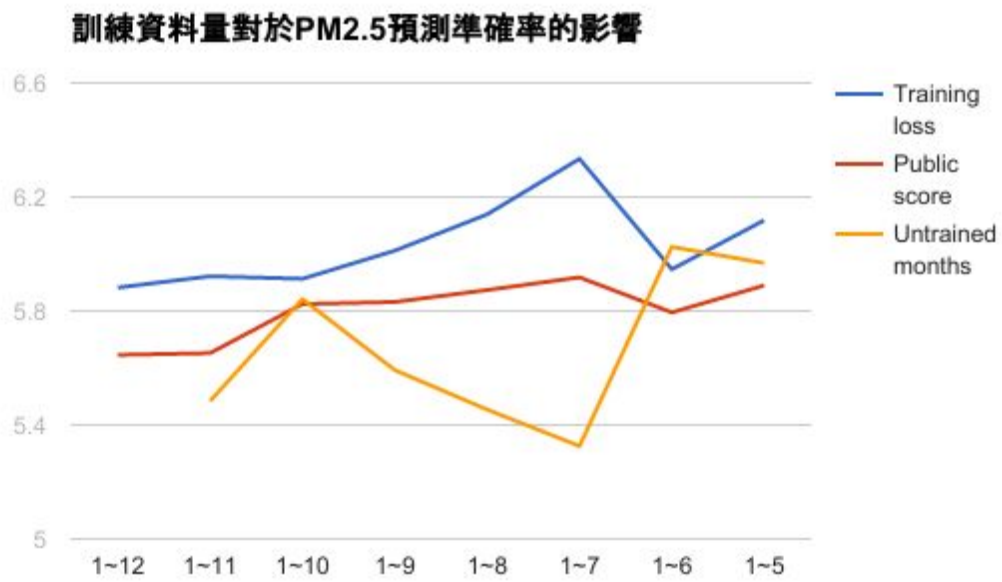
學號：B03902042 系級：資工三 姓名：宋子維

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)
答：
原本取前9個小時的18種空氣汙染指標，作為輸入特徵，故總共為162種特徵，後來經過不斷的測試和domain knowledge，**選用PM10, PM2.5, O3, WIND_DIREC, WIND_SPEED, WD_HR, WS_HR, RAINFALL, PM10 ** 2, PM2.5 ** 2，總共為90種特徵，而第二及第四題的測試標準也是用這組特徵，且訓練次數為20000，初始學習率為0.5，並且使用adagrad。**

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：
圖表的橫軸代表拿來訓練的月份，每個月份有471筆資料。從折線圖的趨勢來看，**資料量越大，public score越低**，但值得一提的是，這和「用什麼資料來訓練」有很大的關係：可以看到1~7和1~6這兩個點，若不訓練7月反而能使public score下降，猜想可能是因為7月的資料對於public set是noise，才會有這樣的結果出現。




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：
0~17依序代表以下18個測項： AMP_TEMP, CH4, CO, NMHC, NO, NO2, NOx, O3, PM10, PM2.5, RAINFALL, RH, SO2, THC, WD_HR, WIND_DIREC, WIND_SPEED, WS_HR。每一個測項都使用完整的9個小時，而Dot term指的是將兩個測項每小時個別相乘。
**從表格可看出，並不是越多越高次的feature就能讓結果越好，有些模型反而因此產生overfitting（如第三個）。**

| Linear term | Square term | Cubic term | Dot term | training loss | Public score |
|---|---|---|---|---|---|
| 0~17 | None | None | None | 5.68531 | 5.95965 |
| 0~17 | 0~17 | None | None | 5.53272 | 5.85442 |
| 0~17 | 0~17 | 0~17 | None | 5.43825 | 6.09038 |
| 7, 8, 9, 10, 14, 15, 16, 17 | 8, 9 | None | None | 5.88212 | 5.64586 |
| 2, 7, 8, 9, 10, 14, 15, 16, 17 | 8, 9 | None | None | 5.82960 | 5.68664 |
| 2, 7, 8, 9, 10, 12, 14, 15, 16, 17 | 8, 9 | None | None | 5.78344 | 5.65403 |
| 2, 7, 8, 9, 10, 14, 15, 16, 17 | 同一次項 | None | 7 * 9 | 5.72772 | 5.69233 |
| 2, 7, 8, 9, 10, 12, 14, 15, 16, 17 | 同一次項 | None | 7 * 9 | 5.67329 | 5.63437 |

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

| Regularization | Training loss | Public score |
|---|---|---|
| 0.01 | 6.14937 | 5.73088 |
| 0.001 | 5.91889 | 5.65030 |
| 0.0001 | 5.88604 | 5.64798 |
| 0.00001 | 5.88264 | 5.64917 |
| 0.000001 | 5.88216 | 5.64619 |
| 0.0 | 5.88212 | 5.64586 |

可以看到regularization對於public score並沒有太大的幫助，甚至加了regularization會讓RMSE上升，可能是我使用的這組feature對training set並沒有達到overfitting，regularization才會沒什麼功用。

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：

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

(註：在我的hw1_best.sh中，就是用closed-form解出最佳解。)