ML hw1 Report

學號：B05705001 系級： 資管四 姓名：黃意芹

請實做以下兩種不同feature的模型，回答第 (1) ~ (2) 題：

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

(1) 抽全部9小時內的污染源feature當作一次項(加bias)

RMSE

(2) 抽全部9小時內pm2.5的一次項當作feature(加bias)

RMSE	全部9小時內的污染源當作feature	全部9小時內pm2.5當作feature
private	5.65158	5.72749
public	6.02616	5.99585
average	5.83887	5.86167

由得到的RMSE可知，抽全部9小時內的污染源feature當作一次項的表現會比抽全部9小時內pm2.5的一次項當作feature更好，不過差距不是很大。原因可能是因為PM2.5跟其他污染源有關聯，所以增加更多feature對於預測PM2.5會有所幫助。不過因為PM2.5本身就算是最主要的特徵，且其他資料可能有摻雜許多雜亂、異常的值，所以增加其他特徵的進步幅度沒有非常大。

2. 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。

RMSE	全部9小時內的污染源當作feature且有清除掉異常pm2.5	全部9小時內的污染源當作feature沒有清除掉異常pm2.5
private	5.69725	5.65158
public	5.54239	6.02616
average	5.61982	5.83887

除了 NR皆設為0，還有把在字尾的其他符號去除只留數字，我還有做資料清理方法為：先把pm2.5的極端值挑出（pm2.5 < 2 or pm2.5 >= 100），之後算出pm2.5的平均和標準差，將平均+−3.6*標準差以外的training data丟掉。在predict時，若testing data有pm2.5的值在平均+−3.8*標準差以外，則取代成平均值進行預測。會

這麼做的原因在於，pm2.5本身為最重要的特徵，然而資料有一些異常值，所以先清除掉再訓練出的model會比較準確。而testing data因為沒辦法丟掉資料，所以就把異常值取代為平均值進行預測。

從上表RMSE可觀察到，有去除異常pm2.5的model表現明顯較好。這是因為訓練與預測時不會被異常值誤導，所以結果也會比較準確。

3. Refer to math problem

ML HW1

No:
Date: /

1.

1-(a)

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

$$w'=\arg\min_{w} L(w,b)$$

when $\frac{\partial L(w,b)}{\partial w}=0$

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

$$\bar{x}=\frac{\sum_{i=1}^{5}x_i}{5}=3$$

$$\bar{y}=\frac{\sum_{i=1}^{5}y_i}{5}=3.36$$

$$=\frac{(1.2-3.36)(1-3)+(2.4-3.36)(2-3)+(3-3.36)(3-3)+(4.1-3.36)(4-3)+(5.6-3.36)(5-3)}{2^2+1^2+0^2+1^2+2^2}$$

$$=\frac{(-2.16)\cdot(-2)+(-0.96)(-1)+0+(0.74)(1)+(2.24)(2)}{10}$$

$$=\frac{1}{10}[4.32+0.96+0.74+4.48]=\frac{10.5}{10}=1.05$$

$$b'=\bar{y}-w\bar{x}=3.36-1.05\times 3=3.36-3.15=0.21$$

No:

Date: /

1-(b) To find w, b that minimize $L(w,b)$

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

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

$$\Rightarrow \sum_{i=1}^{N} y_i x_i - \sum_{i=1}^{N} wx_i^2 - \sum_{i=1}^{n} bx_i = 0$$

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

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

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

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

$$= \frac{\sum_{i=1}^{N} xy - N\bar{x}\bar{y} - N\bar{x}\bar{y} + N\bar{x}\bar{y}}{\sum x_i^2 - N\bar{x}^2}$$

$$= \frac{\sum xy - \sum x_i\bar{y} - \sum \bar{x}y_i + \sum \bar{x}\bar{y}}{\sum x_i^2 - 2N\bar{x}^2 + N\bar{x}^2}$$

$$\therefore w = \frac{\sum(x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b = \bar{y} - \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}\bar{x}$$

