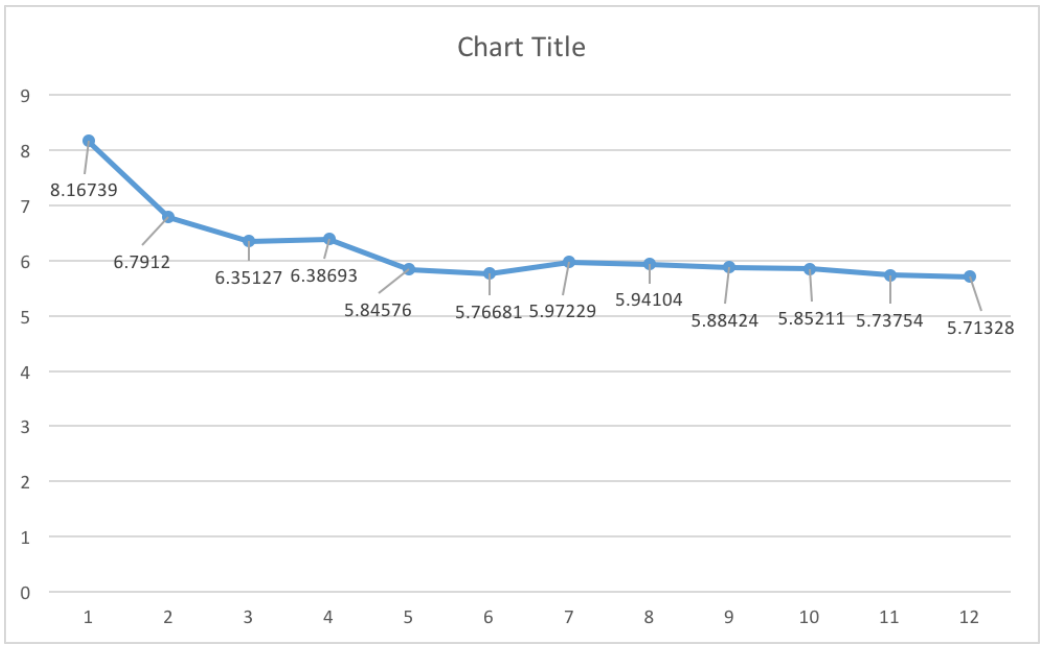
學號：B04902044 系級： 資工二 姓名：朱柏澄

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)
答：
我把每一種feature對PM2.5計算相關係數，再根據相關程度大略選出一個輸入特徵，然後一次增加或減少一個feature，觀察最後RMSE的大小、以及RMSE的收斂速度。經過不斷的測試，得到最好的輸入特徵。
最後我的模型的輸入特徵有：PM2.5*9, PM10*5, RAINFALL*3, O3*3, NO2*1, WIND_SPEED*2, PM2.5^2 *4, PM10^2 *2

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：



橫軸為training的資料量（月）、縱軸為RMSE
如圖，準確率在資料量有一定數量以後，會逐漸趨於穩定，並有緩慢上升的趨勢。但是並非只要增加資料量，就可以增進準確率，圖中就有發生資料量增加而準確率下降。一般而言增加資料量，可以使學習的效果更好，準確率更高，但新加入的資料，會使整體資料的分佈稍微改變，有時候反而會造成準確率稍微下降。

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：
feature過少的模型，預測的準確率會比較低，加入一些合適的feature，可以使準確率提高，例如：只有選PM2.5*3會比PM2.5*9的準確率低。
但不斷加入feature並不一定總是會增加準確率，例如把前九天的所有資料當作162維的feature，其準確率只能通過simple baseline。
因此選擇模型時要選擇適當且適量的feature，才能有效地增加準確率。

4. 請討論正規化(regularization)對於PM2.5預測準確率的影響
答：
我試著在不同模型中加入正規化，但並沒有得到明顯進步的結果。正規化在我的模型中對於PM2.5的預測並沒有顯著的影響，因為我的模型沒有高次方的參數，僅有幾項平方項的參數，但在其他模型中，正規化對PM2.5的預測可能會有影響。

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

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