學號：R04945008 系級：生醫電資碩二 姓名：黃思翰

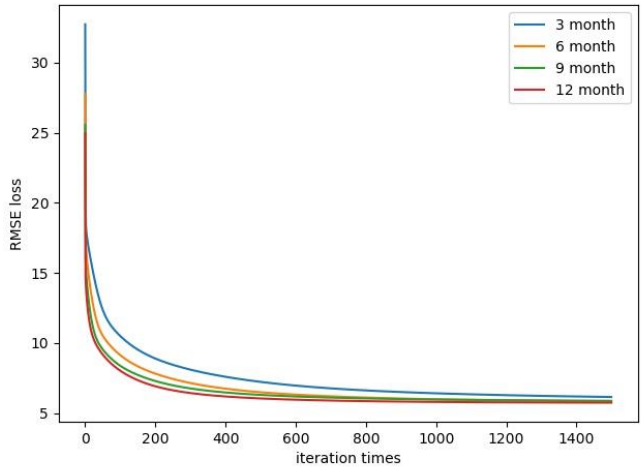
1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：共有兩種，將整份的 train.csv 讀入後，第一種是把每月的各項空氣汙染指標全數納入，共計 18 種不同的指標，每項指標取連續九小時當特徵，第十小時作為驗證用的答案，最後加入一個 1 進去陣列中做為 bias 內積用，所以共有 1+9*18=163 維的特徵，第二種一樣是取連續九小時的資料但僅取 pm2.5 的值，所以共有 1+9*1=10 維的特徵。

2.請作圖比較不同訓練資料量對於 PM2.5 預測準確率的影響

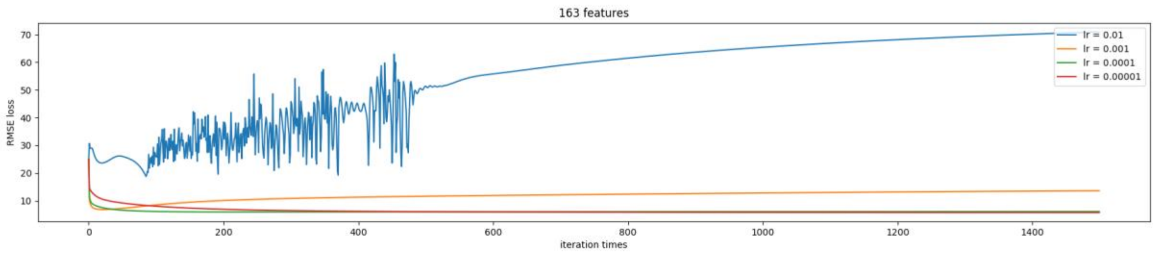
答：以下測試使用的參數如下表，比較了使用 3、6、9 和 12 個月的量的 trainning data 的狀況。從下圖可以觀察到隨著使用的 trainning data 量越多越能提早降低預測的誤差。

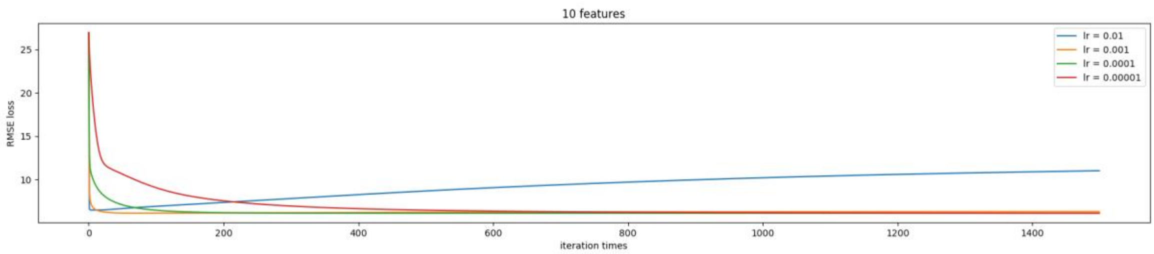
方法	Feature	Iteration times	Learning rate	Optimizer	Beta1	Beta2	Epsilon
Gradient descent	163 維	1500	0.0001	Adam	0.9	0.99	1e-8



