學號：R08942085 系級： 電信丙 姓名：陳芃彣

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

feature(2): private score: 12.94861 public score: 12.92303

feature(1): private score: 5.58970 public score: 5.69885

Data preprocessing我只做了去除PM2.5不合理數值的筆數（PM2.5小於2以及 PM2.5大於100），再分別根據題目要求選取feature，得出的結果如上。可以見得若只單用PM2.5的前九個小時數值作為feature，成績比將所有汙染值考慮在內還差，可見影響PM2.5的因素有很多，其中一定被包含在這18個features當中，因此選取所有污染圍作為features的預測結果會較好。

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。

根據我的kaggle上傳結果“best_v3”來看，有一次的成績為：pubilc: 16.89188, private: 16.59830，後來又有一次的成績為：public: 5.69885, private: 5.58970，兩次成績之間的差別在於data preprocessing。我沒有更動features數目，然後我發現PM2.5當中有很多奇怪的數字，例如：-28或是大於100的數字，因此我將有出現奇怪PM2.5的筆數刪掉，發現成績大幅進步，可見資料的乾淨與否會大幅影響預測效果，若我們有其他污然物的相關背景知識，能夠過濾掉其他features的奇怪數值，預測效果也許會更好。

3.(3%) Refer to math problem

陳芃妙 R08942085

$S=\{(x_i, y_i)\}_{i=1}^{5} = \{(1,1.2), (2,2.4), (3,3.5), (4,4.1), (5,5.6)\}$

1-a

$$L(w,b)=\frac{1}{2\times5}\sum_{i=1}^{5}(y_i-(w^Tx_i+b))^2$$

$$\frac{\partial L}{\partial w}=\frac{1}{5}\sum_{i=1}^{5}(y_i-(w^Tx_i+b))(-x_i)=0 \quad -①$$

$$\frac{\partial L}{\partial b}=\frac{1}{5}\sum_{i=1}^{5}(y_i-(w^Tx_i+b))(-1)=0 \quad -②$$

from 1-b $\rightarrow$ $b=\bar{y}-w^T\bar{x}$

$$w^T=\frac{\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

$$\bar{y}=(1.2+2.4+3.5+4.1+5.6)\div5=3.36$$

$$\bar{x}=(1+2+3+4+5)\div5=3$$

$$w^T=\frac{(1.2-3.36)(1-3)+(2.4-3.36)(2-3)+(3.5-3.36)(3-3)+(4.1-3.36)(4-3)+(5.6-3.36)(5-3)}{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}$$

$$=\frac{1}{10}(4.32+0.96+0.74+4.48)=1.05$$

$$b=3.36-1.05\cdot3=0.21$$

$\therefore\ w^T=1.05,\ b=0.21$ #

1-b

$$\begin{cases}\dfrac{\partial L}{\partial w}=\dfrac{1}{N}\sum_{i=1}^{N}(y_i-(w^Tx_i+b))(-x_i)=0 & -①\\[2ex] \dfrac{\partial L}{\partial b}=\dfrac{1}{N}\sum_{i=1}^{N}(y_i-(w^Tx_i+b))(-1)=0 & -②\end{cases}$$

$$②:\ \sum_{i=1}^{N}(y_i-w^Tx_i-b)=0$$

$$b=\frac{1}{N}\sum_{i=1}^{N}(y_i-w^Tx_i)=\bar{y}-w^T\bar{x} \quad -③$$

$$①:\ \sum_{i=1}^{N}(y_i-(w^Tx_i+b))(x_i)=0$$

$$\sum_{i=1}^{N}(y_ix_i-w^Tx_i^2-bx_i)=0 \quad ,\ ③代入$$

$$\sum_{i=1}^{N}y_ix_i-w^T\sum_{i=1}^{N}x_i^2-\sum_{i=1}^{N}(\bar{y}-w^T\bar{x})x_i=0$$

$$\sum_{i=1}^{N}y_ix_i-w^T\sum_{i=1}^{N}x_i^2-\sum_{i=1}^{N}\bar{y}x_i+w^T\sum_{i=1}^{N}\bar{x}x_i=0$$

$$w^T\left(\sum_{i=1}^{N}\bar{x}x_i-\sum_{i=1}^{N}x_i^2\right)=\sum_{i=1}^{N}\bar{y}x_i-\sum_{i=1}^{N}y_ix_i$$

$$w^T\left(\sum_{i=1}^{N}x_i(\bar{x}-x_i)\right)=\sum_{i=1}^{N}\left(x_i(\bar{y}-y_i)\right)$$

$$\rightarrow w^T=\frac{\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \quad \#$$

$$b=\bar{y}-w^T\bar{x} \quad \#$$

1-C

$$L(w,b) = \frac{1}{2N}\sum_{i=1}^{N}(y_i - (w^Tx_i + b))^2 + \frac{\lambda}{2}\|w\|^2 \qquad \|w\|^2 = w_1^2 + \cdots + w_k^2$$

w^Tw

$$\begin{cases} \frac{\partial L}{\partial w} = \frac{1}{N}\sum_{i=1}^{N}\left[(y_i - w^Tx_i - b)(-x_i)\right] + \lambda w^T = 0 & -① \\ \frac{\partial L}{\partial b} = \frac{1}{N}\sum_{i=1}^{N}\left[(y_i - w^Tx_i - b)(-1)\right] = 0 & -② \end{cases}$$

