學號：B05902017 系級： 資工四 姓名：王盛立

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

$Feature$	$RMSE(public)$	$RMSE(private)$
$162+1$	5.5678	5.3715
$9+1$	5.8641	5.7151

當只用pm2.5當作feature時，可能因為數量太少所以導致預測的結果沒有那麼準確。用上所有的feature時就做得比較準確，儘管有些部分其實可以看出來影響是沒有那麼大可能拿掉結果會再好一點。

2.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。

我覺得負的資料都是不太合理的，所以我將負的資料都拿掉了。另外由於我是將連續的九個小時當成訓練資料，而第十個小時當成預測的結果。因此若是發生第十個小時跟前九個小時都差太多的狀況我也把他拿掉因為我覺得可能是比較奇怪的狀況不利於訓練。

$Data\ preprocess$	$RMSE(public)$	$RMSE(private)$
yes	5.5678	5.3715
no	5.7506	5.5083

3.Refer to math problem

1-(a)

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

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

$= \frac{1}{10}((w+b-1.2)^2 + (2w+b-2.4)^2 + (3w+b-3.6)^2 + (4w+b-4.1)^2 + (5w+b-5.6)^2)$

$\frac{\partial}{\partial w}L_{ssq} = \frac{1}{10}(2\times(w+b-1.2)\times 1 + 2\times(2w+b-2.4)\times 2 + 2\times(3w+b-3.6)\times 3 + 2\times(4w+b-4.1)\times 4$
$+ 2\times(5w+b-5.6)\times 5)$

$= 11w + 3b - 12.18$

$\frac{\partial}{\partial b}L_{ssq} = \frac{1}{10}(2 \times (w + b - 1.2) \times 1 + 2 \times (2w + b - 2.4) \times 1 + 2 \times (3w + b - 3.6) \times 1 + 2 \times (4w + b - 4.1) \times 1$
$+ 2 \times (5w + b - 5.6) \times 5)$

$= 3w + b - 3.36$

Let $\frac{\partial}{\partial b}$ and $\frac{\partial}{\partial w}$ = 0

$$\begin{cases} 11w + 3b - 12.18 = 0 \\ 3w + b - 3.36 = 0 \end{cases}$$

w = 1.05 and b = 0.21

1-(b)

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

$$x = \begin{bmatrix} x_1^T \\ x_2^T \\ x_3^T \\ \vdots \\ x_N^T \end{bmatrix} = \begin{bmatrix} x_{10} & x_{11} & x_{12} & \cdots & x_{1k} \\ x_{20} & x_{21} & x_{22} & \cdots & x_{2k} \\ x_{30} & x_{31} & x_{32} & \cdots & x_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N0} & x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix}$$

$w = \begin{bmatrix} w_0 & w_1 & w_2 & \cdots & w_k \end{bmatrix}$

$y = \begin{bmatrix} y_0 & y_1 & y_2 & \cdots & y_N \end{bmatrix}$

Let $b = w_0 x_0$

L_{ssq} = $\frac{1}{2N}\sum_{i=1}^{N}(\sum_{j=0}^{k} w_j x_{ij} - y_i)^2$

Let $\begin{bmatrix} \frac{\partial}{\partial w_0}L_{ssq} \\ \frac{\partial}{\partial w_1}L_{ssq} \\ \frac{\partial}{\partial w_2}L_{ssq} \\ \vdots \frac{\partial}{\partial w_k}L_{ssq} \end{bmatrix} = 0$

