

學號：B05902018 系級：資工三 姓名：張凱程

1. (2%)記錄誤差值 (RMSE)(根據 kaggle public+private 分數)，討論兩種 feature 的影響

(1) 全部 feature: 5.51664 + 7.07990

(2) PM2.5 feature: 5.79583 + 7.17286

因為(1)的 feature 比(2)多，可以做參考的項目會比較多，因此前者會去找非 PM2.5 的 feature 跟 PM2.5 的相關性，會增加其準確率。

(1) 全部 feature: 5.51664 + 7.07990 \rightarrow 5.88784 + 7.11336

(2) PM2.5 feature: 5.79583 + 7.17286

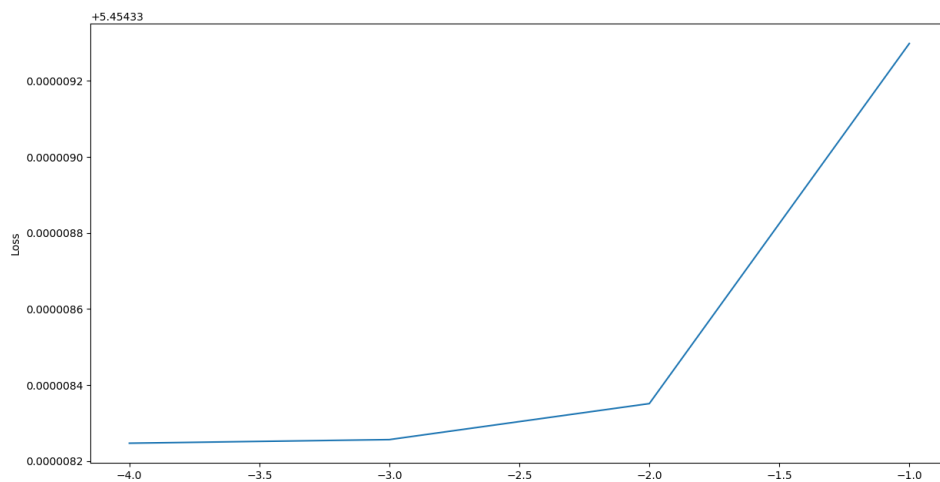
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

(1) 全部 feature: 5.51664 + 7.07990 \rightarrow 5.88784 + 7.11336

(2) PM2.5 feature: 5.79583 + 7.17286 \rightarrow 6.16914 + 7.15599

由於剩下的 4 小時還是有其參考性，他們的 weight 也需要進行學習，所以正確率會比較差。

3. (1%)Regularization on all the weight with $\lambda=0.1$ 、 0.01 、 0.001 、 0.0001 ，並作圖



因為 training 次數大的關係(epoch = 100000)，使得效果不是很明顯，但可以看出 λ 越大，training data 的 loss 會越大，其中 x 軸為 $\log \lambda$ 。

4. (1%)在線性回歸問題中，假設有 N 筆訓練資料，每筆訓練資料的特徵(feature) 為一向量 \mathbf{x}^n ，其標註(label)為一純量 y^n ，模型參數為一向量 \mathbf{w} (此處忽略偏權值 b)，則線性回歸的損失函數(loss function)為 $\sum_{n=1}^N (y^n - \mathbf{x}^n \mathbf{w})^2$ 。若將所有訓練資料的特徵值以矩陣 $\mathbf{X} = [\mathbf{x}^1 \mathbf{x}^2 \dots \mathbf{x}^N]^T$ 表示，所有訓練資料的標註以向量 $\mathbf{y} = [y^1 y^2 \dots y^N]^T$ 表示，請問如何以 \mathbf{X} 和 \mathbf{y} 表示可以最小化損失函數的向量 \mathbf{w} ？請選出正確答案。(其中 $\mathbf{X}^T \mathbf{X}$ 為 invertible)

- a. $(\mathbf{X}^T \mathbf{X}) \mathbf{X}^T \mathbf{y}$
- b. $(\mathbf{X}^T \mathbf{X}) \mathbf{y} \mathbf{X}^T$
- c. $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$
- d. $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{y} \mathbf{X}^T$

Answer: c.