學號：R07725019 系級： 資管碩二 姓名：鄒雅雯

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

(1) 抽全部 9 小時內的污染源 feature 當作一次項(加 bias)：

Kaggle public score ：6.838；private score：5.98513

(2) 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias)：

Kaggle public score ：158.44946；private score：162.31940

從結果顯示，若只使用 pm2.5 作為 feature，其表現非常不佳，因為 pm2.5 的產生還關係到其他的物質，例如 NO、pm10 等等，因此要加入其他因素一起訓練，才能有較好的表現，如(1)。

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。

除了將非數字的符號去除，並將 NAN 補為 0 外，從 pm2.5 中發現有些異常值，例如 pm2.5<0，或是非常大，因此將 pm2.5<2 或是 pm2.5>100 的異常值過濾掉。

(1) 沒有濾掉 pm2.5 異常值：

Kaggle public score ：8.05012；private score：7.83198

(2) 濾掉 pm2.5 異常值(pm2.5<2 or pm2.5>100):

Kaggle public score ：6.838；private score：5.98513

3.(3%) Refer to math problem
https://hackmd.io/RFiu1FsYR5uQTrrpdxUvlw?view

* 1-(a)

$$\frac{\partial L}{\partial w} = \frac{\frac{1}{10}\sum_{i=1}^{5}(y_i-(w^Tx_i+b))^2}{\partial w} = \frac{1}{10}\sum_{i=1}^{5} 2(y_i-(w^Tx_i+b))(-x_i)$$

$$= \frac{1}{5}\sum(-x_i)(y_i-(w^Tx_i+b)) = 0$$

$$\frac{\partial L}{\partial b} = \frac{2}{10}\sum(y_i-(w^Tx_i+b))(-1) = \frac{1}{5}\sum(-1)(y_i-(w^Tx_i+b)) = 0$$

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

$$\rightarrow (-1)(1.2-(w+b)) + (-2)(2.4-(2w+b)) + (-3)(3.5-(3w+b))$$

$$+(-4)(4.1-(4w+b)) + (-5)(5.6-(5w+b)) = 0$$

$$\rightarrow -1.2 + w + b - 4.8 + 4w + 2b - 10.5 + 9w + 3b - 16.4 + 16w + 4b$$

$$-28 + 25w + 5b = 0$$

$$\rightarrow 55w + 15b = 60.9$$

$$\sum_{i=1}^{5}(-1)(y_i-(w^Tx_i+b)) = 0$$

$$\rightarrow -1(1.2-(w+b)) - (2.4-(2w+b)) - (3.5-(3w+b)) - (4.1-(4w+b))$$

$$-(5.6-(5w+b)) = 0$$

$$\rightarrow -1.2 + w + b - 2.4 + 2w + b - 3.5 + 3w + b - 4.1 + 4w + b - 5.6 + 5w + b = 0$$

$$\rightarrow 15w + 5b = 16.8$$

$$\begin{cases} 55w + 15b = 60.9 \\ 15w + 5b = 16.8 \end{cases} \rightarrow \begin{cases} 55w + 15b = 60.9 \\ 45w + 15b = 50.4 \end{cases} \rightarrow 10w = 10.5 \rightarrow w = 1.05$$

$$15.75 + 5b = 16.8 \rightarrow 5b = 1.05 \rightarrow b = 0.21$$

Ans. $\underline{w = 1.05,\ b = 0.21}$ #

* 1-(b)

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

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

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

$$= \frac{-1}{N}\sum(y_i-(w^Tx_i+b)) = 0$$

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

$$\rightarrow \sum_{i=1}^{N}y_i - \sum_{i=1}^{N}w^Tx_i = Nb \rightarrow b = \frac{\sum_{i=1}^{N}y_i - w^T\sum_{i=1}^{N}x_i}{N} = \bar{y} - w^T\bar{x}$$

