學號： B04901056 系級： 電機三 姓名：張承洋

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
All features	7.46836	5.43763
PM2.5	7.4469	5.6239

從表格可以看出，對於 public set 來說，第一種 model(選擇所有汙染源)的誤差稍微大於第二種 model(選擇 PM2.5)，推測是因為第一種 model 選過多不重要的 feature 導致 overfitting。

然而，對於 private set 來說情況卻相反，猜測是 private set 跟 training set 相似度較高，導致複雜度高 model 的「small bias error」特性被凸顯出來，「large variance」特性不明顯所導致，若 testing data 的資料量提高或多測幾次，可能就會有不同的結果。

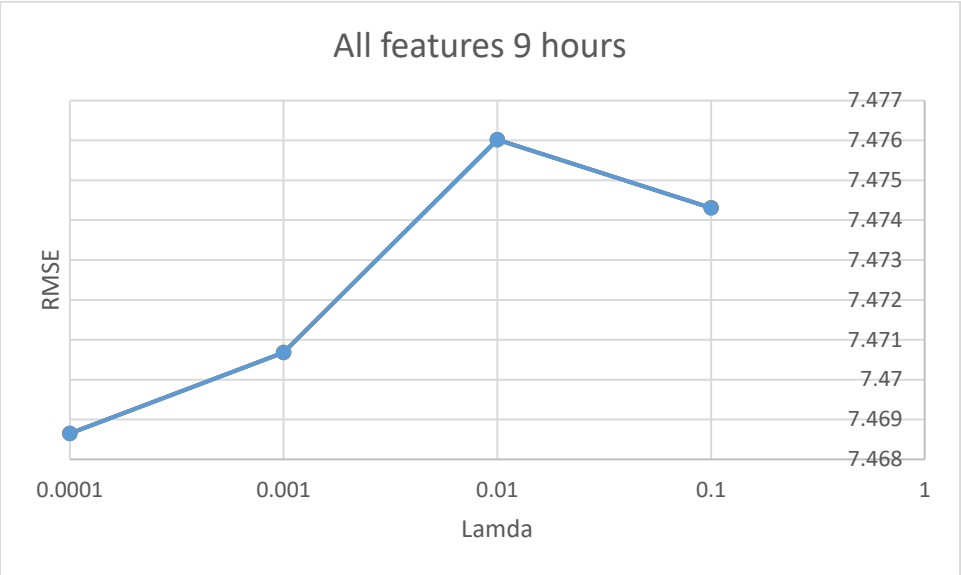
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

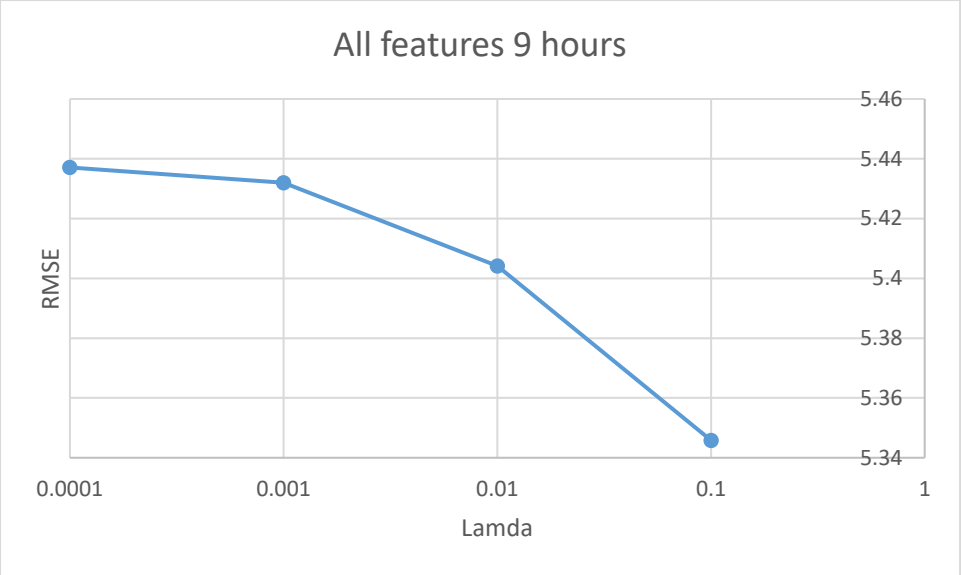
	Public	Private
All features	7.64894	5.41128
PM2.5	8.57338	5.78152

