

學號：B03901149 系級：電機四 姓名：陳咸嘉

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. 記錄誤差值 (RMSE)(根據 kaggle public+private 分數)，討論兩種 feature 的影響 (Repeat 10000 次)

	抽 9 小時內全部污染源	抽 9 小時內 PM2.5
RMSE	9.45786	7.42637

討論：

推測因為並非所有空氣指標都與 PM2.5 有相關性，因此取全部的參數會造成結果偏差。因此取用與 PM2.5 相關係數較高者訓練會有較好的結果。

2. 將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

(Repeat 10000 次)

	抽 5 小時內全部污染源	抽 5 小時內 PM2.5
RMSE	8.65809	7.57924

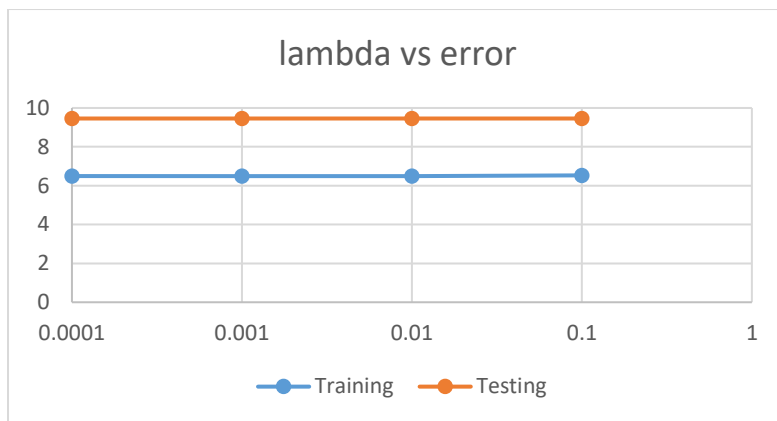
討論：

推測因為 9 小時會造成資料 overfitting，用 5 小時比較不會有這種情形發生。

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

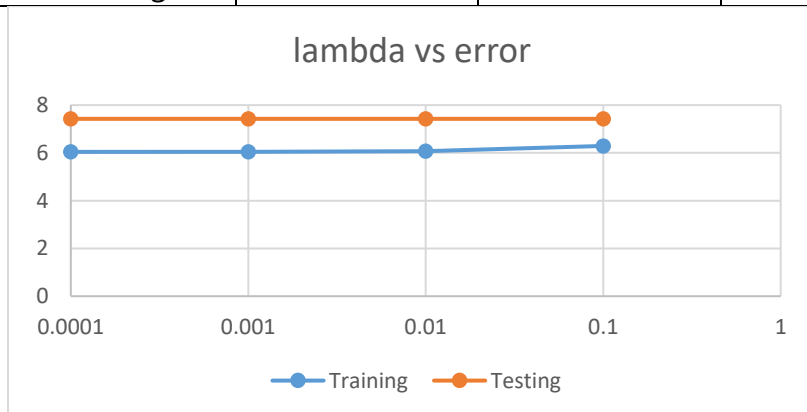
抽 9 小時內全部污染源：(Repeat 10000 次)

λ 值	0.1	0.01	0.001	0.0001
Training	6.53091	6.49802	6.49510	6.49478
Testing	9.45786	9.45786	9.45786	9.45786



抽 9 小時內 PM2.5 : (Repeat 10000 次)

λ值	0.1	0.01	0.001	0.0001
Training	6.29108	6.06657	6.04411	6.04186
Testing	7.42637	7.42637	7.42637	7.42637



Testing 的誤差幾乎不變，training 的變動也很小，可能是因為 Lambda 在 Loss function 的貢獻遠小於 error 的貢獻。

4.

答案為(c)，推導過程如下：

$$\begin{aligned}
 L &= \sum_{n=1}^N (y^n - x^n w)^2 = \|y - Xw\|^2 = (y - Xw)^T (y - Xw) \\
 L &= (y^T - (Xw)^T)(y - Xw) = y^T y - (Xw)^T y - y^T (Xw) + (Xw)^T (Xw) \\
 L &= y^T y - 2(Xw)^T y + (Xw)^T (Xw) \\
 L &= y^T y - 2w^T X^T y + w^T X^T X w \\
 \frac{\partial L}{\partial w} &= -2X^T y + 2X^T X w
 \end{aligned}$$

設偏微分後的值為 0

$$\begin{aligned}
 X^T Y &= X^T X w \\
 w &= (X^T X)^{-1} X^T y
 \end{aligned}$$