學號：B07901169 系級： 電機三 姓名：楊宗桓

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

features	9*15+1	9*1+1
train RMSE	6.6508	6.7849
validation RMSE	23.1599	17.4572
private	6.5948	6.4871
public	6.4325	6.4157

兩種 model 的 preprocess 都只有 fillna 跟令”-”=0 而已，training set 是處理過的資料的前半，validation set 的資料則是後半。

由表格可發現，兩種 model 的 RMSE 其實差不多但都偏高，除了因為沒有作其他的 data preprocess 外，還有因為 features 數量取的不好。features 取的太多會造成 model 太過複雜而產生 overfit，features 取的太少反而會造成 model 太過簡單而使得 training 出來的 function 無法貼近 target function，所以在取 features 時應該要取得剛好而不要過多或過少。

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。

我做了兩部分的 preprocess

(1) Training 中異常值的處理

這裡講的異常值包括 nan、非數值的資料還有一過高或過低不合理的數值首先是過高、過低的數值，我先用 excel 把各種 features 的平均值及標準差算出來，接著把超過平均三個標準差的數值列為異常值，如下表:

	標準差	平均	合理範圍	
SO2	1.443701766	2.22248	0	6.55358
NO	10.38180379	8.37741	0	39.5228
NOx	16.36284301	28.309	0	77.3976
NO2	9.069455914	19.9319	0	47.1402
CO	0.3534674	0.90124	0	1.96164
O3	19.0011114	23.9277	0	80.931
THC	0.324844392	2.43573	1.46119	3.41026
CH4	0.151965181	2.10464	1.64874	2.56053
NMHC	0.253195484	0.33969	0	1.09928
PM10	22.51579179	46.7976	0	114.345
PM2.5	17.03644817	27.7583	0	78.8676
WS	0.388913083	0.67708	0	1.84382
WD	108.5908908	167.778	0	493.551
AT	5.430835586	24.7789	8.48637	41.0714
RH	12.98160211	68.3878	29.443	107.333

接著將包含異常值的那一小時從 training data 中刪除(所以共會刪掉 9 筆 training 的資料)，以 model_1 為例:

train RMSE	validation RMSE	private	public
4.9626	4.9069	5.3997	5.0118

四種 RMSE 都大幅下降，可以得知判斷是否為異常值是蠻重要的 data preprocess

(2) feature extracting

由問題 1 可以知道，features 不能取的過多或過少，為了決定要取那些 feature 我決定先從 test_data1 去分析各種 feature 與 pm2.5 的相關係數，如下圖:

	相關係數
SO2	0.534719198
NO	0.171364568
NOx	0.411852209
NO2	0.544119279
CO	0.103656432
O3	0.050847356
THC	0.365519291
CH4	0.26661908
NMHC	0.315972288
PM10	0.860609886
PM2.5	1
WS	-0.178697961
WD	0.16345381
AT	-0.256999182
RH	-0.052696236

最後我選擇只取 5 個相關係數最高的 features，分別是 SO2、NO2、NOX、PM10、PM2.5

train RMSE	validation RMSE	private	public
4.8524	4.7762	5.0453	4.9306

與(1)中做過異常值分析的 RMSE 相比更低，只不過下降趨勢並沒有如(1)那麼大，所以 feature extracting 其實並沒有貢獻很大的 rmse

(3) testing data preprocess

觀察發現 testing data 中有不少 0 值，推測是觀測異常值，可能是沒有觀測數據所以才填 0，我把這些數值為 0 的改成前後兩小時的平均值，結果:

train RMSE	validation RMSE	private	public
4.8524	4.7762	4.5176	4.7457

可得知這作法可有效降低 testing 的 accuracy，因為被 preprocess 的 testing datac 會更貼近真實的觀測值。

3.(3%) Refer to math problem

https://hackmd.io/RFiu1FsYR5uQTrrpdxUv1w?view

1-(a).