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

$$\rightarrow \sum_{i=1}^{N}y_ix_i - \sum_{i=1}^{N}w^Tx_i^2 - \sum_{i=1}^{N}bx_i = 0$$

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

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

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

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

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

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

#

* 1-(c)

$$\frac{\partial L}{\partial b} = \frac{\frac{1}{2N}\sum(y_i-(w^Tx_i+b))^2+\frac{\lambda}{2}\|w\|^2}{\partial b} = \frac{1}{2N}\sum 2(y_i-(w^Tx_i+b))(-1)=0$$

同 1-(b)

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

$$\frac{\partial L}{\partial w} = \frac{\frac{1}{2N}\sum(y_i-(w^Tx_i+b))^2+\frac{\lambda}{2}\|w\|^2}{\partial w} = \frac{1}{2N}\sum 2(y_i-(w^Tx_i+b))(-x_i)+\frac{\lambda}{2}2w=0$$

$\rightarrow \frac{-1}{N}\sum(y_i-(w^Tx_i+b))(x_i)+\lambda w=0$

$\rightarrow \sum(y_i-(w^Tx_i+b))(x_i)-\lambda wN=0$

$\rightarrow \sum y_ix_i-\sum w^Tx_i^2-\sum bx_i-\lambda wN=0$

$\rightarrow \sum y_ix_i-\sum w^Tx_i^2-\sum(\bar{y}-w^T\bar{x})x_i-\lambda wN=0$

$\rightarrow \sum y_ix_i-w^T\sum x_i^2-\sum\bar{y}x_i+\sum w^T\bar{x}x_i-\lambda wN=0$

$\rightarrow w^T\sum(\bar{x}-x_i)x_i-\lambda wN=\sum(\bar{y}-y_i)x_i$

$\rightarrow w^T(\sum(\bar{x}-x_i)x_i-\lambda N)=\sum(\bar{y}-y_i)x_i$

$$\rightarrow w=\frac{\sum(y_i-\bar{y})x_i}{\sum(x_i-\bar{x})x_i+\lambda N}$$

$$\text{Ans: } w=\frac{\sum_{i=1}^{N}(y_i-\bar{y})x_i}{\sum_{i=1}^{N}(x_i-\bar{x})x_i+\lambda N}$$

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

#

*2

By 1-(c)

$$\frac{\partial L}{\partial w} = \frac{\frac{1}{2N}\sum_{i=1}^{N}(f_{w,b}(x_i)-y_i)^2+\frac{\sigma^2}{2}\|w\|^2}{\partial w} \Rightarrow w = \frac{\sum_{i=1}^{N}(y_i-\bar{y})x_i}{\sum_{i=1}^{N}(x_i-\bar{x})x_i+\sigma^2 N} \text{ —— ①}$$

$$b = \bar{y} - w^T\bar{x} \text{ —— ②}$$

$$\frac{\partial L}{\partial b} = \frac{E[\frac{1}{2N}\sum((w^T(x_i+\eta_i)+b)-y_i)^2]}{\partial b} = E[\frac{1}{2N}\sum 2((w^T(x_i+\eta_i)+b)-y_i)] = 0$$

$$\rightarrow E[\frac{1}{N}\sum((w^T(x_i+\eta_i)+b)-y_i)] = 0 \rightarrow E[\frac{1}{N}\sum w^Tx_i+\frac{1}{N}\sum w^T\eta_i+\frac{1}{N}\sum b-\frac{1}{N}\sum y_i] = 0$$

$$\rightarrow E[w^T\frac{\sum x_i}{N}+w^T\frac{\sum\eta_i}{N}-\frac{\sum y_i}{N}+\frac{Nb}{N}] = 0 \rightarrow w^T\bar{x}+0-\bar{y}+b = 0 \rightarrow b = \bar{y}-w^T\bar{x}$$

