

學號：R05922135 系級：資工碩一 姓名：蔡慶源

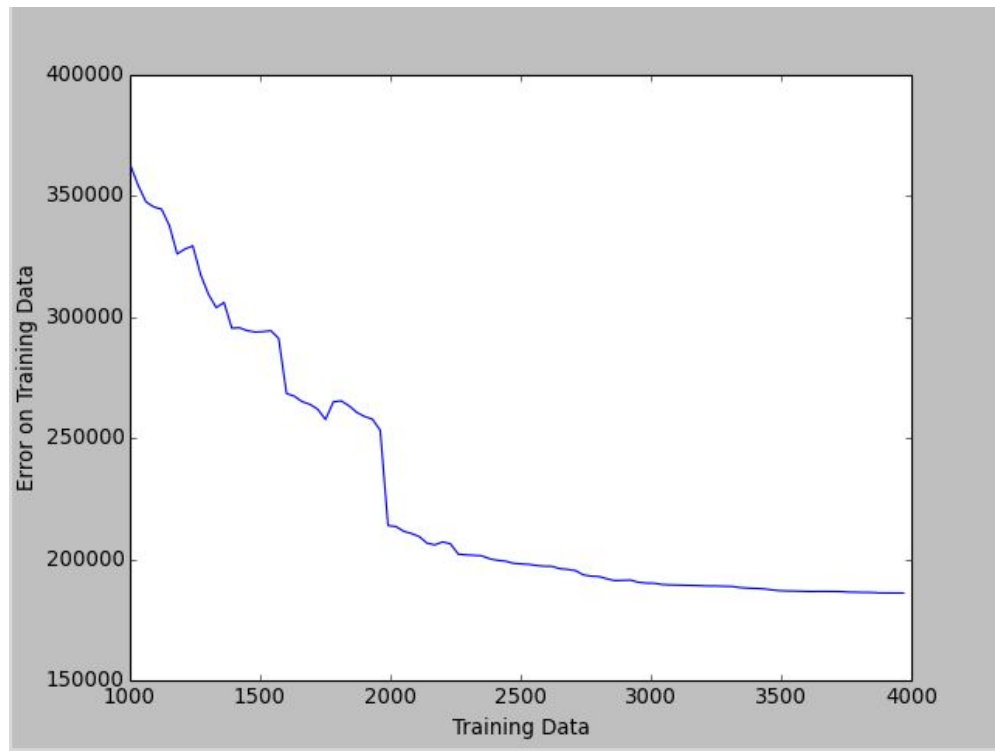
1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：

每10小時為一筆，前九小時所有資料為特徵有162個，取特徵不跨月份。所以一個月共有 $20 \times 24 - 9$ 筆輸入，總共5652筆。

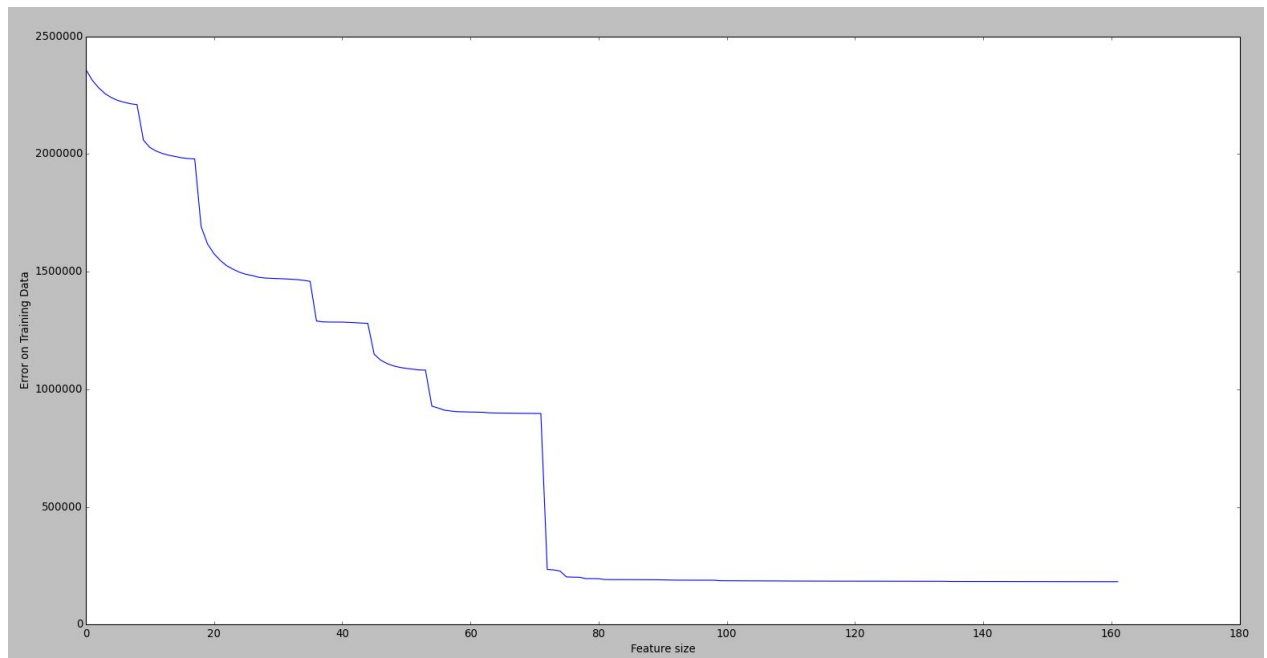
2. 請作圖比較不同訓練資料量對於PM2.5預測準確率的影響

答：



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：



圖為9個小時162個特徵依時間由後到前依序增加。

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

答：

加入正規化會讓各個weight更平滑，但我使用時沒有讓我預測更準確。

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 $X =$

$[x^1 \ x^2 \ \dots \ x^N]$ 表示，所有訓練資料的標註以向量 $y = [y^1 \ y^2 \ \dots \ y^N]^T$ 表示，請以 X 和 y 表示可以最小化損失函數的向量 w 。

答：

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