$$L_{ssq}(w,x) = \frac{1}{10}\sum_{i=1}^{5}(y_i-(wx_i+b))^2$$

$$= \frac{1}{10}\left[(1.2-w-b)^2+(2.4-2w-b)^2+(3.5-3w-b)^2+(4.1-4w-b)^2+(5.6-5w-b)^2\right]$$

$$\frac{\partial L}{\partial w} = \frac{1}{10}\times 2\times\left[(w+b-1.2)+2(2w+b+2.4)+3(3w+b-3.5)+4(4w+b-4.1)+5(5w+b)-5.6\right]$$

$$= 11w+3b-12.18$$

$$\frac{\partial L}{\partial b} = \frac{1}{10}\times 2\times\left[(w+b-1.2)+(w+b-2.4)+(3w+b-3.5)+(4w+b-4.1)+(5w+b-5.6)\right]$$

$$= 3w+b-3.36$$

To match minimum

$$11w+3b-12.18 = 3w+b-3.36 = 0$$

$$\begin{cases}11w+3b=12.18\\ 3w+b=3.36\end{cases} \Rightarrow w=1.05\quad b=0.21 \qquad (w,b)=(1.05,\ 0.21)\ \#$$

1-(b).

1.(b)

$$W = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{pmatrix} \in R^k \qquad x_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ik} \end{pmatrix} \in R^k \qquad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y^N \end{pmatrix} \in R^N$$

let $w' = \begin{pmatrix} w_1 \\ \vdots \\ w_k \\ b \end{pmatrix} \in R^{k+1}$ let $X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} & 1 \\ x_{21} & x_{22} & \cdots & x_{2k} & 1 \\ & & \vdots & & 1 \\ & & & & \vdots \\ x_{N1} & x_{N2} & \cdots & & 1 \end{pmatrix} \in R^{N \times k+1}$

$$y_i = w^T x_i + b = \sum_{j=1}^{k} w_j x_{ij} + b = \sum_{j=1}^{k} w'_j x'_{ij} + b = \sum_{j=1}^{k+1} w'_j x'_{ij}$$

$$L_{ssq} = \frac{1}{2N} \sum_{i=1}^{N} \left(y_i - (w^T x_i + b)\right)^2 \qquad \frac{\partial L}{\partial w_l} = \frac{\partial}{\partial w_l}\left(\frac{1}{2N} \sum_{i=1}^{N} \left(\sum_{j=1}^{k+1} w'_j x'_{ij} - y_i \right)^2 \right)$$

$$= \frac{1}{2N} \sum_{i=1}^{N} \left(\sum_{j=1}^{k+1} w'_j x'_{ij} - y_i \right)^2 \qquad = \frac{2}{2N} \sum_{i=1}^{N} \left(x'_{il} \left(\sum_{j=1}^{k} w'_j x'_{ij} - y_j \right) \right)$$

to meet minimum

$\frac{\partial L}{\partial w_l} = 0$ for every l

$$X^T (X w' - y) = 0 \qquad w' = (X^T X)^{-1} X^T y$$

$$\Rightarrow \frac{1}{N} \sum_{i=1}^{N} \left(x'_{il} \left(\sum_{j=1}^{k+1} w'_j x'_{ij} - y_j \right) \right) = 0 \qquad \Uparrow$$

where $X = \begin{pmatrix} x_1^T & 1 \\ x_2^T & 1 \\ \vdots & \vdots \\ x_N^T & 1 \end{pmatrix}$

$$\Rightarrow \underbrace{\begin{pmatrix} x'_{11} & x'_{21} & x'_{31} & \cdots & x'_{N1} \\ x'_{12} & & & & \vdots \\ x'_{32} & & & & \vdots \\ \vdots & & & & \\ x'_{1k+1} & & \cdots & & x_{N,k+1} \end{pmatrix}}_{X^T} \left[\underbrace{\begin{pmatrix} x'_{10} & \cdots & x'_{1k} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ x'_{N0} & & x'_{Nk} \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} w'_0 \\ w'_1 \\ w'_2 \\ \vdots \\ w'_k \end{pmatrix}}_{w'} - \underbrace{\begin{pmatrix} y'_1 \\ \vdots \\ y'_N \end{pmatrix}}_{y} \right] = 0 \qquad w' = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ b \end{pmatrix}$$

1.(c).

