

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

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

RMSE = 6.7699244452

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

RMSE = 6.59624690677

Feature1: 抽全部的一次項因為 model 可能會有 overfitting 的問題，所以 RMSE 較大

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

- (1) 抽全部 5 小時內的污染源 feature 的一次項(加 bias):

RMSE = 6.66311131487

- (2) 抽全部 5 小時內 pm2.5 的一次項當作 feature(加 bias)

RMSE = 6.74491674708

Feature1 的 error 變低但 feature2 的 error 變高，所以只抽 5 小時不一定會讓 model 變好

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

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

	Train	test
$\lambda=0.1$	RMSE = 5.878287	6.76992444415
$\lambda=0.01$	RMSE = 5.878285	6.76992444508
$\lambda=0.001$	RMSE = 5.878284	6.76992444518
$\lambda=0.0001$	RMSE = 5.878284	6.76992444519

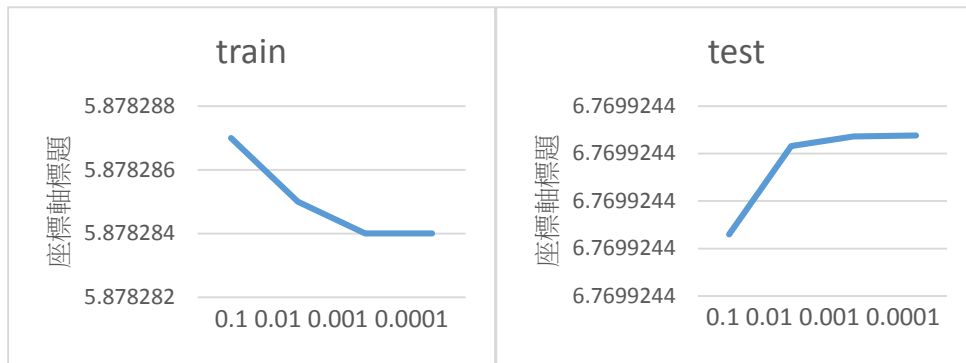
- (2) 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias):

	Train	test
$\lambda=0.1$	RMSE = 6.123029	6.59624640053
$\lambda=0.01$	RMSE = 6.123022	6.5962468561
$\lambda=0.001$	RMSE = 6.123022	6.59624690169

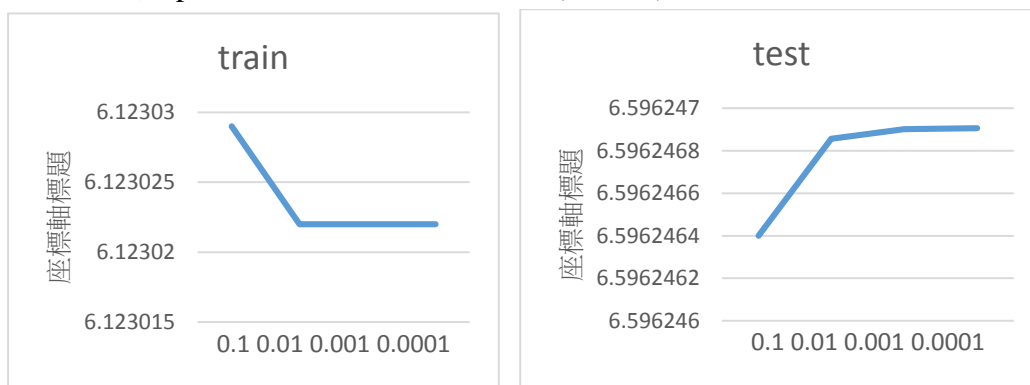
$\lambda=0.0001$  RMSE = 6.123022

6.59624690626

抽全部 9 小時內的污染源 feature 的一次項(加 bias):



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  $n=1N \sum (y^n - x^n w)^2$ 。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)

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

將 gradient 寫成向量形式  $= (-2) * X^T (y - Xw)$

當 gradient(一階微分) = 0 時，存在極值，又 Loss function 是一個 convex 的函數，所以 gradient = 0 時， $w$  的值帶入 Loss function 即為 Loss function 的最小值。

$$\text{Gradient} = (-2) * X^T (y - Xw) = 0$$

$$\Rightarrow X^T y - X^T X w = 0 \Rightarrow X^T y = X^T X w \Rightarrow (X^T X)^{-1} X^T y = w$$

$\Rightarrow$  答案 = c #