%)解釋什麼樣的data preprocessing 可以improve你的training/testing accuracy，ex. 你怎麼挑掉你覺得不適合的data points。請提供數據(RMSE)以佐證你的想法。

將有其中一個數字是超過65的都刪掉。原先的RMSE是18.02，應用後是7.1(大幅成長)

3.(3%) Refer to math problem

https://hackmd.io/RFiulFsYR5uQTrrpdxUvlw?view

1.

NO.

1.(a) $\underset{w,b}{\arg\min} \sum_{i=1}^{5} (y_i - (wx_i+b))^2$, $\because \underset{b}{\arg\min} \sum_{i=1}^{5} (z_i - b)^2 = \frac{1}{5}\sum_{i=1}^{5} z_i$ $\therefore b = \frac{1}{5}\sum_{i=1}^{5} (y_i - wx_i)$

↳ 此時 $\sum (z_i - b)^2 = \sum_{i=1}^{5} (z^2 - \bar{z}^2)$

$\therefore \underset{w}{\arg\min} \left[Var(y_i - wx_i) \right] = \underset{w}{\arg\min} \left[\sum y_i^2 - 2\sum y_i x_i w + \sum x_i^2 w^2 - \frac{1}{5}(\sum y_i - \sum x_i w)^2 \right]$

$= \underset{w}{\arg\min}\ w^2\left[\sum x_i^2 - \frac{1}{5}(\sum x_i)^2\right] - 2w\left[\sum x_i y_i - \frac{1}{5}(\sum x_i \sum y_i)\right]$

$= \dfrac{\sum x_i y_i - \frac{1}{5}(\sum x_i \sum y_i)}{\sum x_i^2 - \frac{1}{5}(\sum x_i)^2} = \dfrac{60.9 - 50.4}{55 - 45} = 1.05$ $\therefore b = \frac{1}{5}(1.05) = 0.21$.

(b) 令 $\tilde{x} = \begin{bmatrix} x \\ 1 \end{bmatrix}$, $\tilde{w} = \begin{bmatrix} w \\ b \end{bmatrix}$, $L_{ssq}(\tilde{w}) = \frac{1}{2N}\sum_{i=1}^{N} (y_i - \tilde{w}^T \tilde{x}_i)^2 = \frac{1}{2N}\|X^T\tilde{w} - y\|^2$.

$L_{sqq}(w + \Delta w) = \|X^T(w + \Delta w) - y\|^2 = L_{sqq}(w) + 2\Delta w^T X(X^T w - y) + \Delta w^T X X^T \Delta w$

$\therefore \nabla_w L_{sqq}(w) = 2X(X^T w - y)$ $\Rightarrow \nabla L_{sqq}(w) = 0 \Leftrightarrow w = (XX^T)^{-1} X y$,

Ans. $\tilde{w} = (XX^T)^{-1} X y$. where $\tilde{w} = \begin{bmatrix} w \\ b \end{bmatrix}$, and $X = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N]$ and $\tilde{x}_i = \begin{bmatrix} x_i \\ 1 \end{bmatrix}$

(c) $L_{reg}(w,b) = \frac{1}{2N}\left(\sum_{i=1}^{N} (y_i - (w^T x_i + b))^2 + N\lambda\|w\|^2\right) = \frac{1}{2N}\left(\|X^T\tilde{w} - y\|^2 + N\lambda\|\tilde{w}\|^2\right)$

$\therefore L_{reg}(w) = \frac{1}{2N}\left(\|X^T w - y\|^2 + N\lambda\|w\|^2\right)$, $\nabla_w L_{reg}(w) = 2X(X^T w - y) + 2N\lambda w$

let $\nabla_w L_{reg}(w) = 0 \Leftrightarrow [(XX^T + \lambda N I)w - Xy] = 0 \Leftrightarrow w = (XX^T + \lambda N I)^{-1} X y$

$\therefore \tilde{w} = \begin{bmatrix} w \\ b \end{bmatrix}$, and $\tilde{w} = (XX^T + \lambda N I)^{-1} X y$

2.