3. 請比較不同複雜度的模型對於 PM2.5 預測準確率的影響

答：



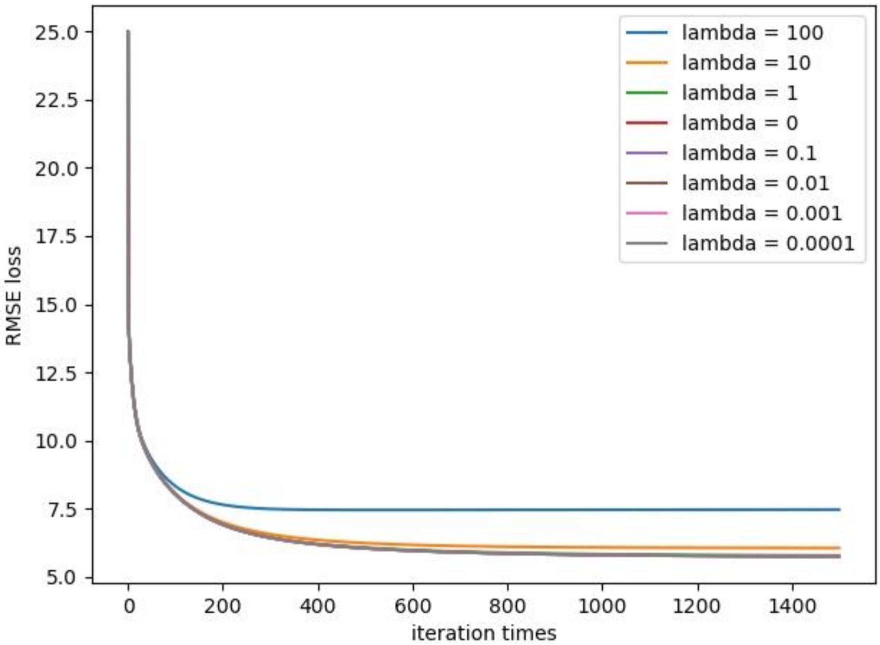


上圖比較了使用 18 種空氣指標連續取樣九小時後共產生 163 維的模型與僅使用 PM2.5 的 10 維模型在不同的 learning rate 下所得出的 RMSE 比較。從結果上來看，僅使用 PM2.5 的模型在相同的 learning rate 時擁有比較高的預測正確率，同時也比較快收斂。

4. 請討論正規化(regularization)對於 PM2.5 預測準確率的影響

答：以下測試使用的參數如下表，比較使用不同 lambda 時對於預測誤差的影響。由於 lambda 在小於 1 後對於誤差的影響非常非常低，在圖上會幾乎相疊在一起，造成看似只有三條線的錯覺。從結果來看，正規化在本例子中影響不大，lambda 過大會造成誤差增加，lambda 過小僅使 RMSE 於小數點第 3 位有變化。

方法	Feature	Iteration times	Learning rate	Optimizer	Beta1	Beta2	Epsilon
Gradient descent	163 維	1500	0.0001	Adam	0.9	0.99	1e-8



5. 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$\sum_{n=1}^{N}(y^n - w \cdot x^n)^2$ 。若將所有訓練資料的特徵值以矩陣 X = [x^1 x^2 … x^N] 表示，所有訓練資料的標註以向量 y = [y^1 y^2 … y^N]^T 表示，請以 X 和 y 表示可以最小化損失函數的向量 w 。

答：假設 loss function = $\sum_{n=1}^{N}(y^n - w \cdot x^n)^2$ ，y(N*1 維)、X(N*d 維)、w(d*1 維)，將 loss function 寫成矩陣形式 $|y - X \cdot w|^2$，而本題所求的最小化 loss function 的向量 w 等同於求出能使 ∇L(w) = 0 的向量 w。其中，

$$\nabla L = -2 \cdot (y - X \cdot w)^T \cdot X = -2 \cdot (y^T \cdot X - X^T \cdot X \cdot w^T)$$

而最終找出能使 ∇L(w) = 0 的 w 為

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