學號：B06507002 系級： 材料三 姓名：林柏勳

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

model \score	public	private
(1) all features	5.83085	5.45078
(2) PM2.5	5.96996	5.77371

所有 feature 加在一起的 model 會比單純用 PM2.5 這個 feature 建立的 model 還要來的好，原因在於除了 PM2.5 這個主要變因之外，還有很多其他因素會嚴重影響 PM2.5 的預測值(觀測值的重要程度，可以用 batch normalization 後的 weight 值大小看出)，例如: CO、NO、NO_2、O_3。

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。

preprocess\model	(1) all features		(2) PM2.5	
RMSE	public	private	public	private
(1)	10.85783	5.85780	6.24000	6.09963
(2)	7.22502	5.45240	6.06858	5.77200
(3)	5.83085	5.45078	5.96996	5.77371

(1) 把 nan 改成 0

(2) 刪掉包含 PM2.5> 500 的數據點(包含 feature 裡面有 PM2.5>500 的數據點)

(3) 刪掉包含 PM2.5> 200 的數據點(包含 feature 裡面有 PM2.5>200 的數據點)

3.(3%) Refer to math problem

1 - (a)

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

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

①

$\frac{\partial L_{ssq}}{\partial w} = \frac{1}{\cancel{10}_5}\sum_{i=1}^{5}\cancel{2}(y_i - (w x_i + b))(-1)x_i = 0$

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

$\sum_{i=1}^{5}(y_i x_i - w x_i^2 - b x_i) = 0$

②

$\frac{\partial L_{ssq}}{\partial b} = \cancel{\frac{1}{5}}\sum_{i=1}^{5}(y_i - (w x_i + b))\cancel{(-1)} = 0$

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

① $60.9 - 55w - 15b = 0$

② $16.8 - 15w - 5b = 0$

$50.4 - 45w - 15b = 0$

$10.5 - 10w = 0 \quad w = 1.05 \quad b = 0.21$

$y = 1.05x + 0.21$

1 - (b) $\quad (\vec{w}, b) \in R^k \times R$

$$L_{ssq}(\vec{w}, b) = \frac{1}{2N}\sum_{i=1}^{N}\left(y_i - (\vec{w}^T\vec{x}_i + b)\right)^2$$

let $\tilde{w} = \begin{bmatrix} b \\ \vec{w}\end{bmatrix} \in R^{k+1}$ $\qquad \tilde{w}^T\tilde{x}_i = [b \ \vec{w}^T]\begin{bmatrix}1\\ \vec{x}_i\end{bmatrix}$

$\tilde{x}_i = \begin{bmatrix}1\\ \vec{x}_i\end{bmatrix} \in R^{k+1}$ $\quad \Rightarrow \quad = b + \vec{w}^T\vec{x}_i$

$$L_{ssq}(\vec{w}, \vec{b}) = L_{ssq}(\tilde{w}) = \frac{1}{2N}\sum_{i=1}^{N}(y_i - \tilde{w}^T\tilde{x}_i)^2$$

$$= \frac{1}{2N}\|\vec{Y} - \tilde{X}^T\tilde{w}\|^2$$

let $\tilde{X} = \left[\tilde{x}_1 \ \tilde{x}_2 \cdots \tilde{x}_N\right]_{(k+1)\times N}$ $\qquad = \frac{1}{2N}(\vec{Y}-\tilde{X}^T\tilde{w})^T(\vec{Y}-\tilde{X}^T\tilde{w})$

$\vec{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}_{N\times 1}$ $\qquad \frac{\partial L_{ssq}}{\partial\tilde{w}} = \frac{1}{2N}\left\{\frac{\partial(\vec{Y} - \tilde{X}^T\tilde{w})}{\partial \tilde{w}}(\vec{Y} - \tilde{X}^T \tilde{w})\right.$

$$\left. + \frac{\partial(\vec{Y} - \tilde{X}^T\tilde{w})}{\partial\hat{w}}(\vec{Y} - \tilde{X}^T \tilde{w})\right\} = 0$$

$$\Rightarrow \frac{\partial(\vec{Y} - \tilde{X}^T\hat{w})}{\partial\tilde{w}}(\vec{Y} - \tilde{X}^T\tilde{w}) = 0$$

$$-\tilde{X}(\vec{Y} - \tilde{X}^T\tilde{w}) = 0 \Rightarrow \hat{X}\vec{Y} = \tilde{X}\tilde{X}^T\hat{w}$$

$$\tilde{w} = (\tilde{X}\tilde{X}^T)^{-1}\hat{X}\vec{Y} = \begin{bmatrix} b \\ \vec{w}\end{bmatrix} \quad ※$$

1 - (C)

$$L_{reg}(\vec{w}, b) = \frac{1}{2N} \sum_{i=1}^{N} (y_i - (\vec{w}^T \vec{x}_i + b))^2 + \frac{\lambda}{2} \|\vec{w}\|^2$$

