

學號：B03701221 系級： 工管三 姓名：王逸庭

請實做以下兩種不同 feature 的模型，回答第 (1) ~ (3) 題：

1. 抽全部 9 小時內的污染源 feature 的一次項(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，其他的數值不要做任何更動
- b. 所有 advanced 的 gradient descent 技術(如: adam, adagrad 等) 都是可以用的

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

RMSE(9hr):

All variable: 6.540106507

pm2.5: 6.576447856

拿全部的污染源當 feature 的表現會比只抽 PM2.5 表現較好，但差距卻不大。  
代表 PM2.5 和結果有高度相關，對結果有極高的預測力。但其他的變因仍有其他擁有預測力的 feature，故抽全部的誤差會比僅抽 PM2.5 的誤差來得小。

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

<b>allvariable_5.csv</b> just now by <a href="#">kelly wang</a> <a href="#">add submission details</a>	<b>5.40813</b>	<b>7.63484</b>
<b>allvariable.csv</b> 6 minutes ago by <a href="#">kelly wang</a> <a href="#">add submission details</a>	<b>5.53788</b>	<b>7.40796</b>
<b>pm2.5_5.csv</b> just now by <a href="#">kelly wang</a> <a href="#">add submission details</a>	<b>5.77478</b>	<b>7.52774</b>
<b>pm2.5.csv</b> a minute ago by <a href="#">kelly wang</a> <a href="#">add submission details</a>	<b>5.61220</b>	<b>7.41637</b>

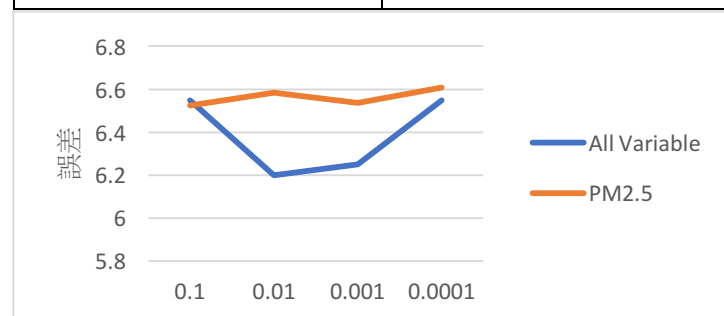
	RMSE(9hr)	RMSE(5hr)
All variable	6.540106507	6.615839022
PM2.5	6.576447856	6.708761195

從上述表格，我們取 9 小時來做訓練的效果比 5 小時好，這代表多的 4 個小時的 feature 仍有對結果有貢獻度；但至多也只有減少 0.3 的誤差，這代表主要預測的貢獻度仍集中後面的 5 個小時。

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

<b>allvariable_reg_4.csv</b> 23 minutes ago by <a href="#">kelly wang</a> <a href="#">add submission details</a>	<b>5.63623</b>	<b>7.34683</b>
<b>allvariable_reg_3.csv</b> 23 minutes ago by <a href="#">kelly wang</a> <a href="#">add submission details</a>	<b>5.73312</b>	<b>6.72948</b>
<b>allvariable_reg_2.csv</b> 23 minutes ago by <a href="#">kelly wang</a> <a href="#">add submission details</a>	<b>5.51478</b>	<b>6.81252</b>
<b>allvariable_reg_1.csv</b> 35 minutes ago by <a href="#">kelly wang</a> <a href="#">add submission details</a>	<b>5.76454</b>	<b>7.17193</b>
<b>pm2.5_reg_4.csv</b> 2 minutes ago by <a href="#">kelly wang</a> <a href="#">add submission details</a>	<b>5.60483</b>	<b>7.47597</b>
<b>pm2.5_reg_3.csv</b> 3 minutes ago by <a href="#">kelly wang</a> <a href="#">add submission details</a>	<b>5.58746</b>	<b>7.36209</b>
<b>pm2.5_reg_2.csv</b> 3 minutes ago by <a href="#">kelly wang</a> <a href="#">add submission details</a>	<b>5.64123</b>	<b>7.40714</b>
<b>pm2.5_reg_1.csv</b> 3 minutes ago by <a href="#">kelly wang</a> <a href="#">add submission details</a>	<b>5.56488</b>	<b>7.35690</b>

$\lambda$	All Variable	PM2.5
0.1	6.547633147	6.522724393
0.01	6.197710351	6.583661553
0.001	6.251182528	6.535291823
0.0001	6.547633147	6.606975358



從上圖可以發現 Regularization 對於變數較多影響較多，同時 $\lambda$ 太高或太小都無法優化此模型。

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  $x^i$ ，其標註(label)為一存量  $y^i$ ，模型參數為一向量  $w$  (此處忽略偏權值  $b$ )，則線性回歸的損失函數(loss function)為  $\sum_{i=1}^N (y^i - x^i w)^2$ 。若將所有訓練資料的特徵值以矩陣  $X = [x^1 \ x^2 \ \dots \ x^N]^T$  表示，所有訓練資料的標註以向量  $y = [y^1 \ y^2 \ \dots \ y^N]^T$  表示，請問如何以  $X$  和  $y$  表示可以最小化損失函數的向量  $w$ ？請寫下算式並選出正確答案。(其中  $X^T X$  為 invertible)

- a.  $(X^T X) X^T y$
- b.  $(X^T X)^{-1} X^T y$
- c.  $(X^T X)^{-1} X^T y$
- d.  $(X^T X)^2 X^T y$

Ans: **C**

$$E = ||y - Xw||^2$$

E: error vector