ML HW1 Report

學號：B06705024 系級：資管四 姓名：郭宇軒

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

	Private	Public
p = 9*15 + 1	6.78769	6.51321
p = 9*1 + 1	5.05615	4.91156

根據上表，我們可以發現，抽全部污染源當作 features 的 model 不管 private 和 public RMSE 都比 只抽 PM2.5 當作污染源的大。

我們可以推測，PM2.5 本身即為最重要的 feature，因此加上了其他的 feature 對 training model 來說可能會增加許多異常值反而讓 RMSE 上升，因此資料的預處理非常重要。

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。

1. 將 missing value 全部補上欄位平均值；並將 testing data中，所有為 0 的值視為遺失值，補上欄位的平均值作預測：

	Private	Public
p = 9*15 + 1	4.83374	4.99790
p = 9*1 + 1	4.77207	5.02830

可以看到進行完這次處理後， all feature 的 public 和 private 的 RMSE都顯著下降，而只有 PM2.5 的 model 則只有 private score 下降，因此適當的 data preprocess 後，其餘的 feature 其實都可以幫助我們進行預測。

2.加入高次項，加入 PM2.5 的平方項以及次方項：

	Private	Public
p = 9*15 + 1	5.00435	4.97564
p = 9*1 + 1	4.72236	4.92891

我們可以發現， all feature model 只有 public 的 RMSE下降，而只有 PM2.5 的 model 則public 和 private score 都下降。

這個處理是為了最後通過 public 的 strong baseline 後加入的，但可以發現只有public score會下降，但對於 private 來說反而會升高，可能是因為太高次項的 feature 已經會造成 overfitting了。因此最後在選擇 model 的時候應該考慮 validation 的結果而不是只相信 kaggle 的 public score啊。

3.(3%) Refer to math problem
https://hackmd.io/RFiu1FsYR5uQTrrpdxUvIw?view

1. (a).

$$L_{ssq}(w,b) = \frac{1}{2\times 5}\sum_{i=1}^{5}(y_i - (w^T x_i + b))^2 \qquad y = wx+b$$

$$\boxed{\begin{array}{l} w = \arg\min L(w,b) \\ \text{when } \dfrac{\partial L(w,b)}{\partial w} = 0 \end{array}} \Rightarrow w' = \frac{\sum S_{xx} S_{yy}}{\sum S_{xx}^2} = \frac{2\times 2.16 + 1\times 0.96 + 0 + 1\times 0.74 + 2\cdot 2.24}{2^2+1^2+0+1^2+2^2}$$

$\mu_x = 3 \quad \mu_y = 3.36$

$$= \frac{10.5}{10} = 1.05$$

$b = y - wx \quad \text{if } (\mu_x, \mu_y)$

$= 3.36 - 1.05\times 3 = 0.21$

$\Rightarrow (w,b) = (1.05, 0.21)$ *

1. (b.) We want to minimize $L_{ssq}(w,b)$

$$\Rightarrow \frac{\partial L_{ssq}(w,b)}{\partial w} = 0$$

$$\Rightarrow -2\cdot\frac{1}{2N}\sum_{i=1}^{N}(y_i - w^T x_i - b)(x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{N}(y_i - w^T x_i - b)(x_i) = 0$$

$$= \sum_{i=1}^{N}(y_i - w^T x_i - \bar{y} + w^T\bar{x})(x_i) = 0$$

$$= \sum_{i=1}^{N}(y_i x_i) - \sum_{i=1}^{N} w^T (x_i)^2 - \sum_{i=1}^{N}(\bar{y} x_i) + w^T\sum_{i=1}^{N}(\bar{x} x_i) = 0$$

$$\Rightarrow w^* = \frac{\sum y_i x_i - \sum \bar{y} x_i}{\sum (x_i)^2 - \sum(\bar{x} x_i)} = \frac{\sum x_i y_i - \bar{y}\sum x_i}{\sum x_i x_i - \bar{x}\sum x_i} = \frac{\sum x_i y_i - n\bar{y}\bar{x}}{n\bar{x}\bar{x} - \sum x_i x_i}$$

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

$$\Rightarrow \quad (w,b) = \left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} , \quad \bar{y} - w^T\bar{x} \right) \#$$

$$\because \Sigma wx = \Sigma \bar{y} - b\bar{x}$$

$$\Rightarrow \sum_{i=1}^{N}(y_i x_i) - \sum_{i=1}^{N}(\bar{y}-b\bar{x})x_i - \sum_{i=1}^{N}(b x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{N}(y_i x_i) - \sum_{i=1}^{N}(\bar{y}x_i) + \sum_{i=1}^{N}(b\bar{x}x_i) - \sum_{i=1}^{N}(b x_i) = 0$$

