

學號：B04902092 系級：資工三 姓名：張均銘

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.

全部污染源 feature 一次項 RMSE: (iterate : 10000)

public+private: 13.33409

pm2.5 feature 一次項 RMSE: (iterate:10000)

public+private : 13.06733

差別：pm2.5 直接 train 會比較高其實也蠻直觀的，畢竟 pm2.5 的數值本來就該跟前幾個小時 pm2.5 的值最有關係，而真實情況我也相信普遍第 10 個小時 pm2.5 的值不會跟第九小時的差太多。畢竟大範圍很難瞬間改變他的值。那至於全部污染源的 RMSE 比較低我認為是因為許多氣體與 pm2.5 沒有直接關係，所以 train 時會增加誤差，例如某氣體假設 NO 好了，他這一小時跟下一小時此氣體變化很大也不代表 pm2.5 就會變化大，但是拿進去 train 的話就代表說 NO 的變化 pm2.5 也會受影響，因此把這些對 pm2.5 影響不大的變因拿去 train 就會造成誤差。

2.

全部污染源 feature 一次項 RMSE(iterate 10000):

public+private : 42.52437

pm2.5 feature 一次項 RMSE(iterate 10000):

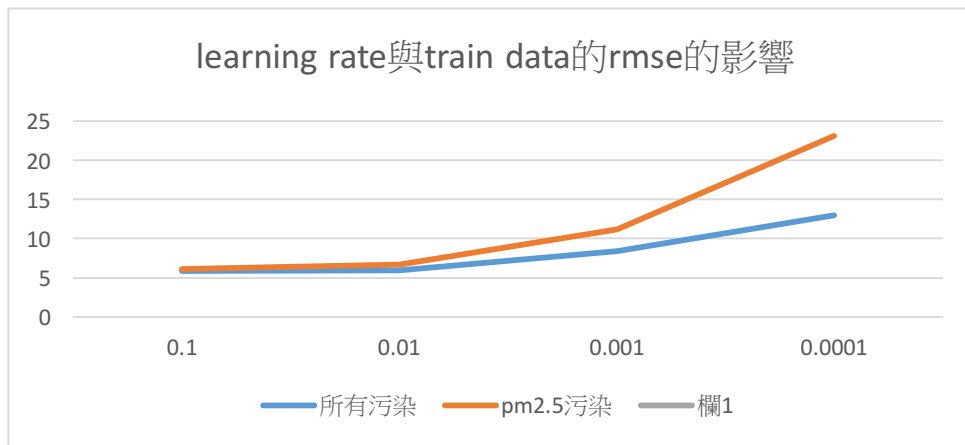
public+private : 39.30038

與上述差別：首先全部資料拿去 train 誤差還是比 pm2.5 大，合理。那因為實際上 pm2.5 第 10 個小時的資料可能與前 9hr 甚至更久的資料都有關係，因此如果只拿前 5hr 下去 train 的話，雖然點變多，但是每個 sample 參數變少，我覺得就有點像 bias 變大的感覺，點的分佈更沒有規則，所以會不準。

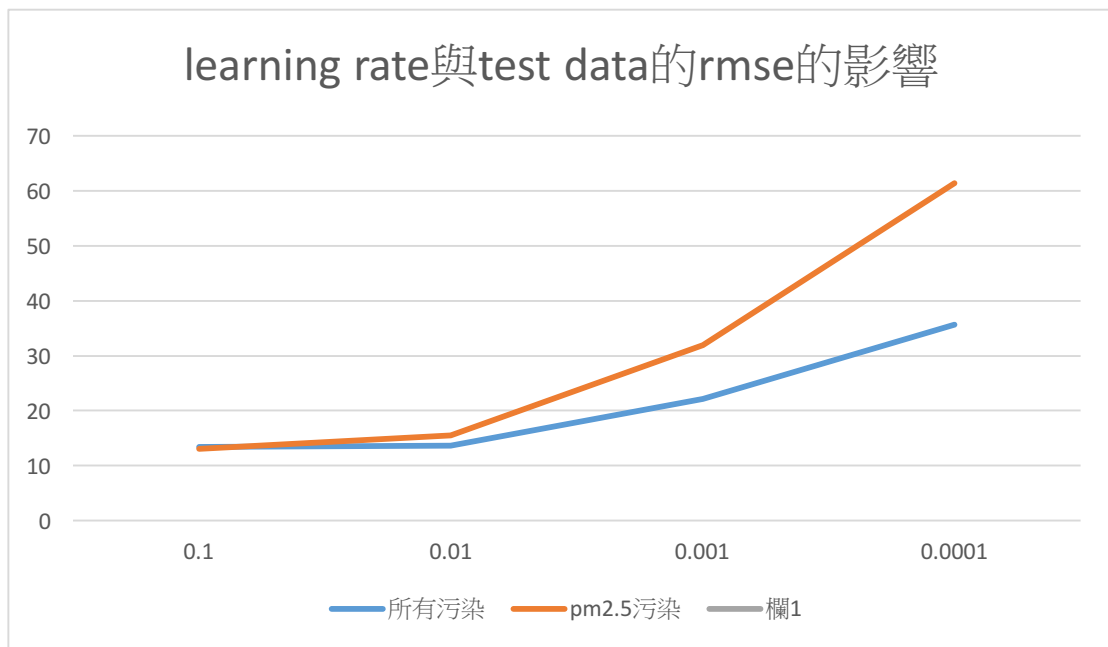
3.

train: iteration: 10000 次

此圖是 train 出來的 model 輸入 train 的 data 得到的 rmse 的值



下圖是不同 learning rate 丟上 kaggle 後得到的 public rmse + private rmse 的到的值



4. (c)

Loss function = $y - Xw$

Loss function 可以當成某種常數矩陣，或者說假設 train 到極致 Loss function 的值=0 時，等時可以改寫成 $y = Xw$ 故 $w = y * x^{-1} = (x^T x)^{-1} * x^T * y$