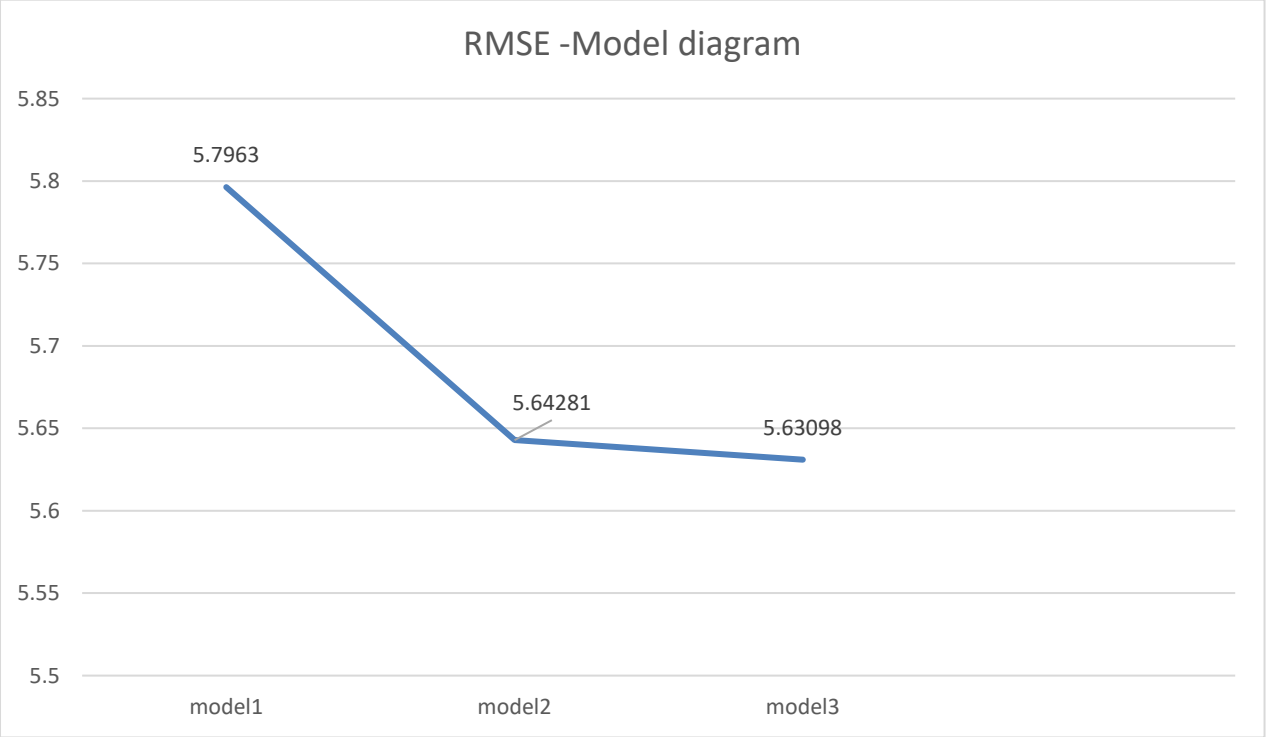
學號：B03901030 系級： 電機三 姓名：蕭晨豪

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)
答：
在 hw1_best 中我取前九個小時的 PM2.5 一次項以及前兩小時的 PM2.5 的平方項作為 feature