$$\Rightarrow \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} x_{i_0}(\sum_{j=0}^{K} w_j x_{ij} - y_i) \\ \frac{1}{N}\sum_{i=1}^{N} x_{i_1}(\sum_{j=0}^{K} w_j x_{ij} - y_i) \\ \frac{1}{N}\sum_{i=1}^{N} x_{i_2}(\sum_{j=0}^{K} w_j x_{ij} - y_i) \\ \vdots \\ \frac{1}{N}\sum_{i=k}^{N} x_{i_0}(\sum_{j=0}^{K} w_j x_{ij} - y_i) \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} x_{10} & x_{20} & x_{30} & \cdots & x_{N0} \\ x_{11} & x_{21} & x_{31} & \cdots & x_{N1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{N2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1k} & x_{2k} & x_{3k} & \cdots & x_{Nk} \end{bmatrix} (\begin{bmatrix} x_{10} & x_{11} & x_{12} & \cdots & x_{1k} \\ x_{20} & x_{21} & x_{22} & \cdots & x_{2k} \\ x_{30} & x_{31} & x_{32} & \cdots & x_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N0} & x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix} - \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}) = 0$$

$\Rightarrow x^T(xw - y) = 0$

$x^T y = (x^T x w)$

$w = (x^T x)^{-1} x^T y$

1-(c)

$w^T x_i + b = \sum_{j=0}^{k} w_j x_{ij}$

$w^2 = \sum_{i=1}^{k} w_i^2 = \sum_{i=0}^{k} w_i^2 - b^2$

$L_{reg} = \frac{1}{2N} \sum_{i=1}^{N} (y_i - (w^T x_i + b)^2) + \frac{\lambda}{2} \|\mathbf{w}\|^2 = \frac{1}{2N} \sum_{i=1}^{N} (y_i - \sum_{j=0}^{k} w_j x_{ij})^2 + \frac{\lambda}{2} \sum_{i=0}^{k} (w_i^2 - b^2)$

$\frac{\partial}{\partial w_n} L_{reg} = \frac{\partial}{\partial w_n} (\frac{1}{2N} \sum_{i=1}^{N} (\sum_{j=0}^{k} w_j x_{ij} - y_i)^2 + \frac{\lambda}{2} \sum_{i=0}^{k} (w_i^2 - b^2))$
$= \frac{1}{2N} \sum_{i=1}^{N} 2(\sum_{j=0}^{k} w_j x_{ij} - y_i) x_{in}) + \frac{\lambda}{2} 2 w_n = \frac{1}{N} \sum_{i=1}^{N} (\sum_{j=0}^{k} w_j x_{ij} - y_i) x_{in}) + \lambda w_n$

Let $\begin{bmatrix} \frac{\partial}{\partial w_0} L_{reg} \\ \frac{\partial}{\partial w_1} L_{reg} \\ \frac{\partial}{\partial w_2} L_{reg} \\ \vdots \frac{\partial}{\partial w_k} L_{reg} \end{bmatrix} = 0$

$\Rightarrow \begin{bmatrix} \frac{1}{N} \sum_{i=1}^{N} x_{i_0} (\sum_{j=0}^{K} w_j x_{ij} - y_i) + \lambda w_0 \\ \frac{1}{N} \sum_{i=1}^{N} x_{i_1} (\sum_{j=0}^{K} w_j x_{ij} - y_i) + \lambda w_1 \\ \frac{1}{N} \sum_{i=1}^{N} x_{i_2} (\sum_{j=0}^{K} w_j x_{ij} - y_i) + \lambda w_2 \\ \vdots \\ \frac{1}{N} \sum_{i=k}^{N} x_{i_0} (\sum_{j=0}^{K} w_j x_{ij} - y_i) + \lambda w_k \end{bmatrix} = 0$

$\Rightarrow \frac{1}{N} (\begin{bmatrix} x_{10} & x_{20} & x_{30} & \cdots & x_{N0} \\ x_{11} & x_{21} & x_{31} & \cdots & x_{N1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{N2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1k} & x_{2k} & x_{3k} & \cdots & x_{Nk} \end{bmatrix} (\begin{bmatrix} x_{10} & x_{11} & x_{12} & \cdots & x_{1k} \\ x_{20} & x_{21} & x_{22} & \cdots & x_{2k} \\ x_{30} & x_{31} & x_{32} & \cdots & x_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N0} & x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix} - \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix})) + \lambda \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = 0$

