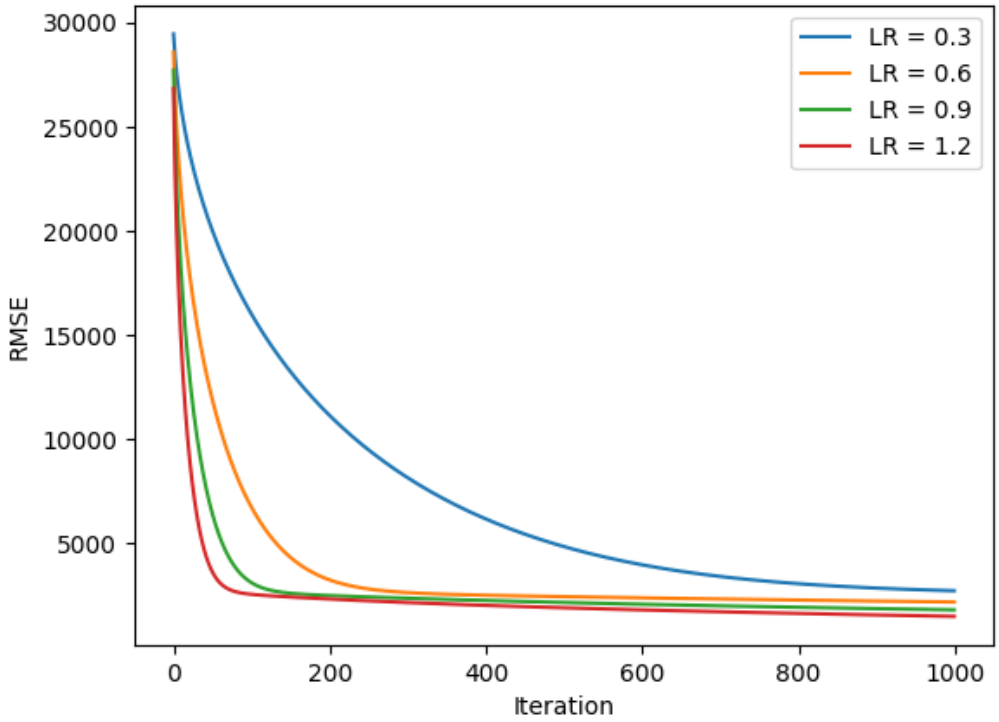
Homework 1 Report - PM2.5 Prediction

學號：b04203058　系級：化學三　　姓名: 蘇軒

- Report.pdf 檔名錯誤（-1%）
- 學號系級姓名錯誤（-0.5%）

1. (1%) 請分別使用至少 **4** 種不同數值的 learning rate 進行 training（其他參數需一致），對其作圖，並且討論其收斂過程差異。



可以很明顯的發現越小的 leaning rate 會使其在收斂的過程曲線變得比較圓滑並且比較慢收斂到穩定且固定的 RMSE; 相反的，雖然較大的 learning rate 在收斂的過程比較陡沒有這麼的平滑，可是他比較快的收斂到幾乎固定的值。

2. (1%) 請分別使用每筆 data9 小時內所有 feature 的一次項（含 bias 項）以及每筆 data9 小時內 PM2.5 的一次項（含 bias 項）進行 training，比較並討論這兩種模型的 root mean-square error（根據 kaggle 上的 public/private score）。

將數據做 sliding window 取每筆九個小時出來的一次項月份不相連並做 1E6 個 iteration (learning rate = 1)得到以下數據:

	Public	Private
全 feature	8.80762	8.82028
Pm2.5	9.05513	8.75951

很明顯地發現 Pm2.5 做出來的效果比全 feature 的效果還要來的低，推測原因是因為如果只有單單 Pm2.5 noise 太多了 數據並沒有好好地拿去處理使得我們做出來的結果並不好，而全 feature 則因為大量的 data 稀釋了這項缺點。

