學號：R06922118 系級： 資工碩一 姓名：吳政軒

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

（1） 6.77912（全部 feature）

（2） 6.89455（PM2.5）

只抽取 PM2.5 的 feature 表現較好，可能是 18 個 feature 中有一些比較沒有預測力

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

（1） 6.70777（全部 feature）

（2） 7.10312（PM2.5）

用所有 feature 的 model 如果只用前 5 小時會預測的比較好，但是只用 PM2.5 的 model 就要用全部 9 小時預測比較好

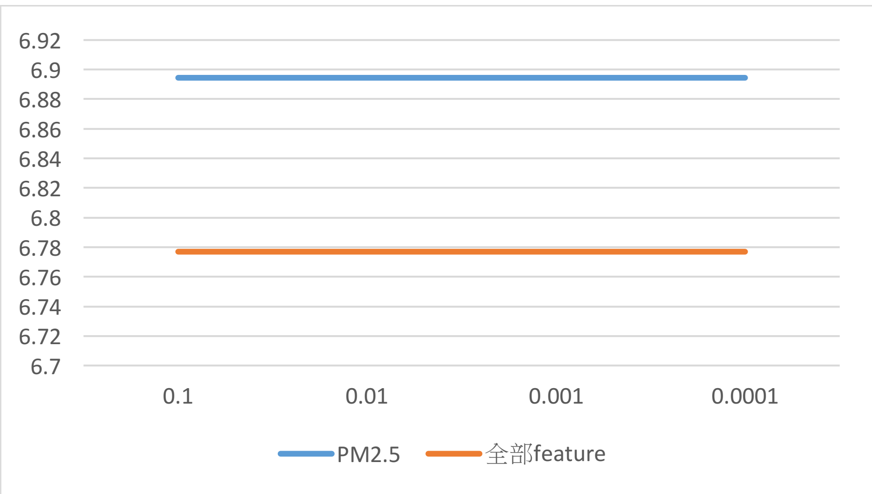
3. (1%)Regularization on all the weight with λ=0.1、0.01、0.001、0.0001，並作圖

pm2.5

6.89455	6.89455	6.89455	6.89455

全部 feature

6.77680	6.77680	6.77680	6.77680



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$\sum_{□=1}^{□}(□^{□} - □^{□} \cdot □)^2$ 。若將所有訓練資料的特徵值以矩陣 $X = [x^1\ x^2 \ldots x^N]^T$ 表示，所有訓練資料的標註以向量 $y = [y^1\ y^2 \ldots y^N]^T$ 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) $(X^TX)^{-2}X^Ty$

欲使||Xw-y||為最小，$\forall z, < y - Xw, Xz > = 0$
$\rightarrow < y, Xz > - < Xw, Xz > = 0$
$\rightarrow < y, Xz > = < Xw, Xz >$
$\rightarrow (Xz)^T y = (Xz)^T (Xw)$
$\rightarrow z^T X^T y = z^T X^T X w$
$\rightarrow < X^T y, z > = < X^T X w, z >$
$\rightarrow X^T y = X^T X w$
$\rightarrow \mathrm{w} = (X^T X)^{-1} X^T y$，故選(c)