$\Rightarrow \frac{1}{N} (x^T (xw - y)) + \lambda w = 0$

$\Rightarrow (x^T x + N \lambda I) w = x^T y$

$w = (x^T x + N \lambda I)^{-1} x^T y$

2

$f_{\mathbf{w},b}(\mathbf{x}_i + \eta_i) = w^T (x_i + \eta_i) + b = f_{\mathbf{w},b}(\mathbf{x}_i) + w^T \eta_i$

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

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

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

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

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

$\mathbb{E} \left[\sum_{i=1}^{N} (2(f_{\mathbf{w},b}(\mathbf{x}_i) - y_i)(w^T \eta_i)) \right]$ = $\sum_{i=1}^{N} (2(f_{\mathbf{w},b}(\mathbf{x}_i) - y_i) \mathbb{E} \left[w^T \eta_i \right]$ =
$\sum_{i=1}^{N} (2(f_{\mathbf{w},b}(\mathbf{x}_i) - y_i) \sum_{j=1}^{k} (w_j \mathbb{E} \left[\eta_{ij} \right]])$ = 0

$\mathbb{E}\left[\sum_{i=1}^{N}(w^T\eta_i)^2)\right] = \sum_{i=1}^{N}\mathbb{E}\left[(w^T\eta_i)^2)\right] = \sum_{i=1}^{N}\mathbb{E}\left[(\sum_{j=1}^{k}w_j\eta_{ij})^2\right]$
$= \sum_{i=1}^{N}\mathbb{E}\left[\sum_{j=1}^{k}\sum_{j'=1}^{k}w_jw_{j'}\eta_{ij}\eta_{ij'}\right] = \sum_{i=1}^{N}(\sum_{j=1}^{k}\sum_{j'=1}^{k}w_jw_{j'}\mathbb{E}\left[\eta_{ij}\eta_{ij'}\right])$
$= \sum_{i=1}^{N}(\sum_{j=1}^{k}\sum_{j'=1}^{k}(w_jw_{j'}\delta_{i,i'}\delta_{j,j'}\sigma^2)) = \sum_{i=1}^{N}(\sum_{j=1}^{k}w_j^2\sigma^2) = \sum_{i=1}^{N}\sigma^2|\mathbf{w}\|^2$
$= N\sigma^2|\mathbf{w}\|^2$

$\Rightarrow \tilde{L}_{ssq}(\mathbf{w},b) = \frac{1}{2N}(\sum_{i=1}^{N}(f_{\mathbf{w},b}(\mathbf{x}_i)-y_i)^2 + 0 + N\sigma^2|\mathbf{w}\|^2)$
$= \frac{1}{2N}\sum_{i=1}^{N}(f_{\mathbf{w},b}(\mathbf{x}_i)-y_i)^2 + \frac{\sigma^2}{2}\|\mathbf{w}\|^2$

3-(a)

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

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

3-(b)

$\frac{\partial}{\partial\alpha_n}L_{test} = \frac{\partial}{\partial\alpha_n}(\frac{1}{N}\sum_{i=1}^{N}(\sum_{j=1}^{k}\alpha_k g_k(x_i)-y_i)^2) = \frac{2}{N}(\sum_{i=1}^{N}(\alpha_k g_k(x_i)-y_i)g_n(x_i))$

Let $\begin{bmatrix}\frac{\partial}{\partial\alpha_0}L_{test}\\ \frac{\partial}{\partial\alpha_1}L_{test}\\ \frac{\partial}{\partial\alpha_2}L_{test}\\ \vdots\frac{\partial}{\partial\alpha_k}L_{test}\end{bmatrix} = 0$

$\Rightarrow \begin{bmatrix}\frac{2}{N}(\sum_{i=1}^{N}(\alpha_k g_k(x_i)-y_i)g_1(x_i))\\ \frac{2}{N}(\sum_{i=1}^{N}(\alpha_k g_k(x_i)-y_i)g_2(x_i))\\ \frac{2}{N}(\sum_{i=1}^{N}(\alpha_k g_k(x_i)-y_i)g_3(x_i))\\ \vdots\\ \frac{2}{N}(\sum_{i=1}^{N}(\alpha_k g_k(x_i)-y_i)g_n(x_i))\end{bmatrix} = 0$