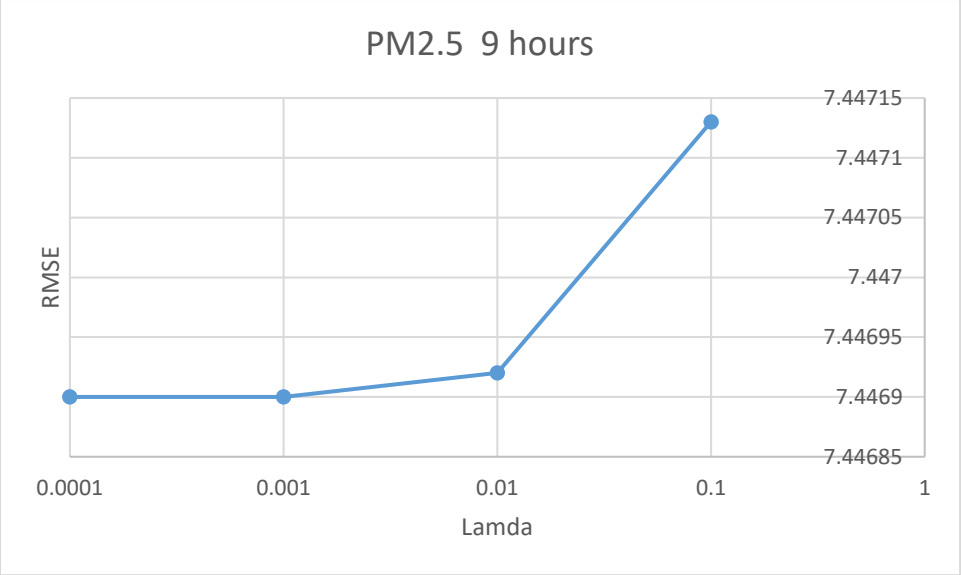
比較兩種 model 的 error 變化量可以發現，第一種 model 的 public set error 只有微微上升，private set error 甚至有下降的趨勢，推測原因是 overfitting 的現象得到改善所致。

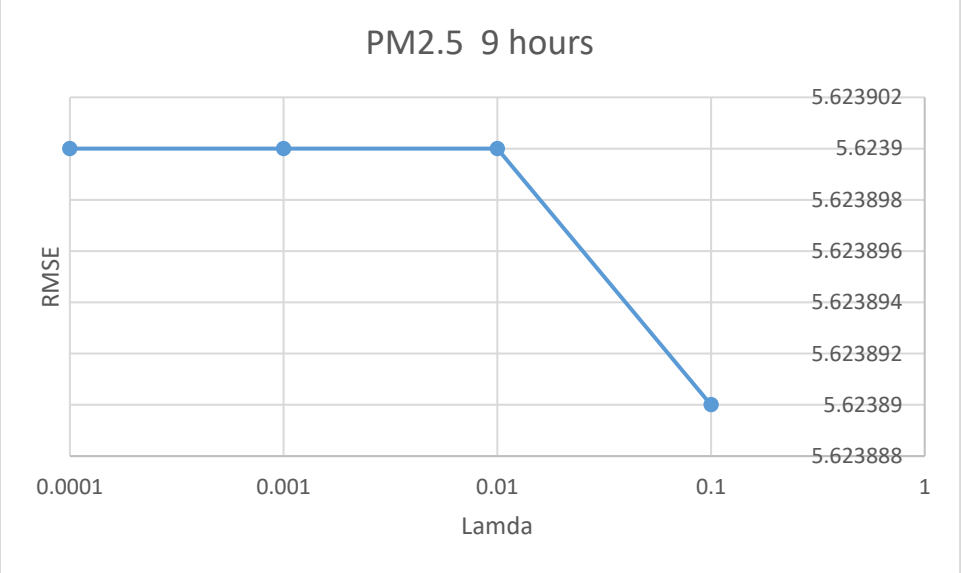
第二種 model 的 error 都有明顯的提升，猜測過去 9 小時的 PM2.5 是預測未來 PM2.5 的主要因素，我個人在實作時甚至將 PM2.5 的二次項(過去 4 小時)也納入考量，得出的 error 無論是在 training set 或是 testing set 都有明顯下降。

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









=>適當的 lamda 可以增加與測的準確度，但是過高或太低會有反效果

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$\sum_{n=1}^{N} (y^n - x^n \cdot w)^2$ 。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^TX 為 invertible)

(a) $(X^TX)X^Ty$

(b) $(X^TX)^{-0}X^Ty$

(c) $(X^TX)^{-1}X^Ty$

(d) $(X^TX)^{-2}X^Ty$

Ans:

Let $Y-Xw=0$

$Y=Xw$

$X^TY= X^TXw$

So $w=(X^TX)^{-1}(X^TY)$ 故選(c)