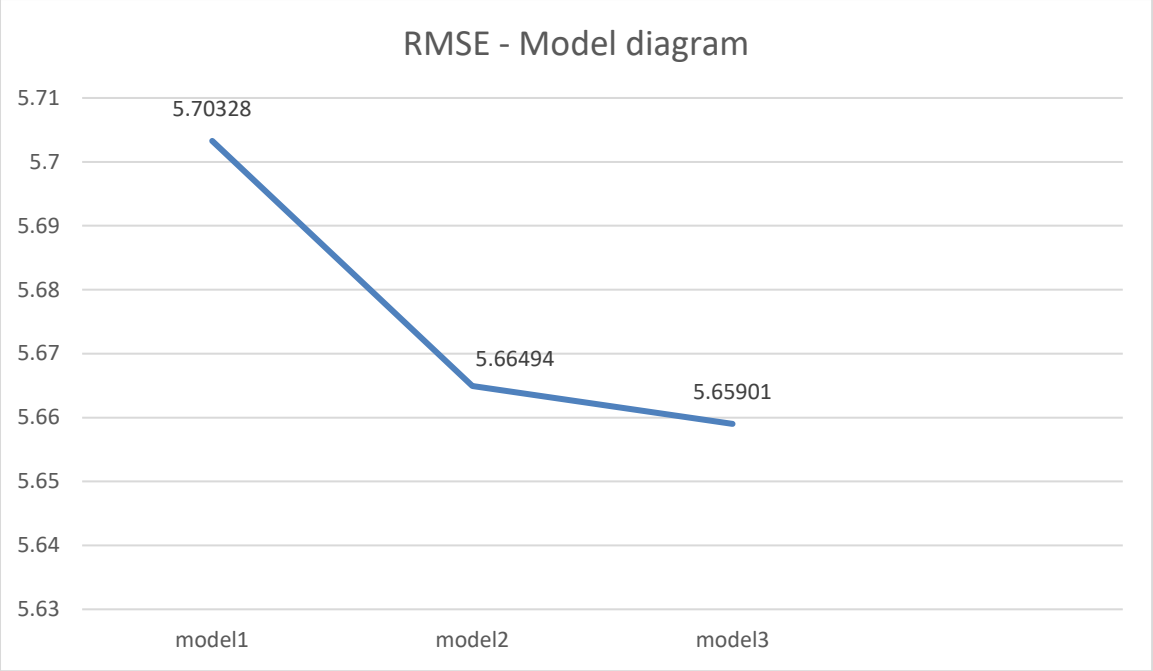
2.請作圖比較不同訓練資料量對於 PM2.5 預測準確率的影響
答：
以下三筆資料均使用 stochastic training, iteration 60 萬次，在最後 1000 次時不使用 stochastic 而使用所有的值做一般的 gradient descent
model 1:十二個月中每個月隨機取五天
model 2:十二個月中每個月隨機取十天
model 3:十二個月中每個月二十天全取



在此方法下越多的訓練資料量達到越好的訓練效果

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：

我主要比較了三種:
model 1:取前九個小時的 PM2.5 一次項當作 feature
model 2: 取前九個小時的 PM2.5 一次項以及前一小時的 PM2.5 的平方項當作 feature
model 3: 取前九個小時的 PM2.5 一次項以及前兩小時的 PM2.5 的平方項當作 feature



此三種 model 的 training 方式同第 2 題，可看出越複雜的 model 有越好的效果

4. 請討論正規化(regularization)對於 PM2.5 預測準確率的影響
答：
當我的 feature 取前九個小時的 PM2.5 一次項時，不論我的 lambda(regularization 參數)取多少 performance 都沒有進步(大部分都比沒有 regularization 差一點點)
但當我的 feature 取前九個小時除了 rainfall 以外的所有參數的一次項時，regularization 將原本的 RMSE 從 6.11286 降至 6.1009(取 lambda = 0.05)

我認為這是因為 regularization 主要是加一個 penalty 在取過多的參數以及過高的 weight 上面，當我取很多參數時 regularization 就會顯現其效果，可是當我只取前九個小時的 PM2.5 一次項時，可能因為參數已經夠少了，regularization 並不會讓效果變好

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)為一純量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：
在取出特徵後我們的預測結果為Xw，因此我們想找到 w 使$(y - Xw)^T(y - Xw)$最小(也就是讓我們定義的 loss function 最小)，將 loss function 對 w 做偏微分後得到$-2X^T(y - Xw)$，使其為零得到$w = (X^TX)^{-1}X^Ty$