

## Machine Learning HW1 report

B04507009 電機三 何吉瑞

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

	public	private
所有汙染源	7.46222	5.52818
只有 pm2.5	7.44013	5.62719

兩者結果相近，原因可能是其他有利預測參數(如 pm10)和沒有適當轉換的參數(如角度沒有取成水平垂直分量)相互抵銷

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

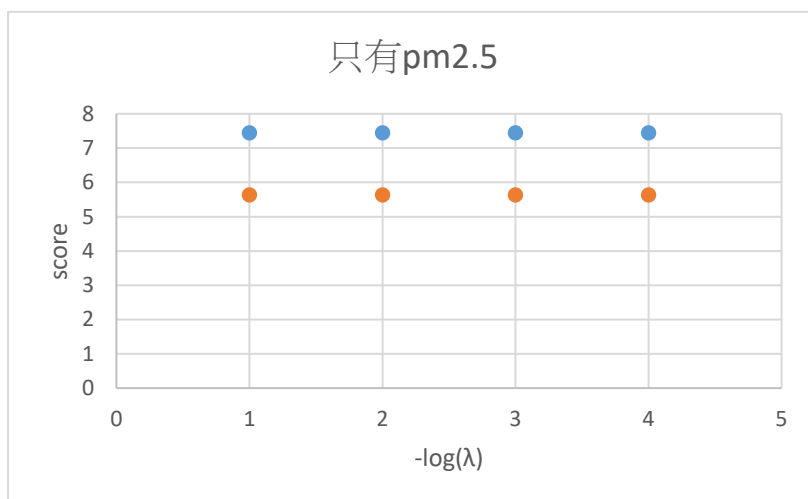
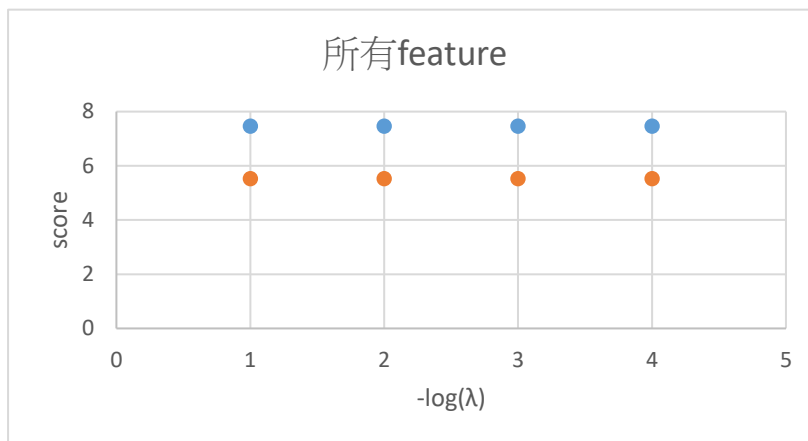
	public	private
所有汙染源 9 小時	7.46222	5.52818
所有汙染源 5 小時	7.65918	5.44081
只有 pm2.5 9 小時	7.44013	5.62719
只有 pm2.5 5 小時	7.57904	5.79187

9 小時的分析結果略優於 5 小時，所有汙染源的因素相對複雜，在 pm2.5 的 case 中比較能看出 9 小時表現較好

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

		public	private
所有污染源	$\lambda=0.1$	7.46268	5.52570
	$\lambda=0.01$	7.46227	5.52793
	$\lambda=0.001$	7.46222	5.52815
	$\lambda=0.0001$	7.46222	5.52817
只有 pm2.5	$\lambda=0.1$	7.44070	5.62750
	$\lambda=0.01$	7.44018	5.62722
	$\lambda=0.001$	7.44013	5.62719
	$\lambda=0.0001$	7.44013	5.62719

大致而言差距不大



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  $\sum_{n=1}^N (y^n - x^n w)^2$ 。若將所有訓練資料的特徵值以矩陣  $X = [x^1 \ x^2 \ \cdots \ x^N]^T$  表示，所有訓練資料的標

註以向量  $y = [y^1 \ y^2 \ \cdots \ y^n]^T$  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)

- a.  $(X^T X)Xy$
- b.  $(X^T X)^{-1}Xy$
- c.  $(X^T X)^{-1}X^T y$
- d.  $(X^T X)^{-1}X^T y$

$$e = (y - wX)$$

$$e^T e = (y - wX)^T (y - wX)$$

$$\frac{\partial e^T e}{\partial w} = -2X^T (y - wX) = 2(X^T Xw - X^T y)$$

令偏微分為 0 求最小值

$$X^T Xw - X^T y = 0$$

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

答案為(c)