

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

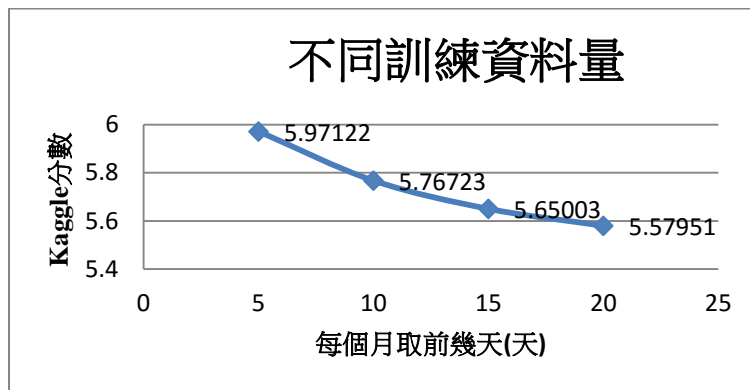
答：

我是取前九小時 PM2.5 的資料當作 feature，其中前兩小時我有額外加入平方項。經測試得知其他數據對我的 kaggle 分數並無太多影響，為了降低計算複雜度故將其省略。

```
ip = np.array([1]+train_data[n:n+9]+[train_data[n+7]**2]+[train_data[n+8]**2])
```

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

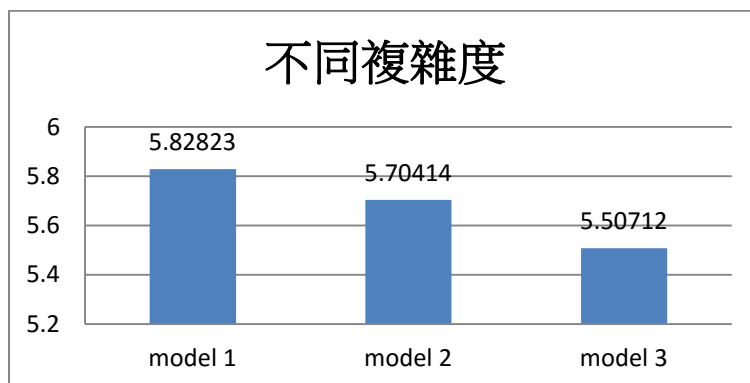
答：



可以看出取的天數有明顯的變化，但變化幅度越來越小。此外因為我是用 stochastic 的方式去做 linear regression，所以天數越少分數的 variance 也越大。

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：



Model 1：前 9 小時 PM2.5 一次項。

Model 2：前 9 小時 PM2.5 一次項加前一小時二次項。

Model 3：前 9 小時 PM2.5 一次項加前兩小時二次項。

可以看出加二次項對於找 model 有很顯著的優化，只加最後一小時則可能會讓這個係數 overtrained，需要有另一個二次項稍微平衡。

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

答：

在這次的作業中，因為我的 feature 只使用前 9 小時的 PM2.5 和前 2 小時 PM2.5 的平方項，所以項數並不多的情況下，小的正規化係數(lb)對於 kaggle 測試結果影響不大，但如果 lb 過大，則會造成效能嚴重下降，可能原因是因為他壓縮 weight 自由的空間，而由變大的 bias 取代。

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

依照我們設定的 model 跟 loss function 可以把這個題目轉換成下列式子：

$$\min_w (y - X^T w)^T (y - X^T w)$$

右式展開可得到:

$$y^T y - y^T X w - w^T X^T y + w^T X^T X w$$

將其對 w 偏微分得到

$$-2X^T (y - X w)$$

使其=0 得到

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