Homework1 Report – PM 2.5 Prediction

B04703117 財金五 謝昊辰

請實做以下兩種不同 feature 的模型，回答第 (1) ~ (2) 題：

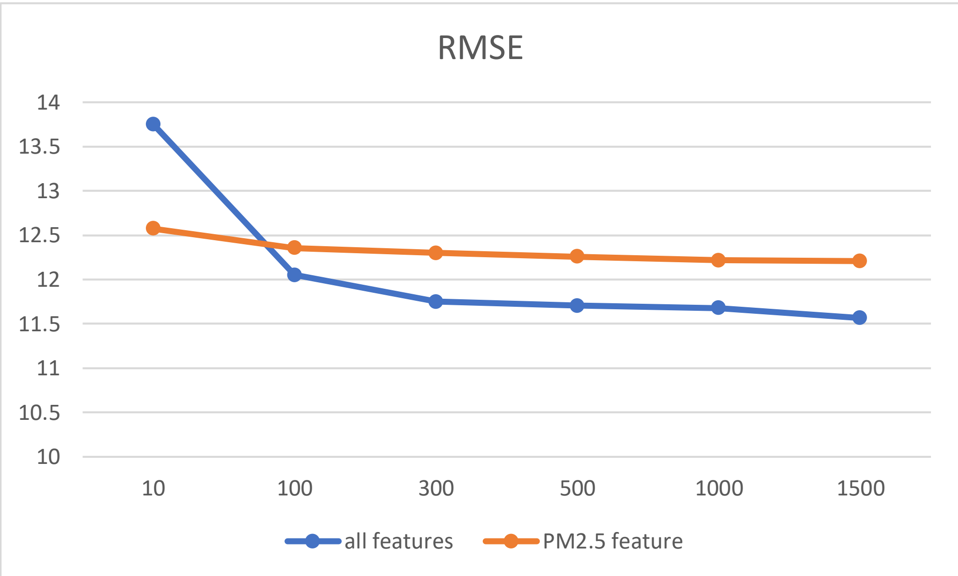
1. 抽全部 9 小時內的污染源 feature 當作一次項(加 bias)
2. 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias)

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

下圖為做不同次數的 gradient descent 的 RMSE(public + private)情形，可以看到只有 PM2.5 當 feature 時，GD 次數少時 error 甚至比較低，可能原因是因為 feature 的量比較少時，比較容易去 fit data，但 feature 量比較大時，卻可以得到比較好的 RMSE。



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。

在這次的資料中有發現 PM10 以及 PM2.5 中皆有不合理的值，e.g. 1000 和負數，將 PM10 以及 PM2.5 中不合理的值去除，可以避免在做 linear regression 時被極端值影響回歸直線的挑選，若沒有將這些值去除，利用 Adam Optimizer 上傳 kaggle 三次得到的 RMSE 平均(private + public)為 6.690，相較於去除極端值的 RMSE 平均(private + public) 5.648，相差 1.042。

1. Collaborator: B04901140 連潔琳

1.

(a)

$$\frac{\partial L}{\partial w} = \frac{2}{10}\sum_{i=1}^{5}\left(y_i - (w^T x_i + b)\right)(-x_i) = 0$$

$$\rightarrow (1.2-(w\times1+b))\times(-1) + (2.4-(w\times2+b))\times(-2) + (3.5-(w\times3+b))\times(-3)$$
$$+ (4.1-(w\times4+b))\times(-4) + (5.6-(w\times5+b))\times(-5) = -60.9 + 55w + 15b = 0$$

$$\frac{\partial L}{\partial b} = \frac{2}{10}\sum_{i=1}^{5}\left(y_i - (w^T\cdot x_i + b)\right) = 0$$

$$\rightarrow (1.2-(w\times1+b)) + (2.4-(w\times2+b)) + (3.5-(w\times3+b)) + (4.1-(w\times4+b))$$
$$+ (5.6-(w\times5+b)) = 16.8 - 15w - 5b = 0$$

$$\begin{cases} 55w+15b = 60.9 \\ 15w+5b = 16.8 \end{cases} \rightarrow \begin{cases} 55w+15b = 60.9 \\ 45w+15b = 50.4 \end{cases} \Rightarrow 10w = 10.5 \rightarrow \boxed{\begin{matrix} w = 1.05 \\ b = 0.21 \end{matrix}}$$

