

學號：R05942017 系級：電信碩二 姓名：黃梓鳴

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

All feature:Public: 7.53216Private: 5.32099 總分=12.85135

只有 PM2.5: Public: 7.74276 Private: 5.43349 總分=13.17165

a).發現在 training 過程中，用所有 feature 下去訓練的 RMSE 值會從比較大的值開始緩慢收斂(如 15)，只用 PM2.5 的值會從比較小的值開始緩慢收斂(如 11)

b)在 training 時，若 training 的 iteration 次數較少(如 200)，則只用 PM2.5 的 feature 其 RMSE 會較高

c)發現 iteration 次數一高(如 10000)，只用 PM2.5 的 RMSE 收斂速度會降很快，到後面幾乎不會在減少，而用所有 Feature 的仍會緩慢下降

d)為了確保有完整的 fit 到，我把 iteration 設 5 萬次下去跑，結果看來是使用所有 feature 的 RMSE 較小，但是比只用 PM2.5 的小一點點兒以，資料量卻大 18 倍，有點不划算

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

All feature:Public:7.78113 Private:5.30711 總分=13.08824

只有 PM2.5: Public:8.08528 Private:5.64844 總分=13.73372

也是 iteration 五萬次

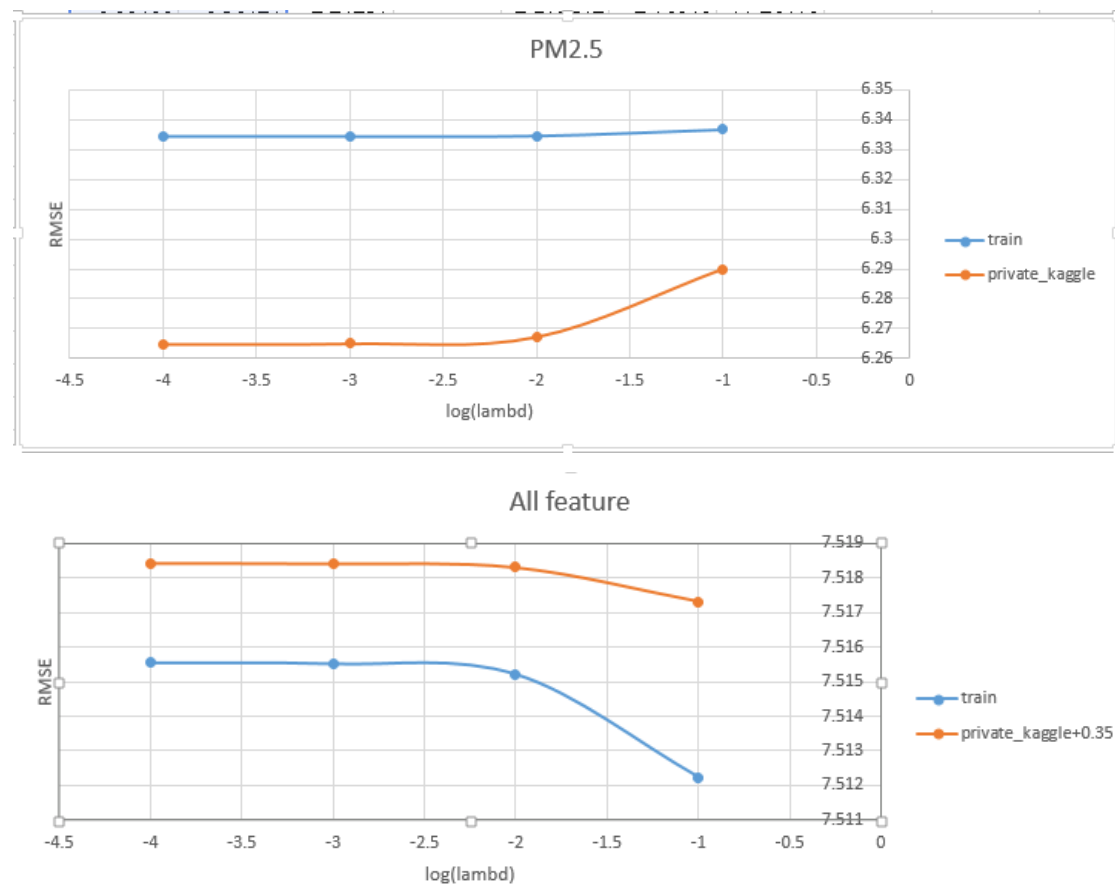
a) 發現若只用前五小時的話，使用全部 feature 的 RMSE 比只用 PM2.5 的 RMSE 相對於取九小時的好一點(九小時的總分相差 0.3203 五小時的相差 0.64548)

b) 取前五小時跟前九小時的差異很小，代表後四小時的參數影響非常小

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

隨著調整 lambd，RMSE 的確有變化，但是變化非常的小，為了讓圖更好的呈現 RMSE 隨著 lambd 的變化，我把 lambd 全部取 log，且取 private 的資料(因為 private 的 RMSE 值離 train 的比較近)，可以放大比例尺看出變化。

而取所有 feature 的 private 值離 train 的又還有點距離，於是把 private 的值通通 +0.35，讓兩 private 的值跟 train 很近，較可以看出變化



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 $n=1N \sum y^n - x^n w^2$ 。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)

- $(X^T X) X^T y$
- $(X^T X)^{-1} X^T y$
- $(X^T X)^{-1} X^T y$

d. $(X^T X)^{-1} X^T y$

$$y = y_1 + y_2$$

$$y_1 = W W^T y$$

$$y_2 = (I - W W^T) y$$

$$\|Xw - y\|^2 = \|Xw - y_1 + (-y_2)\|^2 = \|Xw - y_1\|^2 + \|y_2\|^2$$

故若且唯若 w 是 $Xw = y_1 = W W^T y$ 解時， w 即為最小二乘解，

$$\text{即 } w = X^+ y = (X^H X)^+ X^H y, \text{ 答案為 } \angle$$