學號：B05702095 系級： 會計四 姓名：黃禹翔

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

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。

一開始我將缺值用各變數的平均填補，在公開測試集上的 RMSE 大約在 6.44206，後來改成跟助教一樣用 0 填補所有缺值，結果成績居然反而變好，來到 5.65706，老實說做到這裏我也滿困惑的，按照 OLS 的想法，如果是用 0 填補應該會使係數的估計偏移，但不知道為什麼用 0 填補反而變好了。

3.(3%) Refer to math problem
https://hackmd.io/RFiu1FsYR5uQTrrpdxUvlw?view

HW1 – Handwriting part:

1-b.

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

令 $\underline{\underline{X}} = \begin{bmatrix} -x_1- & 1 \\ x_2 & \vdots \\ x_N & 1 \end{bmatrix}$ $\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_N \end{pmatrix}$ $\underline{\theta} = \begin{pmatrix} \underline{w} \\ b \end{pmatrix}$

$$L(w,b) = \frac{1}{2N}(\underline{\underline{X}}\underline{\theta} - \underline{y})^T(\underline{\underline{X}}\underline{\theta} - \underline{y})$$

故 $L(w,b)$ 的最小值等同於

故

$$J(\theta) = (\underline{\underline{X}}\underline{\theta} - \underline{y})^T(\underline{\underline{X}}\underline{\theta} - \underline{y})$$

的最小值

$$\Rightarrow J(\theta) = (\underline{\underline{X}}\underline{\theta})^T(\underline{\underline{X}}\underline{\theta}) - (\underline{\underline{X}}\underline{\theta})^T\underline{y} - \underline{y}^T(\underline{\underline{X}}\underline{\theta}) + \underline{y}^T\underline{y}$$

$$= (\underline{\underline{X}}\underline{\theta})^T(\underline{\underline{X}}\underline{\theta}) - 2\underline{y}^T(\underline{\underline{X}}\underline{\theta}) + \underline{y}^T\underline{y}$$

$$\Rightarrow J(\underline{\theta}) = \underline{\theta}^T\underline{\underline{X}}^T\underline{\underline{X}}\underline{\theta} - 2\underline{y}^T\underline{\underline{X}}\underline{\theta} + \underline{y}^T\underline{y}$$

$$\Rightarrow \frac{\partial J(\theta)}{\partial \underline{\theta}} = \underline{\underline{X}}^T\underline{\underline{X}}\underline{\theta} + \underline{\underline{X}}^T\underline{\underline{X}}\underline{\theta} - 2\underline{\underline{X}}^T\underline{y}$$

$$= 2\underline{\underline{X}}^T\underline{\underline{X}}\underline{\theta} - 2\underline{\underline{X}}^T\underline{y} = 0$$

$$\Rightarrow 2\underline{\underline{X}}^T\underline{\underline{X}}\underline{\theta} - 2\underline{\underline{X}}^T\underline{y} = 0$$

$$\Rightarrow \underline{\theta} = (\underline{\underline{X}}^T\underline{\underline{X}})^{-1}\underline{\underline{X}}^T\underline{y}$$

⇒ Linear Regression Model:

$$= f(\underline{x}) = \underline{\theta}^T \cdot \begin{pmatrix} \underline{x} \\ 1 \end{pmatrix}$$ ☆

1-c.

$$L(\underline{w}, b) = \frac{1}{2N}\sum_{i=1}^{N}(y_i - (\underline{w}^T x_i + b))^2 + \frac{\lambda}{2}\|w\|^2$$

Set $\underline{\underline{X}} = \begin{bmatrix} -x_1- & 1 \\ -x_2- & \vdots \\ x_N & 1 \end{bmatrix}$ $\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_N \end{pmatrix}$

$$\underline{\theta} = \begin{pmatrix} \underline{w} \\ b \end{pmatrix}$$

minimize minimize

$$L(\underline{w}, b) = J(\underline{\theta})$$

$$= (\underline{y} - \underline{\underline{X}}\underline{\theta})^T(\underline{y} - X\theta) + \lambda\underline{\theta}^T\underline{\theta}$$

$$= \underline{\theta}^T\underline{\underline{X}}^T\underline{\underline{X}}\underline{\theta} - 2\underline{y}^T\underline{\underline{X}}\underline{\theta} + \underline{y}^T\underline{y} + \lambda\underline{\theta}^T\underline{\theta}$$

$$\frac{\partial J}{\partial \theta} = 2\underline{\underline{X}}^T\underline{\underline{X}}\underline{\theta} - 2\underline{y}^T\underline{\underline{X}} + 2\lambda\underline{\theta} = 0$$

$$\Rightarrow \underline{\theta} = (\underline{\underline{X}}^T\underline{\underline{X}} + \lambda\underline{\underline{I}})^{-1}\underline{\underline{X}}^T\underline{y}$$

⇒ $F(x) = \underline{\theta}^T X$ is the regression model ☆.