學號：R08725021 系級： 資管碩一 姓名：王鼎元

請實做以下兩種不同feature的模型，回答第 (1) ~ (3) 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1-3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。

在 epoch =1000，batch size = 64，optimizer = Adam 的模型下，下表為在Kaggle 上的RMSE分數：

	Public RMSE score	Private RMSE score
Only pm 2.5	5.98643	5.88359
All feature	5.54471	5.46335

由表可知，取所有污染源當作 feature 可以得到比較適合的模型。

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。

我將所有 feature 拿出來分析，並上網參考每項 feature 的值的正常範圍為何，將遠超過正常範圍以及不合理（負數）的 training data 屏除，實驗結果如下：在 epoch =1000，batch size = 64，optimizer = Adam 的模型下，下表為在Kaggle 上的RMSE分數：

	Public RMSE score	Private RMSE score
Without preprocessing	5.71904	5.52554
With preprocessing	5.54471	5.46335

由表可知，在比較精細的 data preprocessing 後能得到比較好的模型。

3.(3%) Refer to math problem

https://hackmd.io/RFiu1FsYR5uQTrrpdxUvlw?view

1-(a)

$1\text{-}(a)\quad L_{ssq} = \frac{1}{10}\left[1.2 - (W^T + b) + 2.4 - (W^T \cdot 2 + b) + \cdots \right.$

$= \frac{1}{10}\left(16.8 - W^T \cdot 15 - 5b \right).$

$L_{ssq} = \frac{1}{10}\left[(1.2 - W^T - b)^2 + (2.4 - 2W^T - b)^2 + \cdots \right]$

$3.5 - 3W^T - b \qquad 4.1 - 4W^T - b \qquad 5.6 - 5W^T - b$

$= \frac{1}{10}\left[(W^T)^2 + b^2 + 1.44 - 2.4W^T + 2bW^T - 2.4b \right.$

$$\left(\begin{array}{l} (W^T)^2 + b^2 + 1.44 - 2.4W^T - 2.4b + 2bW^T \\ 4(W^T)^2 + b^2 + 5.76 - 9.6W^T - 4.8b + 4bW^T \\ 9(W^T)^2 + b^2 + 12.25 - 21W^T - 7b + 6bW^T \\ 16(W^T)^2 + b^2 + 16.81 - 32.8W^T - 8.2b + 8bW^T \\ 25(W^T)^2 + b^2 + 31.36 - 56W^T - 11.2b + 10bW^T \end{array} \right)$$

$= \frac{1}{10}\left(55(W^T)^2 - 121.8W^T + 5b^2 - 33.6b + 30bW^T + 67.62 \right)$

$110(W^T) - 121.8 + 30b = 0.$

$10b - 33.6 + 30W^T = 0$

$110W^T - 121.8 - 90W^T + 100.8 = 0.$

$20W^T = 21 \rightarrow W^T = \frac{21}{20} = 1.05$

$b = \frac{2.1}{10} = 0.21$

1-(b)

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

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

$$\frac{\partial L}{\partial w} = \frac{1}{N}\sum_{i=1}^{N}(y_i - w^T x_i - \bar{y} + w^T \bar{x})x_i = 0.$$

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

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

※

1-(c)

1-(c) $\frac{\partial L}{\partial b} = 0 \longrightarrow b = \bar{y} - w^T \bar{x}$

$$\frac{\partial L}{\partial w} = \frac{-1}{N} \sum_{i=1}^{N} (y_i - w^T x_i - \bar{y} + w^T \bar{x}) x_i + \lambda w = 0.$$

$$\Rightarrow \quad w = \frac{\bar{y} \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} y_i x_i}{\bar{x} \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} x_i^2 - N\lambda}$$

#

2

2. $\tilde{L} = E[\frac{1}{2N}\sum_{i=1}^{N} (w^T x_i + w^T \eta_i + b - y_i)^2]$

$= E[\frac{1}{2N}\sum_{i=1}^{N} (w^T x_i + b - y_i)^2 + (w^T \eta_i)^2 + 2(w^T x_i w^T \eta_i) + 2(w^T \eta b) - 2(w^T \eta_i y_i)]$

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

$\Rightarrow E[\sum_{i=1}^{N} (w^T \eta_i)^2] = E[\sum_{i=1}^{N} w \eta_i^T \eta_i w^T]$

$\Rightarrow E[\eta_{ij}^2] = \begin{cases} \delta^2, & \text{if } i = j \\ 0, & o/w \end{cases}$

$\Rightarrow E[\eta_{ij}^2] = Tr(\delta^2).$

$\Rightarrow \tilde{L} = \frac{1}{2N}\sum_{i=1}^{N} (w^T x_i + b - y_i)^2 + \frac{1}{2} w w^T \times Tr(\delta^2)$

$= \frac{1}{2N}\sum_{i=1}^{N} (f_{w,b}(x_i) - y_i)^2 + \frac{\delta^2}{2} \|w\|^2$

3-(a)

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

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

$$N s_k = \sum_{i=1}^{N} [g_k(x_i)]^2 , \quad N e_0 = \sum_{i=1}^{N} y_i^2$$

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

3-(b)

3-(b) $L_{test} = \frac{1}{N}\sum_{i=1}^{N}\left(\left(\sum_{k=1}^{K} a_k g_k(x_i)\right)^2 - 2\left(\sum_{k=1}^{K} a_k g_k(x_i)\right)\cdot y_i + y_i^2\right)$

$= \frac{1}{N}\sum_{i=1}^{N}\left(\left(\sum_{k=1}^{K} a_k g_k(x_i)\right)^2 - 2\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K} a_k g_k(x_i)\right)\cdot y_i + \frac{1}{N}\sum_{i=1}^{N} y_i^2\right.$

令 $\frac{\partial L_{test}}{\partial \alpha_1} = \frac{1}{N}\sum_{i=1}^{N} g_1(x_i) \times \sum_{k=1}^{K} a_k g_k(x_i) - \frac{2}{N}\sum_{i=1}^{N} g_1(x_i) y_i = 0.$

⇒ 因 $\sum_{i=1}^{N} g_1(x_i) y_i$ 已知(承上題)，故可視為K元1次方程式。

⇒ 將 L_{test} 對每個 α_k, $k=1\sim K$ 做偏微分，可得到

K個K元1次方程式，即可解出 $\alpha_1, \alpha_2 \cdots \alpha_K$