NO.

$$2.\quad \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N}(f(x_i+\eta_i)-y_i)^2\right] = \mathbb{E}\left[\frac{1}{2N}\|w^T(X+H)-y^T\|^2\right] \quad \text{, where } H=[\eta_1,\eta_2,\cdots,\eta_N]$$

$$= \frac{1}{2N}\mathbb{E}\left[(w^T(X+H)-y^T)((X+H)^Tw-y)\right] = \frac{1}{2N}\mathbb{E}\left[w^T(X+H)(X+H)^Tw\right]+\frac{1}{2N}\mathbb{E}\left[-2w^T(X+H)y\right]+\frac{1}{2N}\mathbb{E}\left[y^Ty\right]$$

$$= \frac{1}{2N}\left[\mathbb{E}[w^TXX^Tw]+\mathbb{E}[w^T(HX^T+XH^T)w]+\mathbb{E}[w^THH^Tw]+\mathbb{E}[-2w^TXy]+\mathbb{E}[-2w^THy]+\mathbb{E}[y^Ty]\right]$$

$$= \frac{1}{2N}\left[w^TXX^Tw + 0 + N\sigma^2 w^Tw - 2(w^TXy) + y^Ty\right]$$

$$= \frac{1}{2N}\left(\|X^Tw-y\|^2 + N\sigma^2\|w\|^2\right) = \frac{1}{2N}\sum_{i=1}^{N}(f(x_i)-y_i)^2 + \frac{\sigma^2}{2}\|w\|^2$$

3.

$$3.(a)\quad e_k = \frac{1}{N}\sum_{i=1}^{N}(g_k(x_i)-y_i)^2 = \frac{1}{N}\left(\sum_{i=1}^{N}g_k(x_i)^2 - 2\sum_{i=1}^{N}g_k(x_i)y_i + \sum_{i=1}^{N}y_i^2\right) = s_k - \frac{2}{N}\sum_{i=1}^{N}g_k(x_i)y_i + e_0$$

$$\therefore \sum_{i=1}^{N}g_k(x_i)y_i = \frac{N}{2}(s_k - e_k + e_0)$$

$$3.(b)\quad \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K}\alpha_k g_k(x_i)-y_i\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{k=0}^{K}\alpha_k(g_k(x_i)-y_i)\right]^2 \text{, where } g_0(x_i)=0,\ \alpha_0 \equiv 1-\sum_{k=1}^{K}\alpha_k$$

$$\frac{\partial L_{test}}{\partial \alpha_n} = \frac{1}{N}\sum_{i=1}^{N}\left[2\alpha_n(g_n(x_i)-y_i)^2 + \sum_{k=0}^{K}\alpha_k(g_k(x_i)-y_i)(g_n(x_i)-y_i) - \alpha_n(g_n(x_i)-y_i)^2\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}[g_n(x_i)-y_i]^2\cdot\alpha_n + \frac{1}{N}\sum_{k=0}^{K}\alpha_k\sum_{i=1}^{N}(g_k(x_i)-y_i)(g_n(x_i)-y_i)$$

$$= \alpha_n e_n + \frac{1}{N}\sum_{i=1}^{N}\left[g_n(x_i)\sum_{k=0}^{K}\alpha_k g_k(x_i) - g_n(x_i)y_i - \sum_{k=0}^{K}\alpha_k g_k(x_i)y_i + y_i^2\right]$$

$$= \alpha_n e_n - \frac{1}{2}(s_n - e_n + e_0) + e_0 + \frac{1}{N}\sum_{i=0}^{N}\left[(g_n(x_i)-y_i)\sum_{k=0}^{K}\alpha_k g_k(x_i)\right]$$

$$\frac{\partial L_{test}}{\partial \alpha_n} = 0 \iff \alpha_n = \frac{1}{2e_n}(s_n - e_n - e_0) - \frac{1}{e_n}\frac{1}{N}\sum_{i=1}^{N}\left[(g_n(x_i)-y_i)\sum_{k=0}^{K}\alpha_k g_k(x_i)\right]$$

$$= \frac{1}{2e_n}(s_n-e_n-e_0) - \frac{1}{e_n}\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K}\alpha_k g_k(x_i)g_n(x_i) + \frac{1}{e_nN}\cdot\sum_{k=0}^{K}\alpha_k\cdot\frac{N}{2}(s_k-e_k+e_0)$$

$$= \frac{1}{2e_n}(s_n-e_n-e_0) + \frac{1}{2e_n}\sum\alpha_k(s_k-e_k+e_0) - \frac{1}{e_nN}\sum_{i=1}^{N}\sum_{k=0}^{K}\alpha_k g_k(x_i)g_n(x_i)$$

$$\therefore \sum_{n=0}^{K}\alpha_n = 1 =$$ [I messed up]