令 $\tilde{w} = \begin{bmatrix} b \\ w \end{bmatrix} \in \mathbb{R}^{k+1}$ $\quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^N$

$$X^T = \begin{bmatrix} 1 & x_1^T \\ & \vdots \\ 1 & x_N^T \end{bmatrix} \in \mathbb{R}^{(k+1)\times N}$$

(b)

$$\Rightarrow \min_{\tilde{w}} \tilde{L}(\tilde{w}) = \|X^T\tilde{w} - y\|^2 \rightarrow \nabla_{\tilde{w}}\tilde{L}(\tilde{w}) = 2X^TX\tilde{w} - 2X^Ty = 0$$

$$\tilde{w} = \begin{bmatrix} b \\ w \end{bmatrix} = (XX^T)^{-1}\cdot X^Ty$$

(c)

$$\nabla L(w,b) = \frac{1}{N}(X^T\tilde{w} - y)\cdot X + \lambda\tilde{w} = 0$$

$$\rightarrow \frac{1}{N}(X^TX\tilde{w} - X^Ty) + \lambda\tilde{w} = 0$$

$$\rightarrow \frac{1}{N}X^TX\tilde{w} + \lambda\tilde{w} = \frac{1}{N}X^Ty$$

$$\rightarrow \tilde{w} = \left(\frac{1}{N}X^TX + \lambda I\right)^{-1}\cdot\frac{1}{N}X^Ty$$

2. Collaborator: r08921a14 曾浩偉

令 $\tilde{w} = \begin{bmatrix} b \\ w \end{bmatrix}$ $\tilde{x} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}$ $\tilde{\eta} = \begin{bmatrix} 0 & \eta_1 \\ \vdots & \vdots \\ 0 & \eta_N \end{bmatrix}$

$$L(w,b) = E\left[\frac{1}{2N}\left((\tilde{x}+\tilde{\eta})^T\tilde{w} - y\right)^2\right]$$

$$\nabla_{\tilde{w}} L(\tilde{w}) = E\left[\frac{1}{N}(\tilde{x}+\tilde{\eta})\left((\tilde{x}+\tilde{\eta})^T\tilde{w} - y\right)\right] = E\left[\frac{1}{N}\left(\|\tilde{x}+\tilde{\eta}\|^2 \times \tilde{w} - (\tilde{x}+\tilde{\eta})^T y\right)\right]$$

$$= \frac{1}{N}\tilde{x}^T\tilde{x}\tilde{w} + \frac{1}{N}\underbrace{E(\tilde{\eta})}_{0}\tilde{x}^T\tilde{w} + \frac{1}{N}\underbrace{E(\tilde{\eta}^2)}_{N\cdot\sigma^2}\tilde{w} - \frac{1}{N}(\tilde{x} + E(\tilde{\eta}))^T y$$

$$= \frac{1}{N}\tilde{x}^T\tilde{x}\tilde{w} + \sigma^2\tilde{w} - \frac{1}{N}(\tilde{x}+0)^T y = 0$$

$\to \tilde{w} = \left(\frac{1}{N}\tilde{x}^T\tilde{x} + \sigma^2 I\right)^{-1}\cdot\frac{1}{N}\tilde{x}^T y$ (同 1-(c) 的答案(將 λ 替換成 σ^2))

3. Collaborator: r08921a14 曾浩偉

3.

(a)

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

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

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

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

(b) 令 $\underset{(k\times N)}{G} = \begin{bmatrix} g_1(x_1) & \cdots & g_1(x_N) \\ \vdots & & \vdots \\ g_k(x_1) & \cdots & g_k(x_N) \end{bmatrix}$ $\underset{(N\times 1)}{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$ $\underset{(k\times 1)}{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix}$

$$\min_{\alpha_1 \ldots \alpha_n} L_{test} = \frac{1}{N}(G^T\alpha - Y)^2$$

$$\partial L_{test}/\partial\alpha = \frac{2}{N}(GG^T\alpha - GY) = 0$$

$$\Rightarrow \alpha = (GG^T)^{-1}GY$$

$$= \begin{bmatrix} s_1 \\ \vdots \\ s_k \end{bmatrix}^{-1} \cdot \begin{bmatrix} \frac{N}{2}(s_1 - e_1 + e_0) \\ \vdots \\ \frac{N}{2}(s_k - e_k + e_0) \end{bmatrix}$$

$$= \begin{bmatrix} s_1^{-1} \times \frac{N}{2}(s_1 - e_1 + e_0) \\ \vdots \\ s_k^{-1} \times \frac{N}{2}(s_k - e_k + e_0) \end{bmatrix}$$