

說明：請各位使用此 template 進行 Report 撰寫，如果想要用其他排版模式也請註明題號以及題目內容（請勿擅自更改題號），最後上傳至 github 前，請務必轉成 PDF 檔，並且命名為 report.pdf，否則將不予計分。

-----閱讀完以上文字請刪除-----

學號：R07941023 系級：光電碩二 姓名：呂彥穎

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)

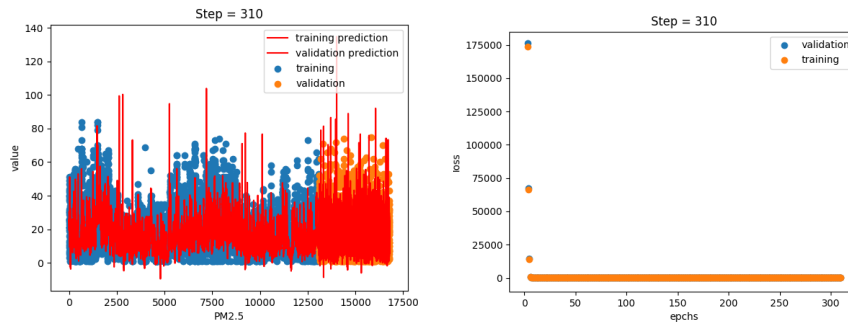
備註：

- 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 $p = 9 \times 18 + 1$ 而(2) 代表 $p = 9 * 1 + 1$

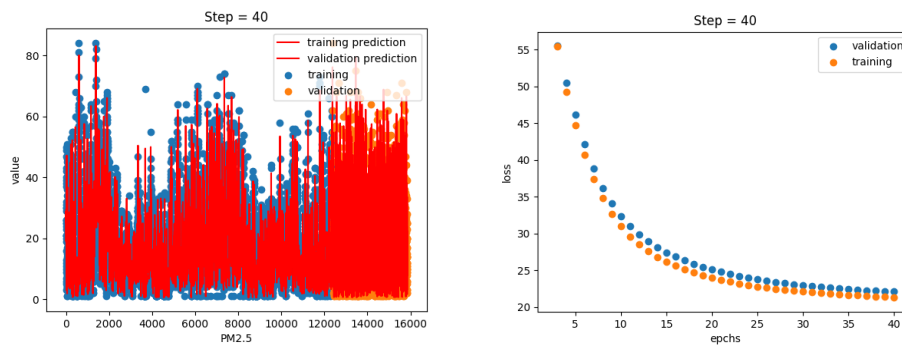
1. (1%)記錄誤差值 (RMSE)(根據 kaggle public+private 分數)，討論兩種 feature 的影響

Sol:

(1) 代表 $p = 9 \times 18 + 1$, kaggle public+private 分數: 6.65461+5.94744



(2) 代表 $p = 9 \times 1 + 1$, kaggle public+private 分數: 6.16929+6.07167



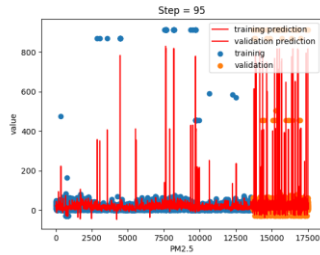
如果過多的考量多餘的 feature, 且每個 feature 的資料前處理不同(除了 PM2.5 項,其餘均未前處理), 其每一項均會增加 weight 的誤差, 在此前提下, 故不是越考慮越多越好.

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。

Sol: 錯誤的 training data 會增加 weight 的誤差

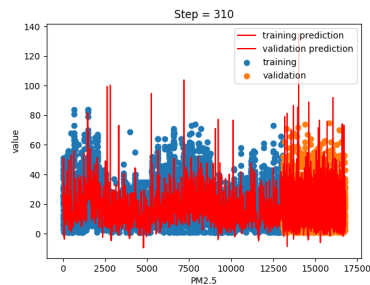
(1) 代表 $p = 9 \times 18 + 1$, kaggle public+private 分數: 5.86819 + 6.22921

未處理 PM2.5 項



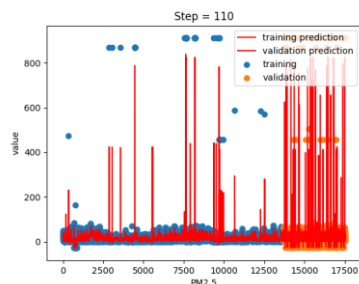
(1) 代表 $p = 9 \times 18 + 1$, kaggle public+private 分數: 6.65461+5.94744

