Homework 1 Report - PM2.5 Prediction

學號：r06521605
系級：土木所電輔組碩一
姓名：許舜翔

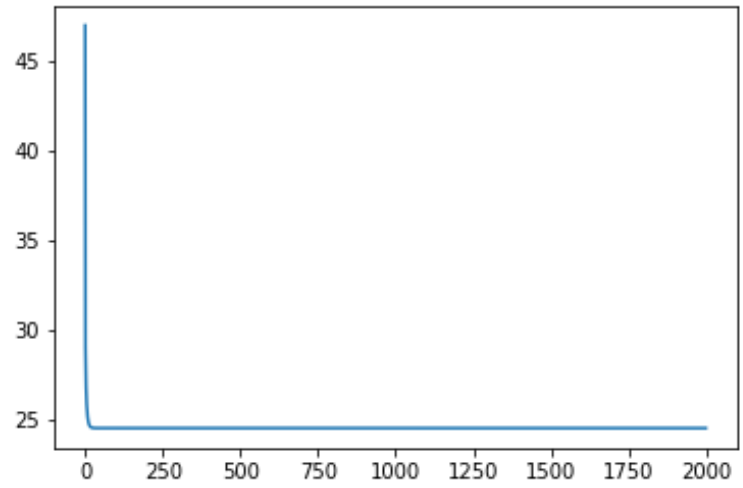
1.

選用不同 *feature* 產生的 *RMS*		
項目	9 小時內所有 *feature* 的一次項 (含 bias 共 163 項)	9 小時內 *PM2.5* 的一次項 (含 bias 共 10 項)
RMS	9.00	9.55

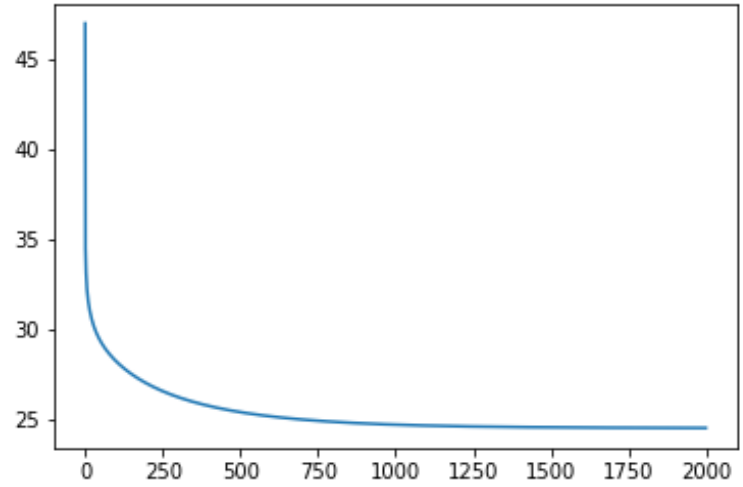
很明顯的可以看出前者的表現較後者為佳，推測可能的原因如下：

- *PM2.5* 值的大小並非一線性連續變化，受到該時段不同環境因素的影響較大，後者較不能反應現況。
- 訓練用的資料為每隔一小時一筆，所以有可能是因為資料間隔過大，其中環境因素變動有起伏，導致 *PM2.5* 值的線性變化預測產生較大的誤差。

