學號：B08123456 系級： 資工所二 姓名：張庭逸

執行方式：

確保hw1.py, hw1.sh, train.csv, test.csv皆在同個目錄下，並執行

sh hw1.sh 檔案名

請實作以下兩種不同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。

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

表格呈現的皆為未預處理過的測試資料，將其中1000組取出作為Validation data，minibatch的epoch為2000，其餘參數照助教第一版本的code，數據如下。

Condition	Kaggle	
	Public	Private
All feature	407.59	401.95
PM 2.5	7.93	6.74

推測選用所有feature時，一是參數太多，model太過簡單，二是2000的epoch可能太少，都可能導致得到的function跟target function差異甚大。

2. (1%)解釋什麼樣的data preprocessing可以improve你的training/testing accuracy，e.g., 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。

重要的preprocessing是去除雜訊。

去除雜訊的方法為：從training data中，針對一個feature，計算所有數字出現的頻率並排序，極端值（極大或極小）中，出現次數太少或與原始資料前後數值相差太大，則判斷為outlier。針對各個feature，在outlier與正常data間取一個合理的border。

以下表格為選擇1,2,3,17項的feature，調整epoch以及調整是否有拿掉極端值。

可看出，有去除雜訊的表現是有微微比只用原始的資料表現的要好。

Condition		Kaggle	
		Public	Private
Epoch:300	Preprocess	7.59871	6.58600
	Original	7.59779	6.58770
Epoch:2000	Preprocess	7.59693	6.56804
	Original	7.58972	6.57709

3.(4%) Refer to math problem

1、

(a) $f_{w,b}(x) = P_{w,b}(C_1|x) = \sigma(-1\times 7 + 2\times 0 + -1\times 3 + 5\times 10 + 1)$

$= \sigma(41)$

$= \frac{1}{1+e^{-41}} = 1$ #

(b)

$L(w,b) = f_{w,b}(x^1)\, f_{w,b}(x^2)\,(1-f_{w,b}(x^3)) \cdots f_{w,b}(x^N)$

$-\ln L(w,b) = \ln f_{w,b}(x^1) + \ln f_{w,b}(x^2) + \ln(1-f_{w,b}(x^3)) \cdots + \ln f_{w,b}(x^N)$

$= \sum_n^N -[\hat{y}^n \ln f_{w,b}(x^n) + (1-\hat{y}^n)\ln(1-f_{w,b}(x^n))]$

$L(w,b) = e^{\sum_n [\hat{y}^n \ln f_{w,b}(x^n) + (1-\hat{y}^n)\ln(1-f_{w,b}(x^n))]}$

#

(c)

$f_{w,b}(x) = \sigma(z) = \frac{1}{1+e^{-z}} \qquad z = w\cdot x + b$

$\frac{-\ln L(w,b)}{\partial w_i} = \sum_n -\left[\hat{y}^n \frac{\ln f_{w,b}(x^n)}{\partial w_i} + (1-\hat{y}^n)\frac{\ln(1-f_{w,b}(x^n))}{\partial w_i}\right]$

其中 $\frac{\partial \ln f_{w,b}(x)}{\partial w_i} = \frac{\partial \ln \sigma(z)}{\partial z}\cdot\frac{\partial z}{\partial w_i}$

$= \frac{1}{\sigma(z)}\frac{\partial \sigma(z)}{\partial z}\cdot\frac{\partial \sum w_i x_i + b}{\partial w_i}$

$= \frac{1}{\cancel{\sigma(z)}}\cdot\cancel{\sigma(z)}(1-\sigma(z))\cdot x_i$

$= (1-\sigma(z))\,x_i$

$\frac{\partial \ln(1-f_{w,b}(x))}{\partial w_i} = \frac{\partial \ln(1-\sigma(z))}{\partial z}\cdot\frac{\partial z}{\partial w_i}$

$= \frac{1}{1-\sigma(z)}\cdot\sigma(z)(1-\sigma(z))\cdot x_i$

$= \sigma(z)\cdot x_i$

$= \sum_n -[\hat{y}^n(1-f_{w,b}(x^n))x_i^n - (1-\hat{y}^n)f_{w,b}(x^n)x_i^n]$

$= \sum_n -(\hat{y}^n - f_{w,b}(x^n))\,x_i^n$

$w_i \leftarrow w_i - \eta\sum_n -(\hat{y}^n - f_{w,b}(x^n))\,x_i^n$ (η = learning rate) #

P:1

2.