②: $\sum_{i=1}^{N}(y_i - w^Tx_i - b) = 0$

$$b = \frac{1}{N}\sum_{i=1}^{N}(y_i - w^Tx_i) = \bar{y} - w^T\bar{x} \quad -③$$

①: $\frac{\partial L}{\partial w} = \frac{1}{N}\sum_{i=1}^{N}\left[(y_i - w^Tx_i - b)(-x_i)\right] + \lambda w^T = 0$

$$\frac{-1}{N}\left[\sum_{i=1}^{N} y_ix_i - w^T\sum_{i=1}^{N}x_i^2 - \sum_{i=1}^{N}(\bar{y} - w^T\bar{x})x_i\right] + \lambda w^T = 0 \leftarrow ③代入$$

$$\sum_{i=1}^{N}y_ix_i - w^T\sum_{i=1}^{N}x_i^2 - \sum_{i=1}^{N}\bar{y}x_i + w^T\sum_{i=1}^{N}\bar{x}x_i - N\lambda w^T = 0$$

$$w^T\left(\sum_{i=1}^{N}\bar{x}x_i - \sum_{i=1}^{N}x_i^2 - N\lambda\right) = \sum_{i=1}^{N}\bar{y}x_i - \sum_{i=1}^{N}y_ix_i$$

$$\therefore \begin{cases} w^T = \dfrac{\sum_{i=1}^{N}x_i(\bar{y} - y_i)}{\sum_{i=1}^{N}(x_i(\bar{x} - x_i) - \lambda)} \\ b = \bar{y} - w^T\bar{x} \end{cases}$$

#

2

$$f_{w,b}(x) = w^T x + b$$

$$L(w,b) = E\left[\frac{1}{2N}\sum_{\hat{n}=1}^{N}\left(f_{w,b}(x_{\hat{n}}+\eta_{\hat{n}}) - y_{\hat{n}}\right)^2\right]$$

$$E[\eta_{\hat{n},j}] = 0$$

$$E[\eta_{\hat{n},j}\,\eta_{\hat{n}',j'}] = \delta_{\hat{n}\hat{n}'}\,\delta_{jj'}\,\sigma^2$$

$$\rightarrow L(w,b) = \frac{1}{2N}\sum_{\hat{n}=1}^{N}\left(f_{w,b}(x_{\hat{n}}) - y_{\hat{n}}\right)^2 + \frac{\sigma^2}{2}\|w\|^2$$

$$\|x\|^2 = x^T x = \mathrm{trace}(x x^T)$$

$$f'_{w,b}(x) = w^T(x+\eta) + b$$

$$f' = f(x) + w^T\eta$$

$$\frac{1}{2N}\sum_{\hat{n}=1}^{N}\left(f(x) + w_{\hat{n}}^T\eta_{\hat{n}} - y_{\hat{n}}\right)^2$$

$$= \frac{1}{2N}\sum_{\hat{n}=1}^{N}\left(f(x) + w_{\hat{n}}^T\eta_{\hat{n}} - y_{\hat{n}}\right)\left(f(x) + w_{\hat{n}}^T\eta_{\hat{n}} - y_{\hat{n}}\right)$$

$$= \frac{1}{2N}\sum_{\hat{n}=1}^{N} f^2(x) + f(x)w_{\hat{n}}^T\eta_{\hat{n}} - f(x)y_{\hat{n}} + f(x)w_{\hat{n}}^T\eta_{\hat{n}} + (w_{\hat{n}}^T\eta_{\hat{n}})^2 - w_{\hat{n}}^T\eta_{\hat{n}}y_{\hat{n}} - f(x)y_{\hat{n}} - w_{\hat{n}}^T\eta_{\hat{n}}y_{\hat{n}} + y_{\hat{n}}^2$$

$$= \frac{1}{2N}\sum_{\hat{n}=1}^{N}\left[f^2(x_{\hat{n}}) + 2f(x_{\hat{n}})w_{\hat{n}}^T\eta_{\hat{n}} + (w_{\hat{n}}^T\eta_{\hat{n}})^2 - 2f(x_{\hat{n}})y_{\hat{n}} - 2w_{\hat{n}}^T\eta_{\hat{n}}y_{\hat{n}} + y_{\hat{n}}^2\right.$$