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

$$= \frac{1}{2N}\sum_{i=1}^{N}\left(\sum_{j=1}^{k+1} w_j' x_{ij}' - y_i\right)^2 + \frac{\lambda}{2}\sum_{i=1}^{k} w_i^2 = \frac{\lambda}{2}\left(\sum_{i=1}^{k+1} w_i'^2 - b^2\right)$$

$$\frac{\partial L}{\partial w_\ell'} = \frac{2}{2N}\sum_{i=1}^{N}\left(x_{i\ell}'\left(\sum_{j=1}^{k+1} w_j' x_{ij}' - y_i\right)\right) + \lambda w_\ell$$

$$\frac{\partial \lambda}{\partial w_\ell} = \cancel{w_1^2}+\cancel{w_2^2}+\cancel{w_3^2}+w_\ell^2+\cdots = 2w_\ell$$

to meet minimum

$$\frac{\partial L}{\partial w_\ell} = 0 = \frac{1}{N}\sum_{i=1}^{N}\left(x_{i\ell}'\left(\sum_{j=1}^{k+1} w_j' x_{ij}' - y_i\right)\right) + \cdots$$

for every ℓ

$$= \frac{1}{N}\times\begin{pmatrix} x_{11}' & x_{21}' & x_{31}' & x_{41}' & \cdots & x_{N1}' \\ x_{12} & & & & & \\ \vdots & & & & & \\ x_{1k+1}' & & & & & x_{Nk+1}' \end{pmatrix}\left(\begin{pmatrix} x_{11}' & \cdots & x_{1k+1}' \\ \vdots & & \vdots \\ x_{N1}' & \cdots & x_{Nk+1}' \end{pmatrix}\begin{pmatrix} w_0' \\ \vdots \\ w_k' \end{pmatrix} - \begin{pmatrix} y_1' \\ \vdots \\ y_N' \end{pmatrix}\right) + \lambda\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{k+1} \end{pmatrix} = 0$$

$$\Rightarrow \frac{1}{N}\left(X^T(Xw'-y)\right)+\lambda w' = 0 \qquad w' = (X^TX+N\lambda)^{-1}X^Ty$$

$$(X^TX+N\lambda)w' = X^Ty$$

$$(X^TX+N\lambda)w' = X^Ty$$

2.

$$\tilde{L}_{ssq} = E\left[\frac{1}{2N}\sum_{i=1}^{N}\left(f_{w,b}(x_i+\eta_i)-y_i\right)^2\right]$$

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

$$= \frac{1}{2N}E\left[\underbrace{\sum_{i=1}^{N}(w^Tx_i+b-y_i)^2}_{\substack{\because\ \text{沒有}\ \eta_i}} + \sum_{i=1}^{N}(w^Tx_i+b-y_i)\,w^T\eta_i\times 2 + \sum_{i=1}^{N}(w^T\eta_i)^2\right]$$

expection of sum = sum of expectation

$$= \sum_{i=1}^{N}(w^Tx_i+b-y_i)^2 \qquad = \sum_{i=1}^{N}(w^Tx_i+b-y_i)\,E\left[w^T\eta_i\times 2\right]$$

$$E\left[\sum_{j=1}^{k} w_j\eta_{ij}\right] = \sum_{j=1}^{k}E[\eta_{ij}] = 0$$

$$E\left[\sum_{i=1}^{N}(w^T\eta_i)^2\right] = \sum_{i=1}\sum_{1\le j\le j'\le k} w_j w_{j'} E[\eta_{ij}\eta_{ij'}]$$