ML HW1-2

No:
Date: /

1-(C)

$$L_2(W,b) = \frac{1}{2N}\sum_{i=1}^{N}\left(y_i-(W^Tx_i+b)\right)^2 + \frac{\lambda}{2}\|w\|^2$$

$$\frac{\partial\left(L_1(W,b)+\frac{\lambda}{2}|w|^2\right)}{\partial W} = \frac{-1}{N}\sum_{i=1}^{N}(y_i-b-Wx_i)x_i + \frac{\partial\frac{\lambda}{2}\|w\|^2}{\partial W} = 0$$

$$\Rightarrow -\frac{1}{N}\sum\left(y_ix_i-bx_i-Wx_i^2\right)+\lambda W=0$$

$$\Rightarrow -\left(\sum y_ix_i-\sum Wx_i^2-\sum bx_i\right)+\lambda N W=0$$

$$\Rightarrow -\sum y_ix_i+\sum\left(\bar{y}-W^T\bar{x}\right)x_i+\sum Wx_i^2+\lambda NW=0$$

$$\Rightarrow -\sum y_ix_i+\sum\bar{y}x_i-W^T\bar{x}\sum x_i+W^T\sum x_i^2+\lambda NW=0$$

$$\Rightarrow -\sum y_ix_i+\sum\bar{y}x_i+W^T\left(\sum x_i(x_i-\bar{x})+\lambda N\right)=0$$

$$\Rightarrow W=\frac{\sum y_ix_i-\sum\bar{y}x_i}{\sum x_i(x_i-\bar{x})+\lambda N}=\frac{\sum(y_i-\bar{y})(x_i-\bar{x})}{\sum(x_i-\bar{x})^2+\lambda N}$$

$$b=\bar{y}-W^T\bar{x}$$

No:

Date: /

2、由1.(2)把 x_i 代换成 $x_i+\eta_i$

→ optimal weight: $W_0 = E\left(\frac{\sum (x_i+\eta_i-\bar{x})(y_i-\bar{y})}{(\sum (x_i+\eta_i-\bar{x})^2}\right)$

由1.(3)知，加上 regularization 后的 optimal weight

$$W_1 = \frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sum (x_i-\bar{x})^2 + \sigma^2 N}$$

若 $W_0 = W_1$，则可证得 minimizing 两者 Loss 后的 W, b 相同

分子：$E\left(\sum_{i=1}^{N}(x_i+\eta_i-\bar{x})(y_i-\bar{y})\right) = \sum_{i=1}^{N}\left[E(x_i+\eta_i-\bar{x})(y_i-\bar{y})\right]$

$$= \sum_{i=1}^{N}\left(E(x_iy_i - x_i\bar{y} + \eta_i y_i - \eta_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y})\right)$$

$\because E(\eta_i)=0$

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

$$= \sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})$$ 同 W_1 的分子

分母：$E\left(\sum_{i=1}^{N}(x_i+\eta_i-\bar{x})^2\right) = \sum_{i=1}^{N}\left[E(x_i-\bar{x}+\eta_i)^2\right]$

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

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

同 W_1 的分母

故两者相等 ✓

No:

Date: /

3.

(a) Let $A = \sum_{i=1}^{N} g_k(x_i) y_i$

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

$$= \frac{1}{N}\left(NS_k - 2\cdot A + Ne_0\right) \Rightarrow Ne_k = NS_k - 2A + Ne_0$$

$$\Rightarrow A = \frac{N(S_k - e_k + e_0)}{2} \quad //$$

(b) Let $F = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K} a_k g_k(x_i) - y_i\right)^2$

$$\frac{\partial F}{\partial a} = \frac{1}{N}\sum_{i=1}^{N}\cdot\left[2\cdot\left(\sum_{k=1}^{K} a_k g_k(x_i) - y_i\right)\cdot\sum_{k=1}^{K} g_k(x_i)\right] = 0$$

$$\rightarrow \sum_{i=1}^{N}\left[\left(\sum_{k=1}^{K} a_k g_k(x_i)\right)\cdot\sum_{i=1}^{N} g_k(x_i) - \sum_{k=1}^{K} g_k(x_i)\cdot y_i\right] = 0$$

$$\rightarrow \sum_{k=1}^{K}\left[\left(\sum_{i=1}^{N} a_k g_k(x_i)\right)\cdot\sum_{i=1}^{N} g_k(x_i) - \sum_{i=1}^{N} g_k(x_i)\cdot y_i\right] = 0$$

$$\rightarrow \sum_{k=1}^{K}\left(a\cdot NS_k - \frac{N}{2}(S_k - e_k + e_0)\right) = 0$$

$$\rightarrow a = \frac{\frac{N}{2}(S_k - e_k + e_0)}{NS_k} = \frac{S_k - e_k + e_0}{2S_k}$$