同②

$$\frac{\partial L}{\partial w} = \frac{E[\frac{1}{2N}\sum((w^T(x_i+\eta_i)+b)-y_i)^2]}{\partial w} = E[\frac{1}{2N}\sum 2((w^T(x_i+\eta_i)+b)-y_i)(x_i+\eta_i)] = 0$$

$$\rightarrow E[\frac{1}{N}\sum w^Tx_i^2+\frac{1}{N}\sum w^Tx_i\eta_i+\frac{1}{N}\sum w^Tx_i\eta_i+\frac{1}{N}\sum w^T\eta_i^2+\frac{1}{N}\sum\bar{y}x_i+\frac{1}{N}\sum\bar{y}\eta_i-\frac{1}{N}\sum w^T\bar{x}x_i-\frac{1}{N}\sum w^T\bar{x}\eta_i = 0$$

$$\rightarrow E[\frac{w^T}{N}\sum(x_i-\bar{x})x_i+\frac{2w^T}{N}\sum x_i\eta_i+\frac{w^T}{N}\sum\eta_i^2] = E[\frac{1}{N}\sum(y_i-\bar{y})x_i+\frac{1}{N}\sum y_i\eta_i-\frac{1}{N}\sum\bar{y}\eta_i]$$

$$\rightarrow \frac{w^T}{N}\sum(x_i-\bar{x})x_i+E[\frac{w^T}{N}\sum\eta_i^2] = \frac{1}{N}\sum(y_i-\bar{y})x_i$$

$$\rightarrow \frac{w^T}{N}\sum(x_i-\bar{x})x_i+w^T\sigma^2 = \frac{1}{N}\sum(y_i-\bar{y})x_i$$

$$\rightarrow w^T(\sum(x_i-\bar{x})x_i+\sigma^2N) = \sum(y_i-\bar{y})x_i$$

$$\rightarrow w^T = \frac{\sum(y_i-\bar{y})x_i}{\sum(x_i-\bar{x})x_i+\sigma^2N} \text{ 同①}$$

$$\text{Ans: } w = \frac{\sum_{i=1}^{N}(y_i-\bar{y})x_i}{\sum_{i=1}^{N}(x_i-\bar{x})x_i+\sigma^2N}$$

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

#

* 3-(a)

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

$$\rightarrow \sum_{i=1}^{N} g_k(x_i)y_i = \frac{\sum_{i=1}^{N} g_k(x_i)^2 + \sum_{i=1}^{N} y_i^2 - Ne_k}{2} = \frac{1}{2}[NS_k + Ne_0 - Ne_k]$$

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

if $k=0$, then $\sum g_0(x_i)y_i = 0$

#

* 3-(b)

$$\alpha = [\alpha_1\ \alpha_2 \cdots \alpha_K],\ g(x)^T = [g_1(x)\ g_2(x) \cdots g_K(x)]$$

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

$$\frac{\partial L}{\partial \alpha} = \frac{\partial\, \frac{1}{N}\sum_{i=1}^{N}(\alpha g(x_i)-y_i)^2}{\partial \alpha} = \frac{1}{N}\sum_{i=1}^{N} 2(\alpha g(x_i) - y_i)g(x_i)$$

$$= \frac{2}{N}\sum_{i=1}^{N}(\alpha g(x_i)-y_i)g(x_i) = 0$$

$$\rightarrow \sum_{i=1}^{N}(\alpha g(x_i) - y_i)g(x_i) = 0 \rightarrow \sum_{i=1}^{N}\alpha g(x_i)^2 - \sum_{i=1}^{N} g(x_i)y_i = 0$$

$$\rightarrow \alpha \sum_{i=1}^{N} g(x_i)^2 = \sum_{i=1}^{N} g(x_i)y_i$$

$$\rightarrow \alpha NS = \frac{N}{2}(s + e_0 - e) \rightarrow \alpha = \frac{\frac{N}{2}(s+e_0-e)}{NS} = \frac{1}{2s}(s+e_0-e)$$

Ans : $\alpha_k = \frac{1}{2s_k}(s_k + e_0 - e_k)$

$k = 1, 2, \cdots, K$

#