· let $\tilde{w} = \begin{bmatrix} b \\ \vec{w} \end{bmatrix}$, $A = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & I_{k \times k} & \\ 0 & & \end{bmatrix}_{(k+1) \times (k+1)}$

$$L_{reg}(\vec{w}, b) = \frac{1}{2N} \|\vec{Y} - \tilde{X}^T \tilde{w}\|^2 + \frac{\lambda}{2} \|A\tilde{w}\|^2$$

$$= L_{reg}(\tilde{w})$$

$$\frac{\partial L_{reg}(\tilde{w})}{\partial \tilde{w}} = \frac{1}{N} \{-\tilde{X}(\vec{Y} - \tilde{X}^T \tilde{w})\} + \frac{\partial}{\partial \tilde{w}} \left\{ \frac{\lambda}{2} (A\tilde{w})^T (A\tilde{w}) \right\}$$

from 1 - (b)

$$= \frac{1}{N} \{-\tilde{X}(\vec{Y} - \tilde{X}^T \tilde{w})\} + \frac{\lambda}{\cancel{2}} \times \cancel{2} \frac{\partial (A\tilde{w})}{\partial \tilde{w}} (A\tilde{w})$$

$$= \frac{1}{N}(-\tilde{X}\vec{Y} + \tilde{X}\tilde{X}^T \tilde{w}) + \lambda A^T A \tilde{w} = 0$$

$$\tilde{X}\tilde{X}^T \tilde{w} + N\lambda A^T A \tilde{w} = \tilde{X}\vec{Y}$$

$$\tilde{w} = \begin{bmatrix} b \\ \vec{w} \end{bmatrix} = (\tilde{X}\tilde{X}^T + N\lambda A^T A)^{-1} \tilde{X} \vec{Y}$$

$$= (\tilde{X}\tilde{X}^T + N\lambda A)^{-1} \tilde{X} \vec{Y}$$

2.

$f_{\vec{w},b} : \mathbb{R}^k \to \mathbb{R}, \quad \vec{w} \in \mathbb{R}^k, \quad b \in \mathbb{R}$

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

$$S = \{(\vec{x}_i, y_i)\}_{i=1}^N, \quad \vec{x}_i \in \mathbb{R}^k, \quad \vec{\eta}_i \in \mathbb{R}^k$$

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

$\mathbb{E}[\eta_{i,j}] = 0 \qquad \mathbb{E}[\eta_{i,j}\eta_{i',j'}] = \delta_{i,i'}\delta_{j,j'}\sigma^2$

pf. $\tilde{L}_{ssq}(\vec{w}, b) = \frac{1}{2N}\sum_{i=1}^N \left(f_{\vec{w},b}(\vec{x}_i) - y_i\right)^2 + \frac{\sigma^2}{2}\|\vec{w}\|^2$

Hint. $\|\vec{x}\|^2 = \vec{x}^T\vec{x} = \mathrm{Trace}(\underset{k\times 1}{\vec{x}}\underset{1\times k}{\vec{x}^T})$

$$\mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^N \left(\vec{w}^T\vec{x}_i + \vec{w}^T\vec{\eta}_i + b - y_i\right)^2\right]$$

$$= \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^N \left(f_{\vec{w},b}(\vec{x}_i) - y_i + \vec{w}^T\vec{\eta}_i\right)^2\right]$$

$$= \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^N \left[\left(f_{\vec{w},b}(\vec{x}_i) - y_i\right)^2 + \left(\vec{w}^T\vec{\eta}_i\right)^2 + \cancel{2\left(f_{\vec{w},b}(\vec{x}_i) - y_i\right)\left(\vec{w}^T\vec{\eta}_i\right)}\right]\right] \xrightarrow{①} 0$$

$$= \frac{1}{2N}\sum_{i=1}^N \left(f_{\vec{w},b}(\vec{x}_i) - y_i\right)^2 + \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^N \mathrm{Trace}\left(\vec{w}^T\vec{\eta}_i\vec{\eta}_i^T\vec{w}\right)\right]$$

① since $\mathbb{E}[\vec{w}^T\vec{\eta}_i] = \mathbb{E}[w_1\eta_{i1} + \cdots w_k\eta_{ik}]$

$= w_1\mathbb{E}[\eta_{i1}] + \cdots w_k\mathbb{E}[\eta_{ik}] = 0$

② $\vec{\eta}_i\vec{\eta}_i^T = \begin{bmatrix}\eta_{i1}\\ \vdots \\ \eta_{ik}\end{bmatrix}\begin{bmatrix}\eta_{i1} & \cdots & \eta_{ik}\end{bmatrix} = I_{k\times k}\sigma^2$

$$= \frac{1}{2N}\sum_{i=1}^N \left(f_{\vec{w},b}(\vec{x}_i) - y_i\right)^2 + \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^N \mathrm{Trace}\left(\sigma^2\vec{w}^T\vec{w}\right)\right]$$

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

3.

K+1 models : $g_0, \dots g_K$

and $g_0(\vec{X}) = 0$

$$e_k = \frac{1}{N}\sum_{i=1}^{N}(g_k(\vec{x}_i) - y_i)^2, \quad k = 0, \dots K$$

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

3-(a) Express $\sum_{i=1}^{N} g_k(\vec{x}_i) y_i$ in terms of $N, e_0, e_1, \dots e_K, s_1, \dots s_K$

$$e_k = \frac{1}{N}\sum_{i=1}^{N}\{(g_k(\vec{x}_i))^2 + y_i^2 - 2 g_k(\vec{x}_i) y_i\}$$

$$= s_k + e_0 - \frac{2}{N}\sum_{i=1}^{N} g_k(\vec{x}_i) y_i$$

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

※

3-(b)

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

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

$$\frac{\partial L_{test}}{\partial \alpha_k} = \frac{1}{N}\sum_{i=1}^{N} 2\left\{\sum_{k=1}^{K} \alpha_k g_k(\vec{x}_i) - y_i\right\} g_k(\vec{x}_i)$$

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

$$= \left(2\sum_{k=1}^{K} s_k \alpha_k\right) + e_k - s_k - e_0$$

We can calculate $\frac{\partial L_{test}}{\partial \alpha_k}$ for $k = 1, \dots K$ to get K equations and solve for $\alpha_1, \dots \alpha_K$ by applying Gaussian Elimination on these system of equations.

Where the equations are :

$$\frac{\partial L_{test}}{\partial \alpha_k} = 0, for\ k = 1, \dots K$$