$\Rightarrow$

$$\begin{bmatrix}g_1(x_1) & g_1(x_2) & g_1(x_3) & \cdots & g_1(x_N)\\ g_2(x_1) & g_2(x_2) & g_2(x_3) & \cdots & g_2(x_N)\\ g_3(x_1) & g_3(x_2) & g_3(x_3) & \cdots & g_3(x_N)\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ g_k(x_1) & g_k(x_2) & g_k(x_3) & \cdots & g_k(x_N)\end{bmatrix}\left(\begin{bmatrix}g_1(x_1) & g_2(x_1) & g_3(x_1) & \cdots & g_k(x_1)\\ g_1(x_2) & g_2(x_2) & g_3(x_2) & \cdots & g_k(x_2)\\ g_1(x_3) & g_2(x_3) & g_3(x_3) & \cdots & g_k(x_3)\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ g_1(x_N) & g_2(x_N) & g_3(x_N) & \cdots & g_k(x_N)\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\\ \vdots\\ \alpha_N\end{bmatrix} - \begin{bmatrix}y_1\\ y_2\\ y_3\\ \vdots\\ y_N\end{bmatrix}\right) = 0$$

Let $Z = \begin{bmatrix}g_1(x_1) & g_2(x_1) & g_3(x_1) & \cdots & g_k(x_1)\\ g_1(x_2) & g_2(x_2) & g_3(x_2) & \cdots & g_k(x_2)\\ g_1(x_3) & g_2(x_3) & g_3(x_3) & \cdots & g_k(x_3)\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ g_1(x_N) & g_2(x_N) & g_3(x_N) & \cdots & g_k(x_N)\end{bmatrix}$

$\Rightarrow z^T z\alpha = z^T y$

From 3-(a) we get $z^T z\alpha = \begin{bmatrix}\frac{N}{2}(S_1 - e_1 + e_0)\\ \frac{N}{2}(S_2 - e_2 + e_0)\\ \frac{N}{2}(S_3 - e_3 + e_0)\\ \vdots\\ \frac{N}{2}(S_k - e_k + e_0)\end{bmatrix}$

$$\Rightarrow \alpha = (z^T z)^{-1} \begin{bmatrix} \frac{N}{2}(S_1 - e_1 + e_0) \\ \frac{N}{2}(S_2 - e_2 + e_0) \\ \frac{N}{2}(S_3 - e_3 + e_0) \\ \vdots \\ \frac{N}{2}(S_k - e_k + e_0) \end{bmatrix}$$

$\Rightarrow$

$$\alpha = (\begin{bmatrix} g_1(x_1) & g_1(x_2) & g_1(x_3) & \cdots & g_1(x_N) \\ g_2(x_1) & g_2(x_2) & g_2(x_3) & \cdots & g_2(x_N) \\ g_3(x_1) & g_3(x_2) & g_3(x_3) & \cdots & g_3(x_N) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g_k(x_1) & g_k(x_2) & g_k(x_3) & \cdots & g_k(x_N) \end{bmatrix} \begin{bmatrix} g_1(x_1) & g_2(x_1) & g_3(x_1) & \cdots & g_k(x_1) \\ g_1(x_2) & g_2(x_2) & g_3(x_2) & \cdots & g_k(x_2) \\ g_1(x_3) & g_2(x_3) & g_3(x_3) & \cdots & g_k(x_3) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g_1(x_N) & g_2(x_N) & g_3(x_N) & \cdots & g_k(x_N) \end{bmatrix})^{-1} \begin{bmatrix} \frac{N}{2}(S_1 - e_1 + e_0) \\ \frac{N}{2}(S_2 - e_2 + e_0) \\ \frac{N}{2}(S_3 - e_3 + e_0) \\ \vdots \\ \frac{N}{2}(S_k - e_k + e_0) \end{bmatrix}$$