

抽全部 9 小時內的污染源 feature 的一次項 : 9×18

抽全部 9 小時內 pm2.5 的一次項當作 feature : 9×1

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

| | Public | Private |
|---------------|--------|---------|
| 9×18 | 6.82 | 6.43 |
| 9×1 | 6.57 | 6.28 |

我將所有 feature normalize 後, 使用 SGD 發現 9×1 的表現不論在 Public 或 Private 都比 9×18 的還要好, 這可能是因為 9×18 有許多不相干的參數如風向或 RH...等等影響訓練的結果, 這可以再 training data 本身的 loss 大小看就的出差別。

另外, 我還發現 Private 的分數普遍都比 Public 高, 這個問題我有仔細想過, 也有在社團上討論這個問題, 我認為可能跟取哪 120 筆資料當 Public 有很大的關係。

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

| | Public | Private |
|--------------------------|--------|---------|
| $9 \text{ hr} \times 18$ | 6.82 | 6.43 |
| $5 \text{ hr} \times 18$ | 6.95 | 6.64 |

$N \times 18$ 的例子中, 9hr 的表現不論在 Public 或 Private 都比 5hr 的好, 可能是因為 5hr 的參數過少造成 Underfitting。

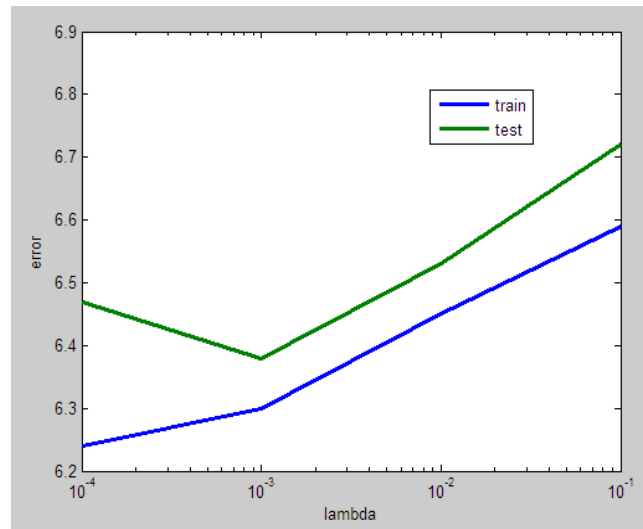
| | Public | Private |
|-------------------------|--------|---------|
| $9 \text{ hr} \times 1$ | 6.57 | 6.28 |
| $5 \text{ hr} \times 1$ | 7.03 | 6.63 |

$N \times 1$ 的例子中, 9hr 的表現不論在 Public 或 Private 也都比 5hr 的好, 而且 5hr Underfitting 的趨勢好像比在 $5 \text{ hr} \times 18$ 的更嚴重, 可能代表 9×1 裡 9hr 的每個參數可能都是非常必要的。

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

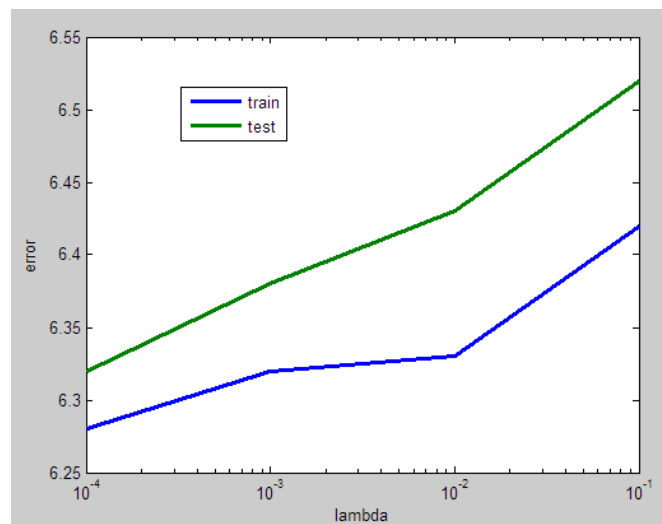
9×18

| λ | Train | Test |
|-----------|-------|------|
| 0.1 | 6.59 | 6.72 |
| 0.01 | 6.45 | 6.53 |
| 0.001 | 6.30 | 6.38 |
| 0.0001 | 6.24 | 6.47 |



9×1

| λ | Train | Test |
|-----------|-------|------|
| 0.1 | 6.42 | 6.52 |
| 0.01 | 6.33 | 6.43 |
| 0.001 | 6.32 | 6.38 |
| 0.0001 | 6.28 | 6.32 |



不論是 9×1 或 9×18 ，train error 都比 test error 小，而且 train error 都會隨 lambda 增加而增加。然而，在 9×18 的例子裡 $\lambda = 0.001$ 似乎可以有效地降低 test error。不過在 9×1 的例子裡就比較沒有這樣的現象。

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 $X = [x^1 \ x^2 \ \dots \ x^N]^T$ 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。

答案應該為 (C)

對於一組輸入 X 得到一組輸出 y ，其最小平方法得出的損失函數的向量 w 可以表示成以下問題： $\min_w |Xw - y|$

根據正交原則，最小平方解 X 滿足正規方程 (normal equation)

符合以下關係： $X^T X w = X^T y$

若 $X^T X$ 可逆，則 $w = (X^T X)^{-1} X^T y$ ，故答案為 (C)

Note :

我在 kaggle best (public = 5.467/ private = 5.50)用了以下方法：

- Feature normalize
- SGD
- 選用 O_3 、 SO_2 、 $PM2.5$ 的一次和二次項
- 1 Layer FC NN (手寫的 back propagation)
- ReLu activation function / 5 hidden layers
- Training test data for weights adjustment (1~8hr : input / 9hr : target)

以上方法都有在 kaggle 上驗證後都是必要的步驟，另外我也有寫 2 Layer FC NN，不過發現 performance 沒有更好所以就沒有繼續寫下去了。