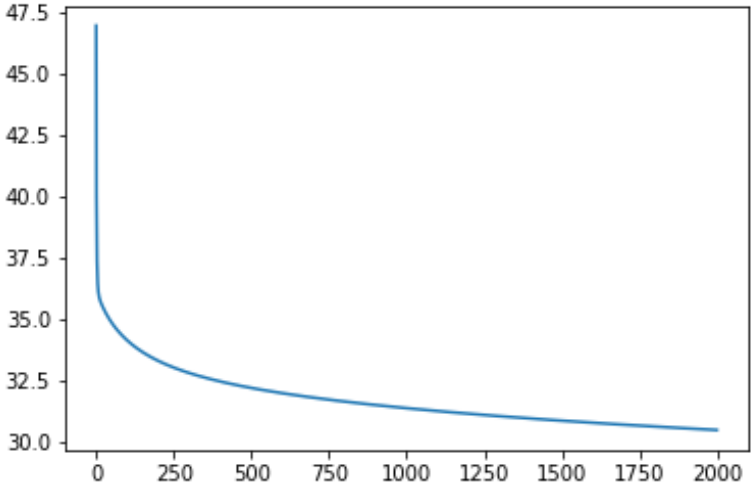
2. 下列為以 *best_hw1* 的模型為基礎，用四種不同數值的 Learning rate 來進行訓練，並觀察收斂情形 ($x = Time/10$, $y = Cost$)。
可以觀察到最後 Cost 應會收斂於 25 左右，而 Learning rate 即是代表收斂的速度；收斂的速度為前幾次最為劇烈，原因應為一開始的假設 (皆為 0) 與實際相差過多，導致初期算出的 gradient，其值皆較大，就算僅以 0.0123 的學習速率訓練，前幾次的變化斜率仍近乎垂直。



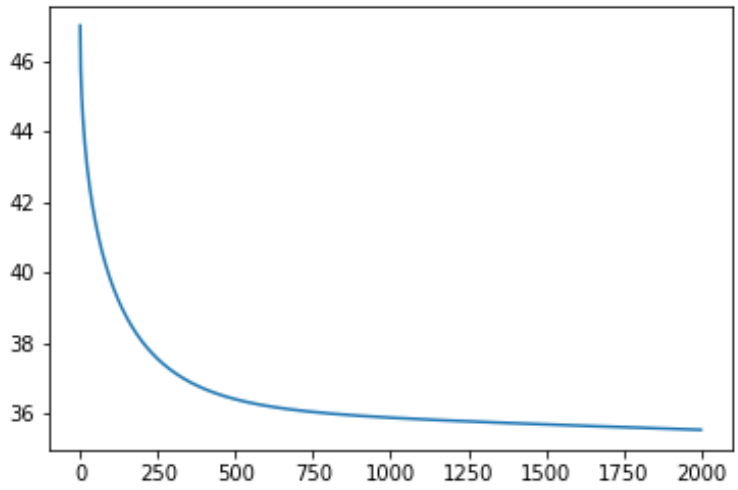
Learning rate = 12.3



Learning rate = 1.23



Learning rate = 0.123



Learning rate = 0.0123

3. Regularization 旨為避免 overfitting 的問題，以下為嘗試不同的 λ 值所得到的 *RMS*（固定 Learning rate = 1.234；Iteration time = 50000）：

λ	Root mean square
0	8.32042
0.1	16.06688
0.5	19.95342
1	21.82169

可以發現隨著 λ 的增加，*RMS* 也隨著增加，推測原因可能是使用的 model 皆為一次項，本來就不會有 Overfitting 的問題，所以增加限制反而讓 ω 無法順利達到最適解。

4. 以下將以針對「Data Preprocessing」、「Features 的選用」、「訓練相關參數」等問題來描述我的 *best_hw1* 是如何實作：

- **Data Preprocessing**
  為了消彌各參數範圍不一造成的影響，所以對各參數進行了 feature scaling，使用後發現模型都能有些微的精度提升（RMS 下降約 0.05~0.1）；另外，我也利用 PM2.5 的平均及標準差，將平均正負兩倍標準差外的資料視為 outlier，將其排除於訓練資料，以提升預測的準確性（RMS 下降約 0.4 左右）。

- **Features 的選用**

  直覺上認為時間隔得越遠，其環境因素應影響越小，甚至影響可能已經內含在 PM2.5 值之中，故僅取前一小時各變因的一次項為 feature，進行訓練後便得到不錯的結果。

- **訓練相關參數**

  在這次的作業中，訓練次數與學習速率參數的改變，僅影響收斂的速度，對精度的提升並無太大的影響，我想一部分應是我假設的模型複雜度不夠高，故算出來 Gradient 的值較單純，不會有卡住、過度震盪或無法盪到谷底等情形發生。