學號：B03705012　系級： 資管三　姓名：張晉華

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：

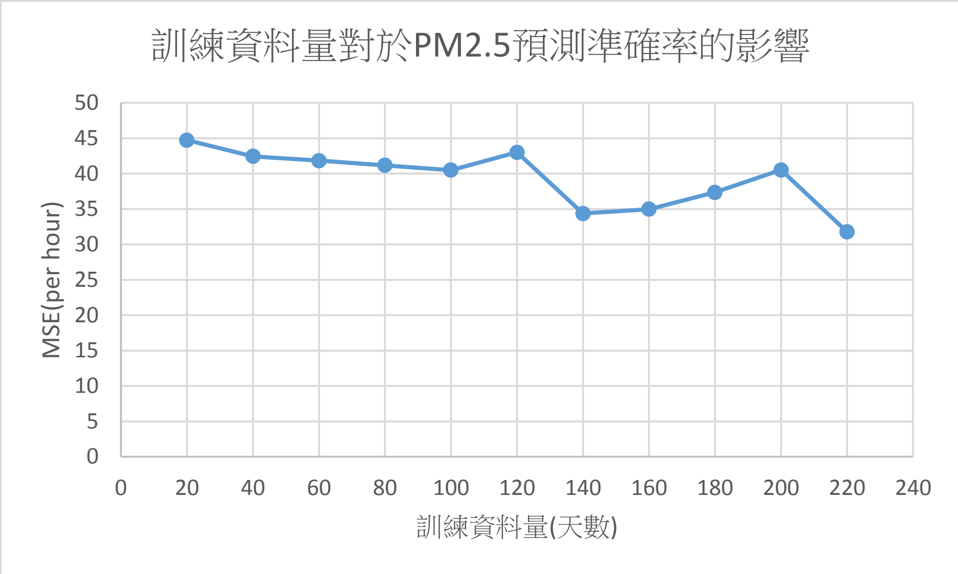
利用過去連續 9 小時的 PM2.5 資料，用 Gradient Descent 實作 linear regression

Learn_rate 初始值為 0.1，若遇到 MSE 變大則降低 Learn_rate 一半

當兩次 MSE 相差不到 0.000001 時視為學習完成

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：



由圖可知訓練的資料量越多 MSE 會有下降的趨勢，但

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：

使用 PM2.5 前 n 個小時的資料	MSE(per hour)
1	51.834710872354385
2	50.919401833046123
3	49.633062069812148
4	45.277861842148653
5	45.051414136794023
6	45.003486846379289
7	44.043806373433981
8	44.009725785595144
9	43.920244207431281

使用越多的 PM2.5 資料精準度會提高。

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

答：

lambda	MSE
0	43.910866873632067
0.01	43.910871948130549
0.1	43.910917618616899
1.0	43.911374323480374

10.0	43.915941372115142
100.0	43.961807908411487
1000.0	44.420373830203843

使用正規化(regularization)對於 PM2.5 預測準確率在這個模型下看來沒有顯著的幫助

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 $X = [x^1\ x^2\ \cdots\ x^N]$ 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，請以 X 和 y 表示可以最小化損失函數的向量 w 。

答：

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

⇨ $\min(y - wX)^2$

⇨ 因為 X 是 linear independent

⇨ $w(XX^T) = yX^T$

⇨ $w = yX^T(XX^T)^{-1}$