處理 PM2.5 項, 值 $PM2.5 > 0$ 以及 $PM2.5 < 100$



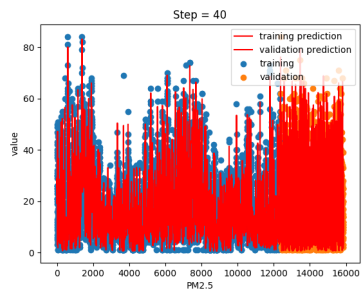
(2) 代表 $p = 9 \times 1 + 1$, kaggle public+private 分數: 6.02156 + 6.15972

未處理 PM2.5 項



(2) 代表 $p = 9 \times 1 + 1$, kaggle public+private 分數: 6.16929+6.07167

處理 PM2.5 項, 值 $PM2.5 > 0$ 以及 $PM2.5 < 100$



3.(3%) Refer to math problem

<https://hackmd.io/RFiu1FsYR5uQTrrpdXUv1w?view>

(p1)

(2) L_{min}

$$\begin{cases} \frac{\partial L}{\partial w} = 0 - (1) \\ \frac{\partial L}{\partial b} = 0 - (2) \end{cases}, L_{sq}(\vec{w}, b) = \frac{1}{2N} \sum_{i=1}^N (y_i - (w^T x_i + b))^2$$

from (1), $\frac{\partial L}{\partial w} = \frac{1}{N} \sum_{i=1}^N x_i (y_i - (w^T x_i + b))$

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

from (2), $\frac{\partial L}{\partial b} = \frac{1}{N} \sum_{i=1}^N (y_i - w^T x_i - b) = 0$

$$b = \bar{y} - \vec{w}^T \bar{x} - (3)$$

$$= \frac{1}{N} \sum_{i=1}^N (x_i y_i - w^T x_i^2 - \bar{y} x_i + \vec{w}^T x_i \bar{x})$$

$$= \frac{1}{N} \sum_{i=1}^N [w^T (x_i \bar{x} - x_i^2) - (\bar{y} x_i - x_i y_i)]$$

$$\Rightarrow w^T \sum_{i=1}^N (x_i \bar{x} - x_i^2) = \bar{y} \sum_{i=1}^N x_i - \sum_{i=1}^N x_i y_i$$

$$\Rightarrow w^T = \frac{\bar{y} \sum_{i=1}^N x_i - \sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i \bar{x} - \sum_{i=1}^N x_i^2}, \text{代入(3)}$$

$$b = \bar{y} - \frac{\bar{y} \sum_{i=1}^N x_i - \sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i \bar{x} - \sum_{i=1}^N x_i^2} \bar{x} - (5)$$

1) $\bar{x} = 3$
 $\bar{y} = 3.36$
 $\sum_{i=1}^5 x_i = 15$
 $\sum_{i=1}^5 x_i y_i = 60.9$
 $\sum_{i=1}^5 x_i^2 = 55$

代入(4), $w^T = \frac{3.36 \times 15 - 60.9}{15 \times 3 - 55} = 1.05$

代入(5), $b = 3.36 - 1.05 \times 3 = 0.21$

(p3)

L_{min}

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \quad \text{--- } \textcircled{1} \\ \frac{\partial L}{\partial b} = 0 \quad \text{--- } \textcircled{2} \end{cases} \quad L(w, b) = \frac{1}{2N} \sum_{i=1}^N (y_i - (w^T x_i + b))^2 + \frac{\lambda}{2} \|w\|^2$$
$$\textcircled{1} \Rightarrow \frac{\partial L}{\partial w} = \frac{1}{N} \sum_{i=1}^N (y_i x_i^T + w x_i^T + b x_i^T) + \lambda \|w\| \quad \text{from } \textcircled{2} \quad \frac{\partial L}{\partial b} = \frac{1}{N} \sum_{i=1}^N (y_i + w x_i^T + b) = 0$$

$$= \frac{1}{N} \sum_{i=1}^N [-y_i x_i^T + w x_i^T + \bar{y} x_i^T - w \bar{x} x_i^T + \lambda w]$$

$$\Rightarrow w \left[\sum_{i=1}^N x_i x_i^T - \bar{x}^T \sum_{i=1}^N x_i^T + \lambda \right] = \sum_{i=1}^N y_i x_i^T - \bar{y} \sum_{i=1}^N x_i^T$$

$$\Rightarrow w = \frac{\sum_{i=1}^N y_i x_i^T - \bar{y} \sum_{i=1}^N x_i^T}{\sum_{i=1}^N x_i x_i^T - \bar{x}^T \sum_{i=1}^N x_i^T + \lambda} \quad \text{A7 } \textcircled{3}$$

$$\Rightarrow b = \bar{y} - \frac{\sum_{i=1}^N y_i x_i^T - \bar{y} \sum_{i=1}^N x_i^T}{\sum_{i=1}^N x_i x_i^T - \bar{x}^T \sum_{i=1}^N x_i^T + \lambda} \bar{x}^T$$