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

$$= \sum_{i=1}^{N} E\left[\left(\sum_{j=1}^{k} w_j \eta_{ij}\right)^2\right]$$

$$= \sum_{i=1}^{N} E\left[\sum_{1\le j\le j'\le k} w_j w_{j'} \eta_{ij}\eta_{ij'}\right]$$

從題目:

$$E[\eta_{ij}\eta_{i'j'}] = \delta_{ii'}\delta_{jj'}\sigma^2$$

$$E[\eta_{ij}\eta_{ij'}] = 1\times\delta_{jj'}\sigma^2$$

$j=j'$ $\quad$ $j\ne j'$

$\delta_{jj'}=0$ $\quad$ $\delta_{jj'}=0$

$$= \sum_{i=1}^{N}\sum_{1\le j\le k} w_j w_j \sigma^2 = \sum_{i=1}^{N}\|W\|^2\sigma^2 = N\|W\|^2\sigma^2$$

$$\tilde{L}_{sq} = \frac{1}{2N}\left(\sum_{i=1}^{N}(\underbrace{w^T x_i + b}_{f_{w,b}(x_i)} - y_i)^2 + 0 + N\|w\|^2\sigma^2\right)$$

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

#

3.

3.

(a) first use e_k to derive $g_k(x_i)y_i$ term

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

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

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

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

#

(b).

To solve $\min_{\alpha_1, \alpha_2 \ldots \alpha_K} L_{test}\left(\sum_{k=1}^{K} \alpha_k g_k\right)$

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

~~find $\frac{\partial L_{test}}{\partial \alpha_l} = 0$ (= minimum) for every $l = 1, 2 \ldots k$~~

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

$$= \frac{2}{N}\sum_{i=1}^{N} g_l(x_i) \sum_{k=1}^{K}\left(\alpha_k g_k(x_i) - y_i\right) = 0 \quad \text{for every } \alpha_l, \; l = 1 \ldots K$$

$$\begin{array}{ccccccc}
\begin{pmatrix} g_1(x_1) & g_1(x_2) & \cdots & g_1(x_N) \\ g_2(x_1) & & & \vdots \\ \vdots & & & \\ g_K(x_1) & \cdots & \cdots & g_K(x_N) \end{pmatrix} & \cdot & \left[\begin{pmatrix} g_1(x_1) & \cdots & g_K(x_1) \\ g_1(x_2) & \cdots & g_K(x_2) \\ \vdots & & \vdots \\ g_1(x_N) & & g_K(x_N) \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}\right] & = & \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\
K \times N & & N \times K \quad K \times 1 \quad N \times 1 & & K \times 1 \\
\Uparrow & & \Uparrow \quad\quad \Uparrow \quad\quad \Uparrow & & \\
\text{let this: } G & & G^T \quad\quad A \quad\quad Y & &
\end{array}$$

$$G(G^T A - Y) = 0$$

$$A = (GG^T)^{-1} G Y$$

$$GY = \begin{bmatrix} \sum_{i=1}^{N} g_1(x_i) y_i \\ \vdots \\ \sum_{i=1}^{N} g_K(x_i) y_i \end{bmatrix} \overset{\text{from (a)}}{=} \frac{N}{2}\begin{bmatrix} s_1 - e_1 + e_0 \\ s_2 - e_2 + e_0 \\ \vdots \\ s_K - e_K + e_0 \end{bmatrix}$$

all these datas can be obtained from known model

$$\therefore \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{bmatrix} = (GG^T)^{-1}\begin{bmatrix} s_1 - e_1 + e_0 \\ s_2 - e_2 + e_0 \\ \vdots \\ s_K - e_K + e_0 \end{bmatrix}\frac{N}{2} \quad \text{where } G = \begin{bmatrix} g_1(x_1) & g_1(x_2) & \cdots & g_1(x_N) \\ \vdots & & & \vdots \\ g_K(x_1) & & & g_K(x_N) \end{bmatrix}$$

if $(GG^T)^{-1}$ exists