

學號：R06922093 系級：資工碩一 姓名：陳禹齊

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
All feature	10.75235	6.96116
PM 2.5	7.40983	5.92390

討論：只使用 PM 2.5 一次項當作 feature 的 RMSE 比較低，可能是因為取所有污染源的 feature，考慮太多與 PM 2.5 不相關的因素，造成為了符合那些無關的污染源使得真正 PM 2.5 的誤差值變大。

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

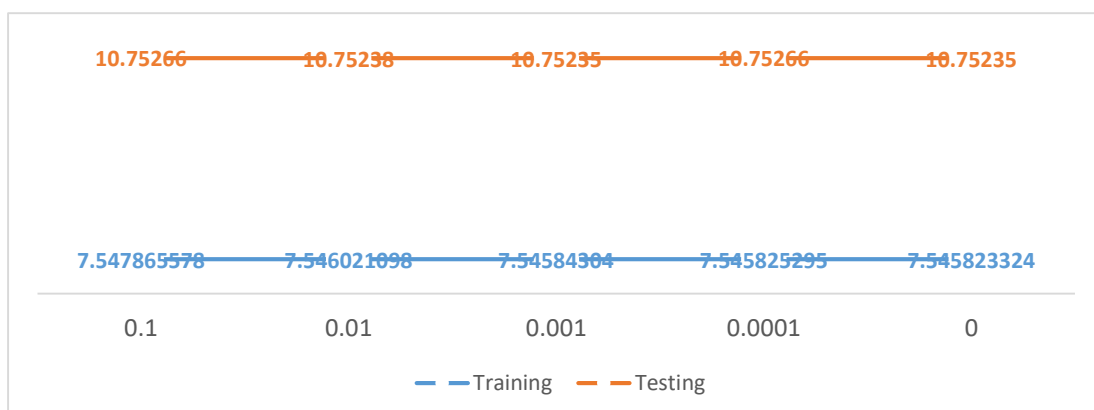
All feature		
	Training	public
抽 9 小時	7.52872	10.75235
抽 5 小時	6.62498	9.45220

PM 2.5		
	Training	public
抽 9 小時	7.52872	10.75235
抽 5 小時	6.62498	9.45220

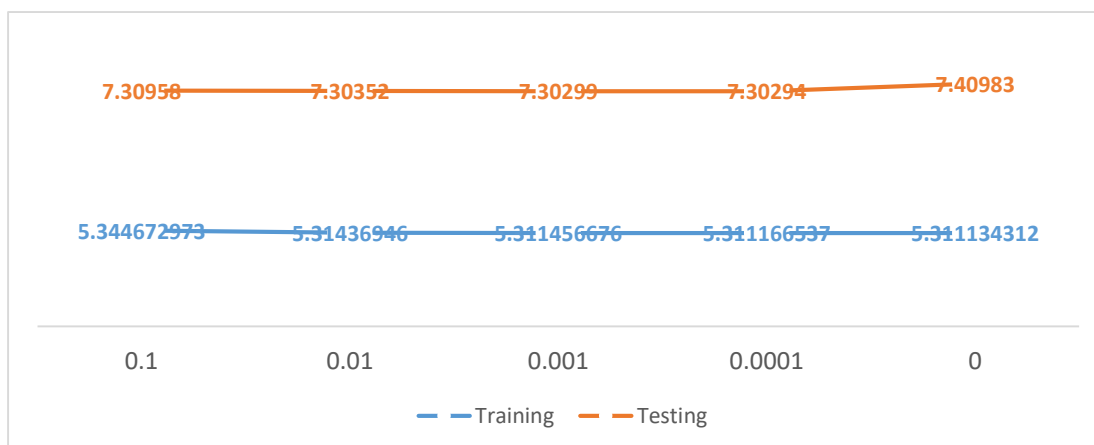
討論：取全部污染源當 feature,因為一開始的 feature 太多,所以只取比較接近的 5 小時來訓練可以讓其他比較沒那麼相關的 feature 影響不會那麼大,然後只取 PM 2.5 因為本來就跟 PM 2.5 相關,所以取越多小時來訓練可以獲得比較好的結果。

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

- i. 抽全部 9 小時內的污染源 feature 的一次項(加 bias):



ii. 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias):



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 $n=1Nyn-xnw^2$ 。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)

- $(X^T X) X^T y$
- $(X^T X)^0 X^T y$
- $(X^T X)^{-1} X^T y$
- $(X^T X)^2 X^T y$

我們的目標是要求出 $X * w = y$ 的最小平方解,

最小平方解發生時, $e = y - Xw$ 有最小長度的平方即 $\min \|y - Xw\|^2$, 從行空間

$CS(X)$ 中取一向量 p 和 $e=y-p$ 正交, p 也就是 y 至 X 的行空間的正交投影。令 P 表示正交投影至行空間 $CS(X)$ 的變換矩陣 $p=Py$, 由正交投影矩陣的性質可得 $Py=P(Py)=P^2y=p$,所以 $P(y-p)=p-p=0$ 。因此 $e=y-p=y-Xw$ 的正交捕餘空間,可得 $X^T e=X^T(y-Xw)=0e$, 則 $X^T Xw=X^T y$ 所求為

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