ML2017 HW1 Report

學號：B02901124 系級： 電機四 姓名：黃柏翔

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

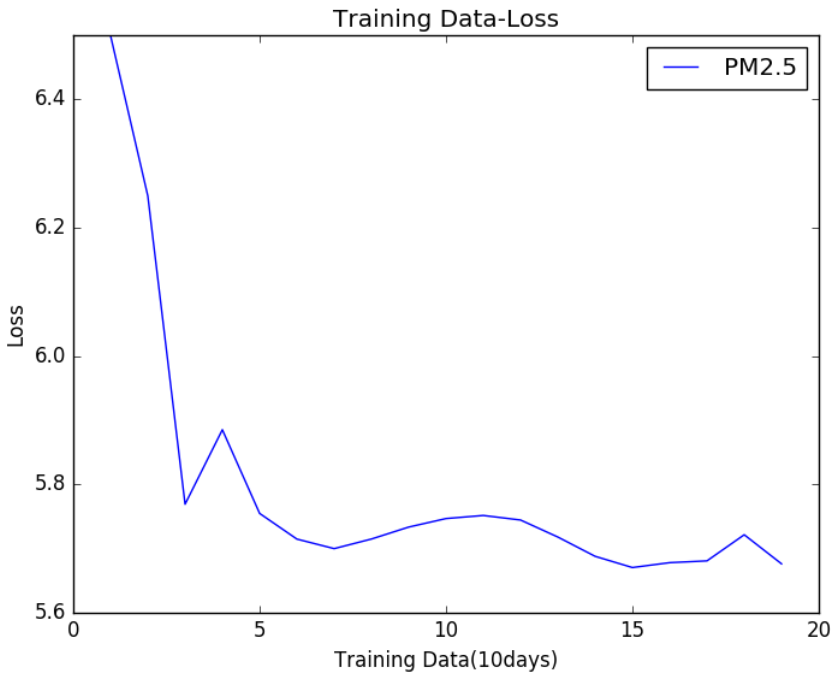
答：

我創建了一個以 18 個統計項目 (PM2.5,RAINFALL,NO2...等等)爲 index 的 dict, 在每個 index 下儲存著每九天的數據爲一個分割的陣列（feature）。以方便我之後選擇 training data 的項目。在數字普遍比較小的項目下 (ex. WIND_SPEED, SO2, THC)，在儲存時我會乘上 10 倍以確保在之後能影響到整個 training 的過程 (雖然之後沒有用到這些 data...)

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：

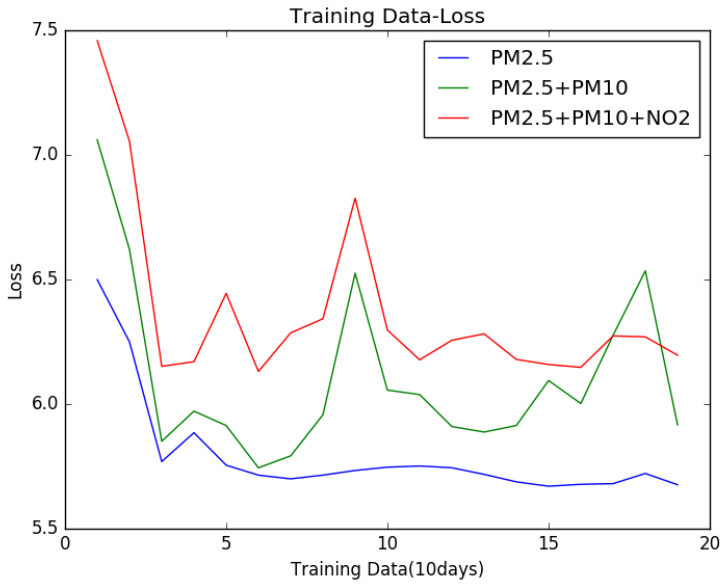
首先將 Training data 切爲 training set 及 validation set (200 天,40 天)，再以 10 天爲一個 mini-batch 的前提下逐漸提升 training data 的量至 200 天爲止作圖。經由 100 次的 iteration 及 1e-5 的 learning rate。由下圖可以看出，在大致上的趨勢下增加 training set 的大小有助於提升預測 PM2.5 的準確率 (Loss 逐漸變小)。