(P2)

$$\begin{aligned}
 \tilde{L}_{sq}(w, b) &= E \left[\frac{1}{2N} \sum_{i=1}^N (f_{w,b}(x_i + \eta_i) - y_i)^2 \right], \quad f_{w,b}(x_i + \eta_i) = w^T (x_i + \eta_i) + b \\
 &= E \left[\frac{1}{2N} \sum_{i=1}^N (w_i^T x_i + w_i^T \eta_i + b - y_i)^2 \right] \\
 &= E \left[\frac{1}{2N} \sum_{i=1}^N (f_{w,b}(x_i) + w_i^T \eta_i - y_i)^2 \right] \\
 &= E \left[\frac{1}{2N} \sum_{i=1}^N \left[(f_{w,b}(x_i) - y_i)^2 + f_{w,b}(x_i) w_i^T \eta_i + (w_i^T \eta_i) f_{w,b}(x_i) + \right. \right. \\
 &\quad \left. \left. + (w_i^T \eta_i)^T w_i^T \eta_i - (w_i^T \eta_i)^T y_i - (y_i w_i^T) \eta_i \right] \right] \\
 &= E[\eta_i, y_i] \quad = E \left[\frac{1}{2N} \sum_{i=1}^N \left[(f_{w,b}(x_i) - y_i)^2 + (w_i^T \eta_i)^T w_i^T \eta_i \right] \right], \quad E[\eta_i, y_i] = \delta_{ij} \delta_{ij} \sigma^2 \\
 &= E \left[\frac{1}{2N} \sum_{i=1}^N \left[(f_{w,b}(x_i) - y_i)^2 + \|\eta_i\|^2 \sigma^2 \right] \right] \\
 &= E \left[\frac{1}{2N} \sum_{i=1}^N \left[(f_{w,b}(x_i) - y_i)^2 \right] + \frac{1}{2} \|\eta_i\|^2 \sigma^2 \right]
 \end{aligned}$$

(p3)

$$\begin{aligned}
 (a) \quad e_k &= \frac{1}{N} \sum_{i=1}^N (g_k(x_i) - y_i)^2 \\
 &= \frac{1}{N} \sum_{i=1}^N [(g_k(x_i))^2 + 2g_k(x_i)y_i + y_i^2] \quad \left\{ \begin{array}{l} s_k = \frac{1}{N} \sum_{i=1}^N (g_k(x_i))^2 \quad - (1) \\ e_0 = \frac{1}{N} \sum_{i=1}^N y_i^2 \quad - (2) \end{array} \right. \\
 &= s_k + e_0 + \frac{1}{N} \sum_{i=1}^N 2g_k(x_i)y_i \\
 \Rightarrow \sum_{i=1}^N g_k(x_i)y_i &= \frac{N(e_k - s_k - e_0)}{2} \quad - (3)
 \end{aligned}$$

$$\begin{aligned}
 (b) \quad \frac{\partial L}{\partial \alpha} &= 0, \quad L = \frac{1}{N} \sum_{i=1}^N \left(\sum_{k=1}^K \alpha_k g_k(x_i) - y_i \right)^2 \\
 &= \frac{1}{N} \sum_{i=1}^N \left(\sum_{k=1}^K \alpha_k g_k^2(x_i) - y_i g_k(x_i) \right) \quad \text{from (1), (2), (3)} \\
 &= \frac{1}{N} \sum_{k=1}^K \left(\alpha_k \sum_{i=1}^N g_k^2(x_i) - \frac{N(e_k - s_k - e_0)}{2} \right) \\
 &= \sum_{k=1}^K \left(2\alpha_k s_k - (e_k - s_k - e_0) \right) \\
 \Rightarrow \alpha_k &= \frac{e_k - s_k - e_0}{2s_k}
 \end{aligned}$$