取E $\longrightarrow$ $$E\left[\frac{1}{2N}\sum_{\hat{n}=1}^{N}\left[f^2(x_{\hat{n}}) - 2f(x_{\hat{n}})y_{\hat{n}} + y_{\hat{n}}^2 + \underbrace{f(x_{\hat{n}})w_{\hat{n}}^T\eta_{\hat{n}}}_{0} - \underbrace{2w_{\hat{n}}^T\eta_{\hat{n}}y_{\hat{n}}}_{0} + (w_{\hat{n}}^T\eta_{\hat{n}})^2\right]\right]$$

$$= E\left[\frac{1}{2N}\sum_{\hat{n}=1}^{N}\left(f(x_{\hat{n}}) - y_{\hat{n}}\right)^2 + \frac{\sigma^2}{2}\|w\|^2\right]$$

#

3-a

express $\sum_{i=1}^{N} g_k(x_i) y_i$

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

$s_k = \frac{1}{N}\sum_{i=1}^{N}(g_k(x_i))^2$

$e_0 = \frac{1}{N}\sum_{i=1}^{N} y_i^2$

$$e_k = \frac{1}{N}\left[(g_k(x_1)-y_1)^2 + (g_k(x_2)-y_2)^2 + \cdots + (g_k(x_N)-y_N)^2\right]$$

$$= \frac{1}{N}\left[g_k^2(x_1) \underline{-2g_k(x_1)y_1} + y_1^2 + g_k^2(x_2) \underline{-2g_k(x_2)y_2} + y_2^2 + g_k^2(x_N) \underline{-2g_k(x_N)y_N} + y_N^2\right.$$

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

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

3-b

$$\min L\left(\sum_{k=1}^{K}\alpha_k g_k\right) = \min\left[\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K}\alpha_k g_k(x_i) - y_i\right)^2\right]$$

$$\frac{\partial L}{\partial \alpha} = \frac{2}{N}\sum_{i=1}^{N}\left[\left(\sum_{k=1}^{K}\alpha_k g_k(x_i) - y_i\right)\sum_{k=1}^{K} g_k(x_i)\right] = 0$$

$$\left(\sum_{k=1}^{K}\alpha_k g_k(x_1) - y_1\right)\sum_{k=1}^{K} g_k(x_1) + \left(\sum_{k=1}^{K}\alpha_k g_k(x_2) - y_2\right)\sum_{k=1}^{K} g_k(x_2) + \cdots + \left(\sum_{k=1}^{K}\alpha_k g_k(x_N) - y_N\right)\sum_{k=1}^{K} g_k(x_N) = 0$$

令 $G = \begin{bmatrix} g_1(x_1) & g_2(x_1) & \cdots & g_K(x_1) \\ \vdots & & & \\ g_1(x_N) & g_2(x_N) & \cdots & g_K(x_N) \end{bmatrix}_{N\times K}$ $\quad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{bmatrix}_{K\times 1}$ $\quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N\times 1}$

$$\therefore \sum_{k=1}^{K}\alpha_k g_k(x_1)\cdot\sum_{k=1}^{K} g_k(x_1) + \sum_{k=1}^{K}\alpha_k g_k(x_2)\cdot\sum_{k=1}^{K} g_k(x_2) + \cdots \sum_{k=1}^{K}\alpha_k g_k(x_N)\cdot\sum_{k=1}^{K}\alpha_k g_k(x_N) = y_1\sum_{k=1}^{K} g_k(x_1) + \cdots + y_N\sum_{k=1}^{K} g_k(x_N)$$

$$\downarrow$$

$$\begin{bmatrix} g_1(x_1) & g_1(x_2) & \cdots & g_1(x_N) \\ \vdots & & & \\ g_K(x_1) & g_K(x_2) & \cdots & g_K(x_N) \end{bmatrix}_{K\times N} \begin{bmatrix} g_1(x_1) & g_2(x_1) & \cdots & g_K(x_1) \\ \vdots & & & \\ g_1(x_N) & g_2(x_N) & \cdots & g_K(x_N) \end{bmatrix}_{N\times K} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{bmatrix}_{K\times 1} = \begin{bmatrix} g_1(x_1) & g_1(x_2) & \cdots & g_1(x_N) \\ \vdots & & & \\ g_K(x_1) & g_K(x_2) & \cdots & g_K(x_N) \end{bmatrix}_{K\times N} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N\times 1}$$

$$\Rightarrow G^T G \alpha = G^T Y$$

$$\Rightarrow \alpha = (G^T G)^{-1} G^T Y \ \#$$