(a) $L_{ssq}(w,b) = \frac{1}{10}\sum_{i=1}^{5}\left(y_i - (w^T x_i + b)\right)^2$

$= \frac{1}{10}\left[(1.5-w-b)^2 + (2.4-2w-b)^2 + (3.5-3w-b)^2 + (4.1-4w-b)^2 + (5.3-5w-b)^2\right]$

$\frac{\partial L_{ssq}(w,b)}{\partial w} = \frac{1}{10}\left[2(1.5-w-b) + 2(2.4-2w-b)\cdot 2 + 2(3.5-3w-b)\cdot 3 + 2(4.1-4w-b)\cdot 4 + 2(5.3-5w-b)\cdot 5\right]$

$= 10.94 - 11w - 3b$

$\frac{\partial L_{ssq}(w,b)}{\partial b} = \frac{1}{10}\left[2(1.5-w-b) + 2(2.4-2w-b) + 2(3.5-3w-b) + 2(4.1-4w-b) + 2(5.3-5w-b)\right]$

$= 3.36 - 3w - b$

令 $\frac{\partial L_{ssq}(w,b)}{\partial w} = 10.94 - 11w - 3b = 0$

$\frac{\partial L_{ssq}(w,b)}{\partial b} = 3.36 - 3w - b = 0$

$\Rightarrow$ $w = 0.43$, $b = 2.07$

$(w,b) = (0.43, 2.07)$ #

(b) 令 $w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_K \end{pmatrix} \in R^K$, $w' = \begin{pmatrix} w_0' \\ w_1' \\ \vdots \\ w_K' \end{pmatrix} = \begin{pmatrix} b \\ w_1 \\ w_2 \\ \vdots \\ w_K \end{pmatrix} \in R^{K+1}$

$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{pmatrix} \in R^K$, $x_i' = \begin{pmatrix} x_{i0}' \\ x_{i1}' \\ \vdots \\ x_{iK}' \end{pmatrix} = \begin{pmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{pmatrix} \in R^{K+1}$

$X = \begin{pmatrix} x_1'^T \\ x_2'^T \\ \vdots \\ x_N'^T \end{pmatrix} = \begin{pmatrix} x_{10}' & x_{11}' & \cdots & x_{1K}' \\ x_{20}' & & & \vdots \\ \vdots & & & \vdots \\ x_{N0}' & \cdots & \cdots & x_{NK}' \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1K} \\ 1 & \vdots & & \vdots \\ \vdots & & & \vdots \\ 1 & x_{N1} & \cdots & x_{NK} \end{pmatrix} \in R^{N\times(K+1)}$

$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \in R^N$

$L_{ssq}(w,b) = \frac{1}{2N}\sum_{i=1}^{N}\left(y_i - (\underbrace{w^T x_i + b}_{\downarrow})\right)^2$

$\sum_{j=1}^{K} w_j x_{ij} + b = \sum_{j=1}^{K} w_j' x_{ij}' + w_0' x_{i0}' = \sum_{j=0}^{K} w_j' x_{ij}'$

No:

Date: /

$$L_{sq}(w,b) = \frac{1}{2N}\sum_{i=1}^{N}\left(y_i - \sum_{j=0}^{K} w_j' x_{ij}'\right)^2 = \frac{1}{2N}\sum_{i=1}^{N}\left(\sum_{j=0}^{K} w_j' x_{ij}' - y_i\right)^2$$

$$\frac{\partial L_{ssq}(w,b)}{\partial w_n'} = \frac{1}{2N}\sum_{i=1}^{N}\left[2\left(\sum_{j=0}^{K} w_j' x_{ij}' - y_i\right) x_{in}'\right] = \frac{1}{N}\sum_{i=1}^{N} x_{in}'\left(\sum_{j=0}^{K} w_j' x_{ij}' - y_i\right)$$

$$\frac{\partial L_{ssq}(w,b)}{\partial w} = \begin{bmatrix} \frac{\partial L_{ssq}(w,b)}{\partial w_0'} \\ \frac{\partial L_{ssq}(w,b)}{\partial w_1'} \\ \vdots \\ \frac{\partial L_{ssq}(w,b)}{\partial w_K'} \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} x_{i0}'\left(\sum_{j=0}^{K} w_j' x_j' - y_i\right) \\ \frac{1}{N}\sum_{i=1}^{N} x_{i1}'\left(\sum_{j=0}^{K} w_j' x_j' - y_i\right) \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} x_{iK}'\left(\sum_{j=0}^{K} w_j' x_j' - y_i\right) \end{bmatrix} = 0$$

$$\frac{1}{N}\begin{bmatrix} x_{10}' & x_{20}' & \cdots & x_{N0}' \\ x_{11}' & & & \vdots \\ \vdots & & & \vdots \\ x_{1K}' & \cdots & \cdots & x_{NK}' \end{bmatrix} \cdot \left(\begin{bmatrix} x_{10}' & x_{11}' & \cdots & x_{1K}' \\ x_{20}' & & & \vdots \\ \vdots & & & \vdots \\ x_{N0}' & \cdots & \cdots & x_{NK}' \end{bmatrix} \begin{bmatrix} w_0' \\ w_1' \\ \vdots \\ w_K' \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \right) = 0$$

