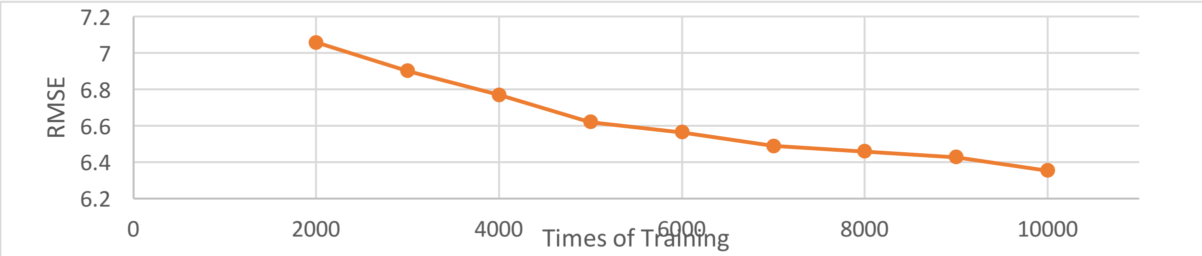
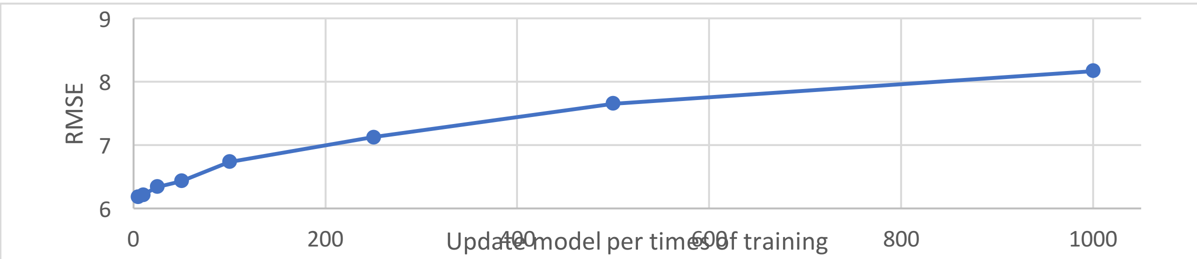
厂 學號：B02404002 系級： 醫技四 姓名：葛竑志

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)
   (1)將 train.csv 以 pandas 讀取，取其中除了 label 以外的 values。
   (2)將 data 以 np.vsplit 切割成 12 個區塊（十二個月），每一區塊再 vsplit 切割成 20 個區塊（二十天）。最後再將每一個月的 data 以 np.hstack 堆疊成連續二十天的逐小時資料。
   (3)針對 RAINFALL 的 NR，使用 np.place 將之取代為‘0.0’。
   (4)將抽取出來的 data 使用 astype(np.float)將 String 轉換為 Float。
   (5)x_train 取每個月的連續 9 小時的 18 項 features（起始時間點自第 0 小時至第 470 小時）。
   (6)y_train 取每個連續 9 小時資料後一小時的 PM2.5 值（起始時間點自第 9 小時至第 479 小時）。
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
   (1)固定相同 Training set 及 Batch 情形下，訓練不同次數。RMSE 由固定 Random Seed 從 Validation batch 中挑選 10 次 500 筆 pair 做驗證。



   (2)固定相同 Training set 及訓練次數，以不同 batch size 更新 model parameter。RMSE 算法同上。



   (3)結果與討論：
      a. 在固定 Training set 與 Batch 環境下，訓練 pair 的次數越多，越能降低 RMSE，但其降低速率會隨著迴圈數增加而減少。
      b. 在固定 Training set 與訓練次數的環境下，每次投入越少 pair 於 model 中並更新參數（batch 越小），越能降低 RMSE。本次最低投入資料量為每 5 筆資料更新一次。
3. 請比較不同複雜度的模型對於 PM2.5 預測準確率的影響
   (1)本次實作中共製作了一次至四次線性回歸的 Model。
   (2)通常越高次的 model 中 learning rate 需要調到很小，避免過於震盪。

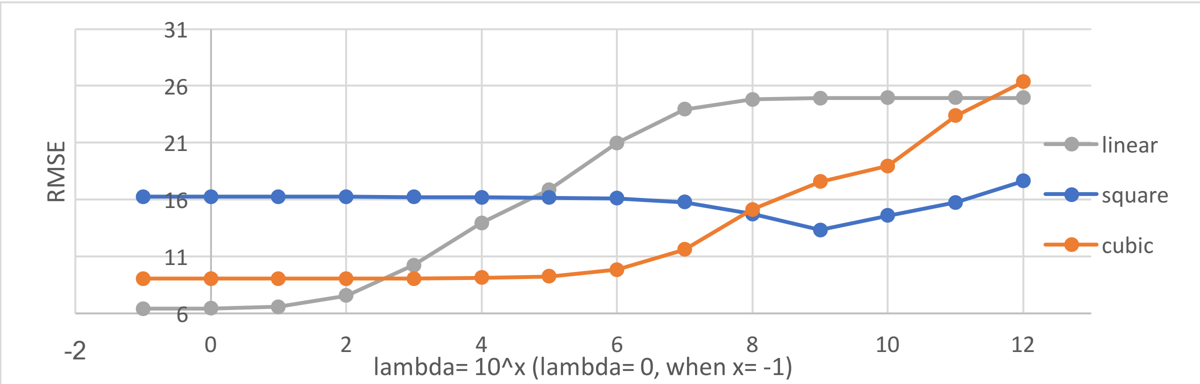
(3)通常 Regularization 可以在高次的 model 中發揮顯著效果，因爲必須要產生 overfitting 才有此必要性。

(4)在此實作中抽取 features 的方式如題一所述，高次向的 model 訓練時間會巨幅增加，也使雜訊增加，不易根據結果調整參數，且不容易可以得到比低次項更好的結果。

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

答：

(1)對所有 Features 採不同 Model(不同次方項)做 Linear regression，測試 regularization 的不同 weight(lambda)將對 RMSE 產生之影響。



(2)結果與討論:

a. 在一次 Linear 回歸與 Cubic 情況下，lambda 增加始終導致 RMSE 上升而無降低趨勢。在 Square 項次的情況時，RMSE 先隨著 lambda 增加而下降，但當 lambda 過大時又會使 RMSE 增加。

b. 在一次 Linear model 中，訓練至最後似乎都不太具有 over fit 的問題，因此 regularization 不太具有效果。而當項次增加時，regularization 有發生作用，但也可能因為採用 feature 數過多（18*9 項），使得此方法無法造成有效之結果。

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 $\mathbf{x}^n$，其標註(**label**)為一存量 $\mathbf{y}^n$，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 $\mathbf{X} = [\mathbf{x}^1\ x^2 \dots x^N]$ 表示，所有訓練資料的標註以向量 $\mathbf{y} = [\mathbf{y}^1\ y^2 \dots y^N]^T$ 表示，請以 **X** 和 **y** 表示可以最小化損失函數的向量 **w** 。

(1)$L = \sum_{n=1}^{N}(y^n)^2 - 2\sum_{n=1}^{N} wy^n x^n + \sum_{n=1}^{N} w^2(x^n)^2$

(2)$\frac{\partial L}{\partial w} = -2\sum_{n=1}^{N} y^n x^n + 2w\sum_{n=1}^{N}(x^n)^2$

(3)min(L) : $\frac{\partial L}{\partial w} = 0$

(4)$\sum_{n=1}^{N} y^n x^n = w\sum_{n=1}^{N}(x^n)^2$

(5)$Xy = wX^TX$

(6)w= $Xy(X^TX)^{-1}$