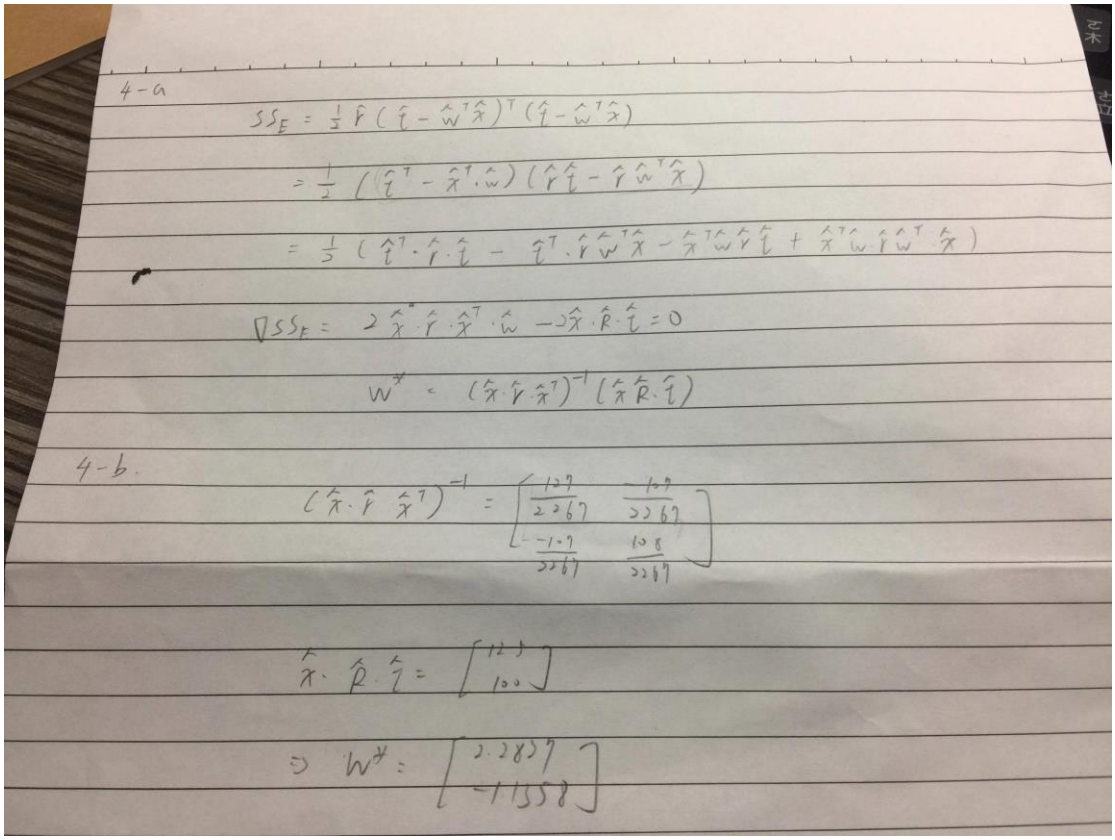
3. (1%)請分別使用至少四種不同數值的 regulization parameter λ 進行 training（其他參數需一至），討論及討論其 RMSE(traning, testing)（testing 根據 kaggle 上的 public/private
score）以及參數 weight 的 L2 norm。