$$X^T \cdot (X \quad w' - y) = 0$$

$$X^T X w' = X^T y$$

$$\begin{pmatrix} b \\ w_1 \\ w_2 \\ \vdots \\ w_K \end{pmatrix} = w' = X^T y = \begin{pmatrix} 1 & x_1^T \\ 1 & x_2^T \\ \vdots & \vdots \\ 1 & x_N^T \end{pmatrix}^T y$$ 為 $L_{sq}(w,b)$ minimize 的一組 (w,b)

(c) 沿用自 (b)，已知 $w^T x_i + b = \sum_{j=0}^{K} w_j' x_{ij}'$

$$\|w\|^2 = \sum_{i=1}^{K} w_i^2 = \sum_{j}^{K} w_i'^2 - b^2$$

$$L_{reg}(w,b) = \frac{1}{2N}\sum_{i=1}^{K}\left(y_i - (w^T x_i - b)\right)^2 + \frac{\lambda}{2}\|w\|^2$$
$$= \frac{1}{2N}\sum_{i=1}^{K}\left(y_i - \sum_{j=0}^{K} w_j' x_{ij}'\right)^2 + \frac{\lambda}{2}\left(\sum_{i=0}^{K} w_i'^2 - b^2\right)$$
$$= \frac{1}{2N}\sum_{i=1}^{K}\left(\sum_{j=0}^{K} w_j' x_{ij}' - y_i\right)^2 + \frac{\lambda}{2}\left(\sum_{i=0}^{K} w_i'^2 - b^2\right)$$

No:

Date: /

$$\frac{\partial}{\partial w_n'} L_{reg}(w,b) = \frac{1}{2N}\sum_{i=1}^{K}\left(2\left(\sum_{j=0}^{K} w_j' x_{ij}' - y_i\right)\cdot x_{in}'\right) + \frac{\lambda}{2}\cdot 2w_n'$$

$$= \frac{1}{N}\sum_{i=1}^{K}\left(x_{in}'\left(\sum_{j=0}^{K} w_j' x_{ij}' - y_i\right)\right) + \lambda w_n'$$

$$\frac{\partial}{\partial w} L_{reg}(w,b) = \begin{bmatrix} \frac{\partial L_{reg}(w,b)}{\partial w_0'} \\ \vdots \\ \frac{\partial L_{reg}(w,b)}{\partial w_K'} \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\sum_{i=0}^{K}\left(x_{i0}'\left(\sum_{j=0}^{K} w_j' x_{ij}' - y_i\right)\right) + \lambda w_0' \\ \vdots \\ \frac{1}{N}\sum_{i=0}^{K}\left(x_{iK}'\left(\sum_{j=0}^{K} w_j' x_{ij}' - y_i\right)\right) + \lambda w_K' \end{bmatrix} = 0$$

$$\frac{1}{N}\begin{bmatrix} x_{10}' & \cdots & x_{N0}' \\ x_{11}' & & \vdots \\ \vdots & & \\ x_{1K}' & \cdots & x_{NK}' \end{bmatrix}\cdot\left(\begin{bmatrix} x_{10}' & \cdots & x_{1K}' \\ \vdots & & \vdots \\ x_{N0}' & \cdots & x_{NK}' \end{bmatrix}\begin{bmatrix} w_0' \\ w_1' \\ \vdots \\ w_K' \end{bmatrix} - \begin{bmatrix} y_1' \\ \vdots \\ y_N' \end{bmatrix}\right) + \lambda\begin{bmatrix} w_0' \\ \vdots \\ w_K' \end{bmatrix} = 0$$

$$\frac{1}{N} \quad X^T \quad (\quad X \quad w' - y) + \lambda w' = 0$$

$$(X^T X + N\lambda I_{K+1})\, w' = X^T y$$

$$\begin{bmatrix} b \\ w_1 \\ w_2 \\ \vdots \\ w_K \end{bmatrix} = w' = (X^T X + N\lambda I_{K+1})^{-1} X^T y$$

$$= \left(\begin{bmatrix} 1 & x_1^T \\ 1 & x_2^T \\ \vdots & \vdots \\ 1 & x_N^T \end{bmatrix}^T \begin{bmatrix} 1 & x_1^T \\ 1 & x_2^T \\ \vdots & \vdots \\ 1 & x_N^T \end{bmatrix} + N\lambda I_{K+1}\right)^{-1} \begin{bmatrix} 1 & x_1^T \\ 1 & x_2^T \\ \vdots & \vdots \\ 1 & x_N^T \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

