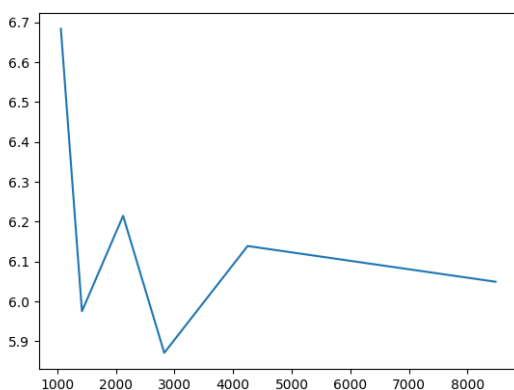


1. 請簡明扼要地闡述你如何抽取模型的輸入特徵：

我自己嘗試幾個組合後現 PM2.5 的一次方與二次方是必要的特徵，而後上網看了幾篇關於 PM2.5 與空氣汙染各項指標的論文，得知 PM2.5 與某些指標有正相關，總共使用的特徵有 PM2.5、PM2.5 二次、臭氧、二氧化硫、風速與風向。

2. 請比較不同同訓練資料量對於 PM2.5 預測準確率的影響：

Training Sample 太小時非常不準確，過大則可能產生 overfit。



RMSE	Training Samples
6.0493129839337669	8478
6.1388976229518928	4248
5.8709375501400443	2826
6.2147114305226951	2124
5.9752134848927341	1422
6.683849101403073	1062

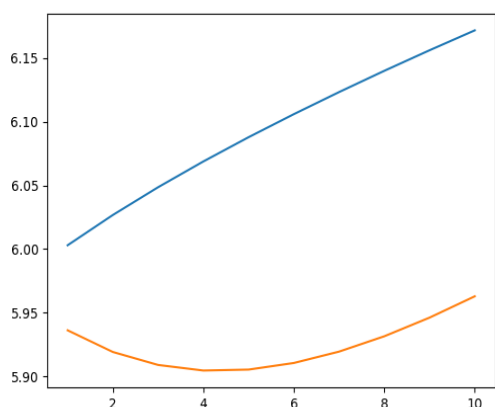
3. 請比較不同複雜度的模型對於 PM2.5 預測準確率的影響 (只用 PM2.5 當特徵)：

最合適本題的模型為二次方模型，三次方以上則會產生嚴重的 overfit 效果而使 error 極大。

RMSE	Degree of x
6.744687732631686	1
6.1985963329715972	2
20.218352868663867	3
629.96590015398954	4

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

正規化與 training error 為正相關是可以理解的，因為 lamda 會懲罰過大的參數，而使收斂速度不如 lambda = 0 的情況。而取適當的 lambda 值則會使 test error 降低，應為解決部分 overfit 的關係，過大的 lambda 則會因收斂太慢而與 training error 一樣逐漸增高。



RMSE (train)	RMSE (test)	Lambda
6.0030082233927775	5.936220242653742	1
6.026789562016841	5.919179236760379	2
6.0486003601768257	5.9089968628631624	3
6.0688438969640606	5.9046113248041454	4
6.0878895791335621	5.9053588396800754	5
6.1059981006618571	5.9105396976281588	6
6.1233451772738396	5.9194712043259914	7
6.140050025715122	5.9315313260943929	8
6.1561954692486962	5.9461741359912512	9
6.1718408426735101	5.9629302995164704	10

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 $\sum_{n=1}^N (y^n - wx^n)^2$ 。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 ：

Let $E = \text{Error}$

$$\sum E^2 = E^T E = (y - Xw)^T (y - Xw)$$

Error 最小值出現在微分等於零

$$\frac{d}{dw} (y - Xw)^T (y - Xw) = 0$$

$$-2X^T (y - Xw) = 0$$

$$X^T y = (X^T X)w$$

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