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

minimize: $\frac{\partial L_{reg}(w,b)}{\partial w} = 0$

$$\Rightarrow -2 \cdot \frac{1}{2N}\sum_{i=1}^{N}(y_i - w^T x_i - b)(x_i) + \lambda w = 0$$

$$\Rightarrow \sum_{i=1}^{N}(y_i - w^T x_i - b)(x_i) - \lambda w N = 0$$

$$\Rightarrow \Sigma y_i x_i - w^T\sum_{i=1}^{N} x_i x_i - \sum_{i=1}^{N} b x_i - \lambda w N = 0$$

$$= \Sigma y_i x_i - w^T\Sigma(x_i)^2 - \Sigma(\bar{y} - w^T\bar{x})x_i - \lambda w N = 0$$

$$\Rightarrow w^T\Sigma(x_i)^2 - \Sigma w^T x_i \bar{x}_i + w\lambda N = \Sigma y_i x_i - \Sigma\bar{y}x_i$$

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

$$\Rightarrow (w,b) = \left(\frac{\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{N}(x_i-\bar{x})^2 + \lambda N} , \bar{y} - w^T\bar{x} \right) \#$$

2. by 1.(b.) replace x_i by $(x_i + \eta_i)$

$\therefore$ minimized weight:

$$W_0 = E\left(\frac{\Sigma(x_i+\eta_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i+\eta_i-\bar{x})^2}\right)$$

and by 1.(c.)

$\therefore$ minimized weight for regularization function.

$$W_1 = E\left(\frac{\Sigma(y_i-\bar{y})(x_i-\bar{x})}{\Sigma(x_i-\bar{x})^2+\lambda N}\right)$$

$$E\left(\Sigma(x_i+\eta_i-\bar{x})(y_i-\bar{y})\right) = \Sigma\left(E(x_i+\eta_i-\bar{x})(y_i-\bar{y})\right)$$

$$= \Sigma\left(E(x_i y_i + \cancel{\eta_i y_i} - \bar{x}y_i - x_i\bar{y} - \cancel{\eta_i\bar{y}} + \bar{x}\bar{y})\right)$$

$$= \Sigma\left(E(x_i y_i - \bar{x}y_i - x_i\bar{y} + \bar{x}\bar{y})\right) = \Sigma E((x_i-\bar{x})(y_i-\bar{y}))$$

$$= W_1$$

$$E\left(\Sigma(x_i+\eta_i-\bar{x})^2\right) = \Sigma\left[E(x_i-\bar{x}+\eta_i)^2\right]$$

$$= \Sigma\left[E(x_i-\bar{x})^2 + 2E(x_i-\bar{x})(\eta_i) + E(\eta_i^2)\right]$$

$$= \sum_{i=1}^{N}\left(E(x_i-\bar{x})^2+\sigma^2\right)$$

$$= \sum_{i=1}^{N}(x_i-\bar{x})^2 + N\sigma^2 = W_1$$

$\Rightarrow W_0 = W_1$ #

3(a.)

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

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

$$= \quad N\cdot S_k - 2\sum_{i=1}^{N}\left(g_k x_i y_i\right) + N e_0$$

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

3(b.) 令 $\alpha_1 \cdots \alpha_k$ $L_{test}\left(\sum_{k=1}^{k}\alpha_k g_k\right)$ 為 H

对 α 微分：$\frac{dH}{d\alpha} =$

$$= \frac{1}{N}\sum_{i=1}^{N} 2\left(\sum_{k=1}^{k}\alpha_k g_k(x_i) - y_i\right)\sum_{k=1}^{k} g_k(x_i) = 0$$

$$\Rightarrow \quad \sum_{i=1}^{N}\sum_{k=1}^{k}\alpha_k g_k(x_i)\cdot\sum_{k=1}^{k} g_k(x_i) = \sum_{i=1}^{N}\sum_{k=1}^{k} g_k(x_i) y_i$$

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

$\Rightarrow$ 當 $\sum_{i=1}^{N}\sum_{k=1}^{k}\alpha_k g_k(x_i)\cdot\sum_{k=1}^{k} g_k(x_i) = \sum_{i=1}^{N}\frac{N(e_0 - e_k + S_k)}{2}$ 時.

$(\alpha_1, \alpha_2 \cdots\cdots \alpha_k)$ are optimal weight ! $\divideontimes$.