學號：B04901066 系級：電機三 姓名：洪國喨

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

model	public score	private score	public+private	((public^2+private^2/2)^0.5
(1)所有污染源	7.46215	5.53190	12.99405	6.56832
(2)僅pm2.5	7.44013	5.62719	13.06732	6.59624

兩者結果其實十分接近，所有污染源model中，其它feature的weight因為只有一次項，其值非常小，所以對於預測結果之影響極小。在public score中，僅pm2.5的model甚至誤差較低。

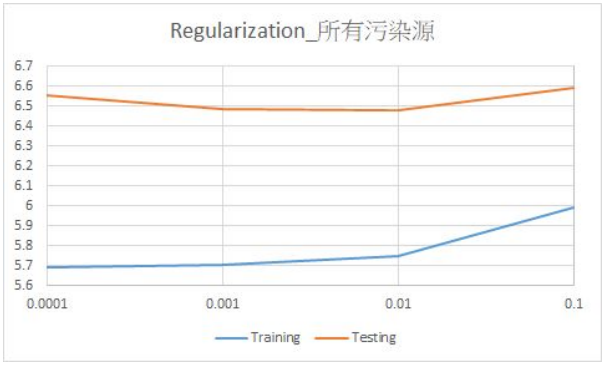
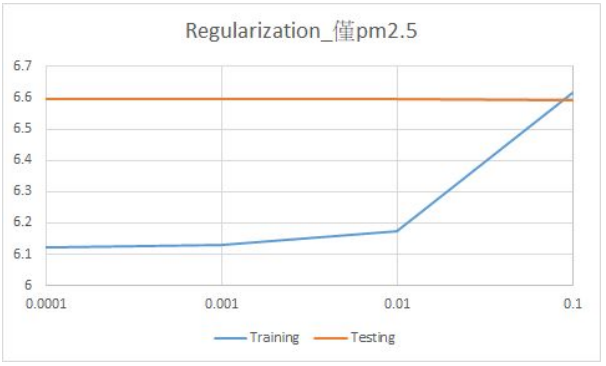
2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

model	public score	private score	public+private	((public^2+private^2/2)^0.5	變化
(1)所有污染源	7.65919	5.44073	13.09992	6.64322	private變低其餘變高
(2)僅pm2.5	7.57904	5.79187	13.37091	6.74491	score皆變高

可知觀測時間更長一些，對於預測準確度提升有幫助，所有污染源model的結果大體上仍較僅pm2.5的model好。所有污染源model中private score在5小時反而變低，推測是因為原始數據中有些pm2.5值是-1(即觀測失敗)，9小時比較容易取到這些資訊，所以5小時反而較準一些。

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

所有污染源					僅pm2.5				
λ	Training	public	private	Testing	λ	Training	public	private	Testing
0.1	5.994212	7.62186	5.37167	6.593466	0.1	6.615253	7.44012	5.6272	6.59624
0.01	5.747677	7.48247	5.28707	6.478444	0.01	6.172248	7.44013	5.62719	6.596241
0.001	5.703113	7.4656	5.3348	6.488269	0.001	6.127944	7.44013	5.62719	6.596241
0.0001	5.691462	7.4585	5.50308	6.554125	0.0001	6.123514	7.44013	5.62719	6.596241

	
(1)所有污染源	(2)僅pm2.5

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^n，其標註(label)為一存量 y^n，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 $\sum_{n=1}^{N}(y^n - x^n \cdot w)^2$ 。若將所有訓練資料的特徵值以矩陣 X = [x^1 x^2 … x^N]^T 表示，所有訓練資料的標註以向量 y = [y^1 y^2 … 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。

答案：(c) $(X^TX)^{-1}X^Ty$

算式：$\sum_{n=1}^{N}(y^n - x^n \cdot w)^2 = \sum_{n=1}^{N}(y^n)^2 - 2\sum_{n=1}^{N}(y^n x^n)w + \sum_{n=1}^{N}(x^n)^2(w)^2$

$= \sum_{n=1}^{N}(x^n)^2((w)^2 - 2\cdot\frac{\sum_{n=1}^{N}(y^n x^n)}{\sum_{n=1}^{N}(x^n)^2} + (\frac{\sum_{n=1}^{N}(y^n x^n)}{\sum_{n=1}^{N}(x^n)^2})^2) + \sum_{n=1}^{N}(y^n)^2 - \cdot\frac{\left(\sum_{n=1}^{N}(y^n x^n)\right)^2}{\sum_{n=1}^{N}(x^n)^2}$

$= \sum_{n=1}^{N}(x^n)^2 \cdot (w - \frac{\sum_{n=1}^{N}(y^n x^n)}{\sum_{n=1}^{N}(x^n)^2}) + \sum_{n=1}^{N}(y^n)^2 - \frac{\left(\sum_{n=1}^{N}(y^n x^n)\right)^2}{\sum_{n=1}^{N}(x^n)^2}$

故當 $w = \frac{\sum_{n=1}^{N}(y^n x^n)}{\sum_{n=1}^{N}(x^n)^2} = (X^TX)^{-1}X^Ty$ 時，

loss function有最小值 $\sum_{n=1}^{N}(y^n)^2 - \frac{\left(\sum_{n=1}^{N}(y^n x^n)\right)^2}{\sum_{n=1}^{N}(x^n)^2} = y^Ty-(X^TX)^{-1}(X^Ty)^2$