

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)

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

	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)
Public Score	<b>7.83378</b>	<b>7.44013</b>
Private Score	<b>5.50413</b>	<b>5.62719</b>
Total RMSE Score	<b>6.76992</b>	<b>6.59624</b>

在都抽九小時的情況下，只抽 pm2.5 所訓練出來的模型表現比較好，以下是我的推測

**抽全部污染源**：18 種特徵前 9 小時的值都使用，可能造成模型過於複雜，Variance 過大，造成最後的 RMSE 分數較差

**只抽取 pm2.5**：pm2.5 是一個很重要的特徵，取 9 小時很正確，且模型相較於前者簡單，Variance 較小，RMSE 分數較好

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

	抽全部 5 小時內的污染源 feature 的一次項(加 bias)	抽全部 5 小時內 pm2.5 的一次 項當作 feature(加 bias)
Public Score	<b>7.74074</b>	<b>7.57651</b>
Private Score	<b>5.37642</b>	<b>5.79427</b>
Total RMSE Score	<b>6.66427</b>	<b>6.74452</b>

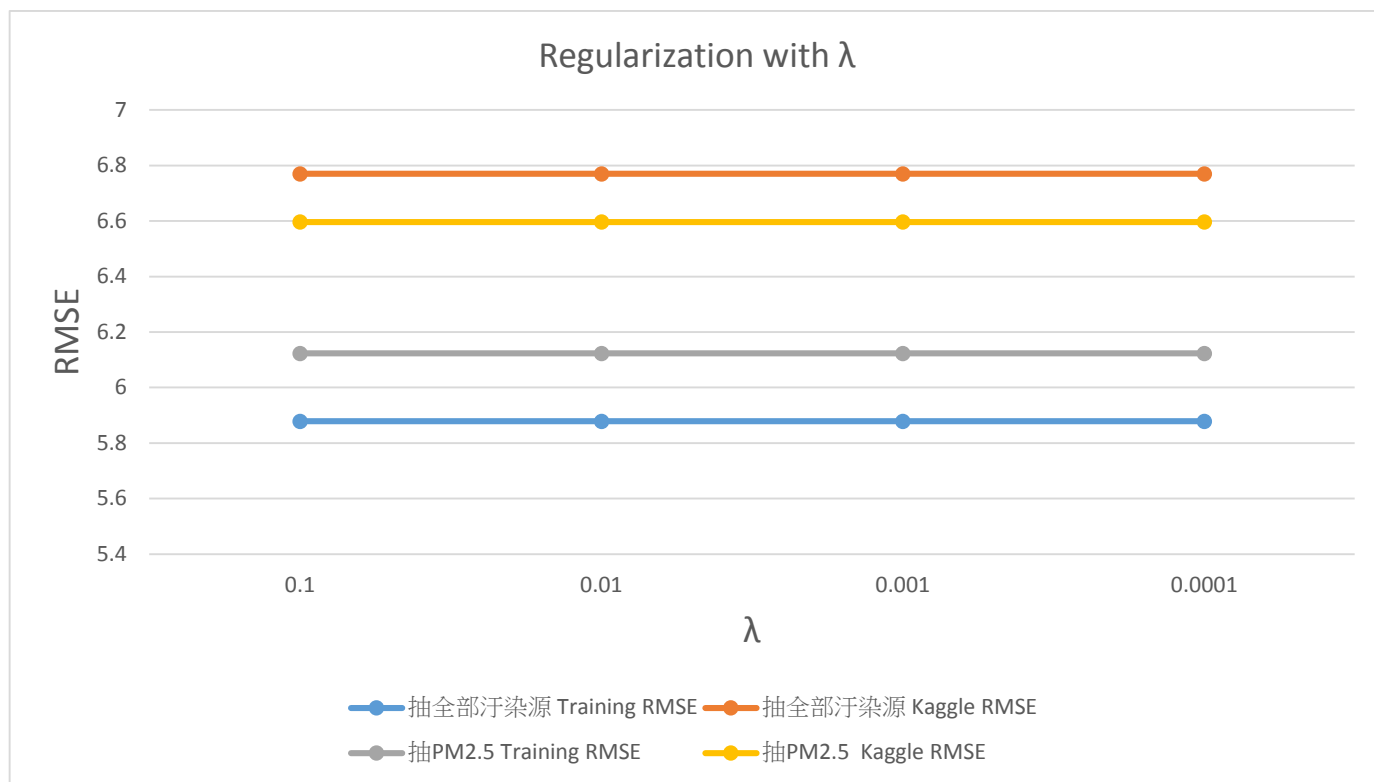
在都抽五小時的情況下，抽全部特徵所訓練出來的模型表現比較好，以下是其變化

**抽全部污染源**：18 種特徵從取前 9 小時改成取前 5 小時，使模型變得較簡單，Variance 縮小，最後 RMSE 分數比取 9 小時還來得好

**只抽取 pm2.5**：pm2.5 是一個很重要的特徵，改取 5 小時不是很好，且造成模型簡單過頭，導致最後的 RMSE 分數比取 9 小時的時候差

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

	抽全部污染源 Training RMSE	抽全部污染源 Kaggle RMSE	抽 PM2.5 Training RMSE	抽 PM2.5 Kaggle RMSE
0.1	<b>5.878286</b>	<b>6.76992</b>	<b>6.123022</b>	<b>6.596247</b>
0.01	<b>5.878284</b>	<b>6.76991</b>	<b>6.123022</b>	<b>6.596241</b>
0.001	<b>5.878284</b>	<b>6.76991</b>	<b>6.123022</b>	<b>6.596241</b>
0.0001	<b>5.878284</b>	<b>6.76991</b>	<b>6.123022</b>	<b>6.596241</b>



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

- a.  $(X^T X) X^T y$
- b.  $(X^T X)^{-1} X^T y$
- c.  $(X^T X)^{-1} X^T y$
- d.  $(X^T X)^{-2} X^T y$

$$\begin{aligned}
 & X = N \times n, \quad w = n \times 1, \quad y = N \times 1, \quad \text{令 } A \text{ 為 } X \text{ 之 column space} \\
 & \|Xw - y\|_{\min} \Rightarrow Xw = \text{Proj}_A y \Rightarrow (y - Xw) \perp A \\
 & \Rightarrow \langle y - Xw, Xz \rangle = 0, \quad \forall z = n \times 1 \\
 & \Rightarrow (Xz)^T (y - Xw) = 0, \quad \forall z = n \times 1 \\
 & \Rightarrow z^T X^T y - z^T X^T Xw = 0, \quad \forall z = n \times 1 \\
 & \Rightarrow z^T (X^T y - X^T Xw) = 0, \quad \forall z = n \times 1 \\
 & \Rightarrow \langle X^T y - X^T Xw, z \rangle = 0, \quad \forall z = n \times 1 \\
 & \Rightarrow X^T y - X^T Xw = 0 \\
 & \Rightarrow w = (X^T X)^{-1} X^T y
 \end{aligned}$$

答案為 c