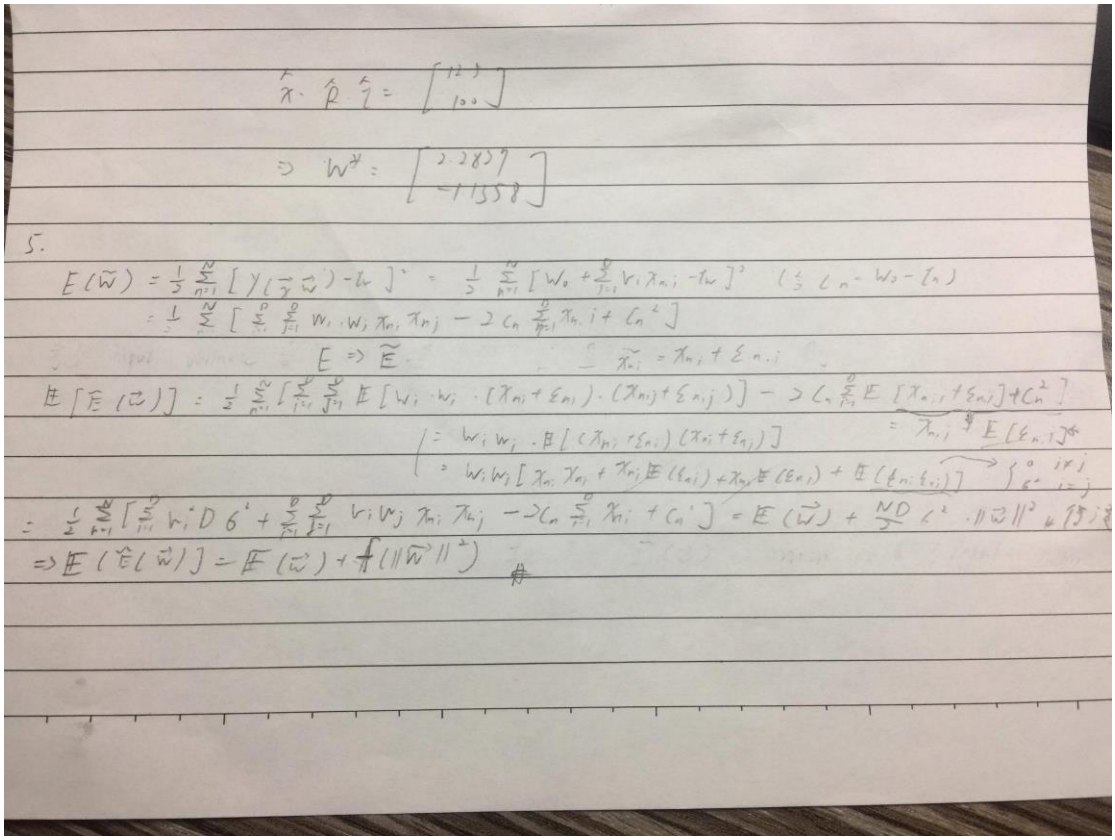
這次只取一個 feature Pm2.5 並且拿去做 sliding window 做 1E5 個 iteration (learning rate = 5)得到以下的數據:

λ	RMSE (public/private)	weight
1	9.05460 / 8.76326	1.91324
10	9.05533 / 8.76397	1.91112
100	9.06421 / 8.76158	1.95779
1000	9.20683 / 8.68981	3.16138

其實四組數據中除了 λ = 1000 的比較大以外 其他三組的 RMSE 還有 weight L2 norm 都差不了太多 第四組變得這麼誇張有可能是因為將 λ 設定的太大使得 gradient 的下降幅度太小孩沒有達到所以達到的最低點比其他三組都還要來的高。此外，我有比較過沒有做 regularization 的情況下有明確的發現 weight 下降地並沒有這麼緩和

4~6 (3%) 請參考數學題目（連結：），將作答過程以各種形式（latex 尤佳）清楚地呈現在 pdf 檔中（手寫再拍照也可以，但請注意解析度）。





6.

$$\frac{d}{d\alpha}\ln(|A|) = \frac{1}{|A|}\frac{d}{d\alpha}|A| = \frac{1}{|A|}|A|\cdot Tr\left[A^{-1}\frac{d}{d\alpha}A\right] = Tr\left(A^{-1}\frac{d}{d\alpha}A\right) \quad \#$$

[from Jacobis formula $\frac{d}{d\alpha}|A| = Tr[adj(A)\frac{dA}{d\alpha}]$

$adj(A) = A^{-1}\cdot|A| \Rightarrow \frac{1}{|A|}\frac{d}{d\alpha}|A| = \frac{1}{|A|}\cdot|A|\cdot Tr\left[A^{-1}\frac{d}{d\alpha}A\right]$]