為 $L_{reg}(w,b)$ minimize 的一組 (w,b)

No:

Date: /

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

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

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

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

$$= \frac{1}{2N}\left(\sum_{i=1}^{N}\left(f_{w,b}(x_i)-y_i\right)^2+\sum_{i=1}^{N}2f_{w,b}(x_i)\underbrace{E\left[w^T\eta_i\right]}_{=0}+E\left[\sum_{i=1}^{N}\left(w^T\eta_i\right)^2\right]\right)$$

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

$$= \sum_{i=1}^{N}\left(\sum_{1\le j,j'\le K}w_j w_{j'}E\left[\eta_{ij}\eta_{ij'}\right]\right) = \sum_{i=1}^{N}\left(\sum_{1\le j,j'\le K}w_j w_{j'}\cdot\delta_{ii}\delta_{jj'}\sigma^2\right)$$

$$= \sum_{i=1}^{N}\left(\sum_{1\le j=j'\le K}w_j w_{j'}\cdot\delta_{ii}\delta_{jj'}\sigma^2\right)+\sum_{i=1}^{N}\left(\sum_{1\le j\ne j'\le K}w_j w_{j'}\delta_{ii}\underbrace{\delta_{jj'}}_{=0}\sigma^2\right)$$

$$= \sum_{i=1}^{N}\left(\sum_{j=1}^{K}w_j\cdot w_j\cdot 1\cdot 1\cdot\sigma^2\right) = \sum_{i=1}^{N}\sigma^2\sum_{j=1}^{K}w_j^2 = N\sigma^2\|w\|^2$$

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

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

No:

Date: /

4.

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

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

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

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

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

(b)

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

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

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

$$\frac{\partial}{\partial\alpha}L_{test} = \begin{bmatrix}\frac{\partial}{\partial\alpha_1}L_{test}\\ \vdots\\ \frac{\partial}{\partial\alpha_K}L_{test}\end{bmatrix} = \begin{bmatrix}\frac{2}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K}\alpha_k g_k(x_i) - y_i\right)\cdot g_1(x_i)\\ \vdots\\ \frac{2}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K}\alpha_k g_k(x_i) - y_i\right)\cdot g_K(x_i)\end{bmatrix} = 0$$

$$\frac{2}{N}\underbrace{\begin{bmatrix}g_1(x_1) & g_1(x_2) & \cdots & g_1(x_N)\\ g_2(x_1) & & & \vdots\\ \vdots & & & \\ g_K(x_1) & \cdots & & g_K(x_N)\end{bmatrix}}_{Z^T}\left(\underbrace{\begin{bmatrix}g_1(x_1) & \cdots & g_K(x_1)\\ \vdots & & \vdots\\ g_1(x_N) & \cdots & g_K(x_N)\end{bmatrix}}_{Z}\underbrace{\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_K\end{bmatrix}}_{\alpha} - \underbrace{\begin{bmatrix}y_1\\ \vdots\\ y_N\end{bmatrix}}_{y}\right) = 0$$

令

$$\frac{2}{N}Z^T(Z\alpha - y) = 0$$

$$Z^TZ\alpha = Z^Ty = \begin{bmatrix}g_1(x_1) & \cdots & g_1(x_N)\\ \vdots & & \vdots\\ g_K(x_1) & \cdots & g_K(x_N)\end{bmatrix}\begin{bmatrix}y_1\\ \vdots\\ y_N\end{bmatrix} = \begin{bmatrix}\sum_{i=1}^{N}g_1(x_i)y_i\\ \vdots\\ \sum_{i=1}^{N}g_K(x_i)y_i\end{bmatrix} = \begin{bmatrix}\frac{N}{2}(s_1+e_0-e_1)\\ \vdots\\ \frac{N}{2}(s_K+e_0-e_K)\end{bmatrix}$$

$$\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_K\end{bmatrix} = \alpha = (Z^TZ)^{-1}Z^Ty = \left(\begin{bmatrix}g_1(x_1) & \cdots & g_1(x_N)\\ \vdots & & \vdots\\ g_K(x_1) & \cdots & g_K(x_N)\end{bmatrix}\begin{bmatrix}g_1(x_1) & \cdots & g_K(x_1)\\ \vdots & & \vdots\\ g_1(x_N) & \cdots & g_K(x_N)\end{bmatrix}\right)^{-1}\begin{bmatrix}\frac{N}{2}(s_1+e_0-e_1)\\ \vdots\\ \frac{N}{2}(s_K+e_0-e_K)\end{bmatrix}$$ #