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：

先找出各項數據與 PM2.5 的相關係數後取出與 PM2.5 的關係最密切的幾項。經由計算得到第 6 項的 NO2 (0.45)與第 9 項的 PM10 (0.77) 最爲相關。並設計模型將這兩項也納入 liner regression 的計算中。並參照上題的方式利用不同大小的 training data 對 Loss 作圖。

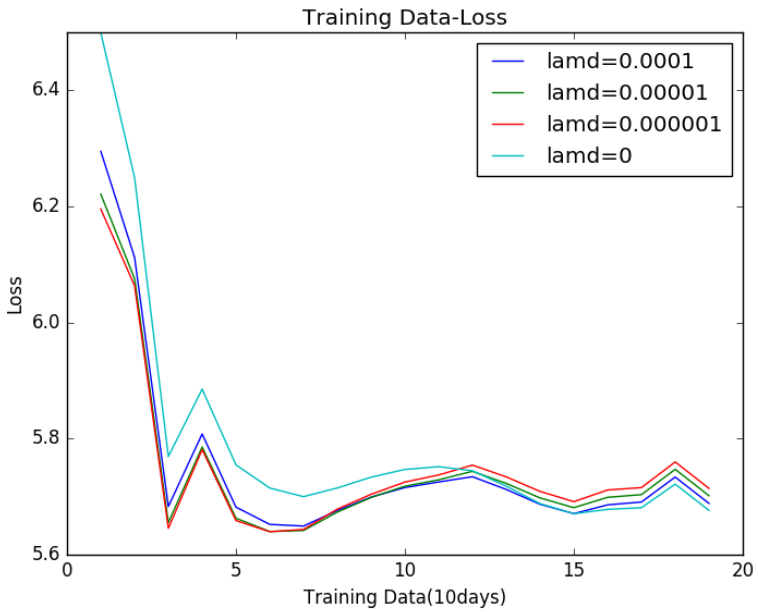


與原先預期不同的是，增加了較多的參考資料對於預測並沒有顯著的幫助，考慮的越多反而更容易造成反效果。

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

答：

在我所進行的實驗中，ragularization 對於預測準確率的影響似乎不太大。不論將權重調整到多少。最後還是會收斂到差不多的值。在 Kaggle 上的分數也都沒有太大的變動。



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

$$\mathbf{Xw} - \mathbf{y} = \begin{pmatrix} \mathbf{w} \cdot \mathbf{x_1} - y_1 \\ \vdots \\ \mathbf{w} \cdot \mathbf{x_m} - y_m \end{pmatrix}$$

$$\sum_{i=1}^{m} (\mathbf{w} \cdot \mathbf{x_i} - y_i)^2 = \|\mathbf{Xw} - \mathbf{y}\|_2^2$$

$$\min_{\mathbf{w}} (\mathbf{Xw} - \mathbf{y})^\top (\mathbf{Xw} - \mathbf{y}) = \min_{\mathbf{w}} \mathbf{w}^\top \mathbf{X}^\top \mathbf{Xw} - 2\mathbf{w}^\top \mathbf{X}^\top \mathbf{y} + \mathbf{y}^\top \mathbf{y}$$

$$\bigtriangledown_{\mathbf{w}} = 2\mathbf{X}^\top \mathbf{Xw} - 2\mathbf{X}^\top \mathbf{y} = 0$$

$$\begin{aligned} \left(\mathbf{X}^\top \mathbf{X}\right) \mathbf{w} &= \mathbf{X}^\top \mathbf{y} \\ \mathbf{w} &= \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{y} \end{aligned}$$