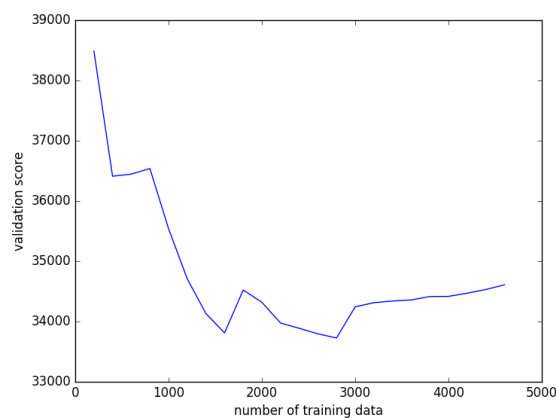


1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：每個月的資料從第 10 筆 pm2.5 資料開始往後每一筆資料和前 9 筆資料作為一個 **vector**。後來有嘗試將當下資料和前一個小時資料相減得到差值，總共會再多 9 筆資料，這個動作好像對最後結果有些許提升。

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

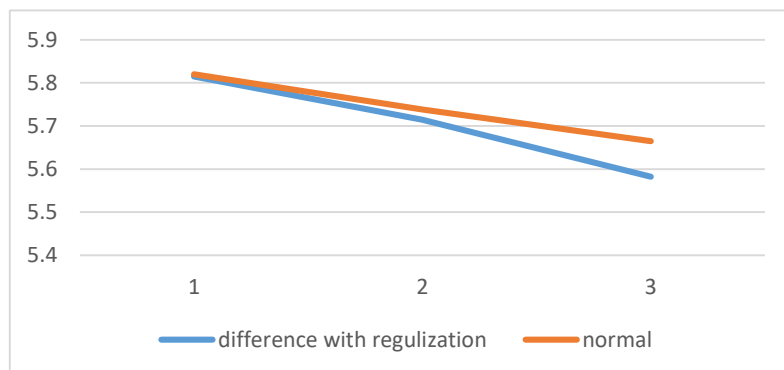
答：為了能夠讓結果快一點，只採用了 **first order linear mode**。每次測試分別增加 200 筆資料。停止條件是 $\text{gradient} \leq 0.001$ 。並切出了 1000 筆資料作為 **validation data**。



從圖上可以看出，基本趨勢是 **validation score** (**validation data loss**) 不斷下降，可能是剛開始由於數據太少而出現 **overfitting**。儘管中間有一些波動，我趨向認為是 **variance** 的問題導致的。而在圖最後的尾部，因為再多的數據也沒有辦法讓模型更好，而趨於平緩。

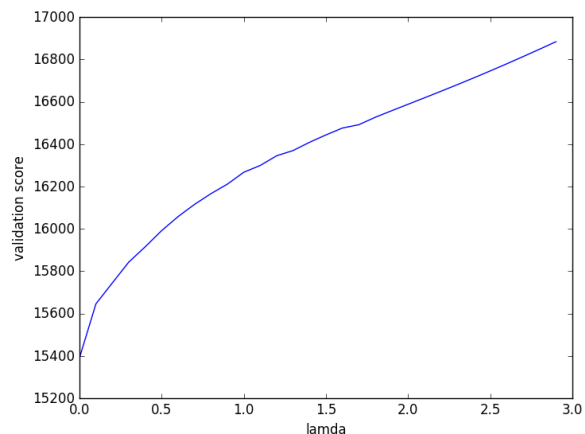
3. 請比較不同複雜度的模型對於 PM2.5 預測準確率的影響

答：由於 **kaggle** 中的 **test** 可以有 **validation** 的效果，因為上傳了各種不同模型的分數，於是直接用 **kaggle** 中 **test** 來作為結果。圖中兩條線，橘色代表直接用三種模型 (**first order, second order, third order**) 來實做。藍色代表加上這個小時 **PM2.5** 和上一小時數值的差值 ($x_i - x_{i-1}$)，當然，有加上些許的 **regularization**。

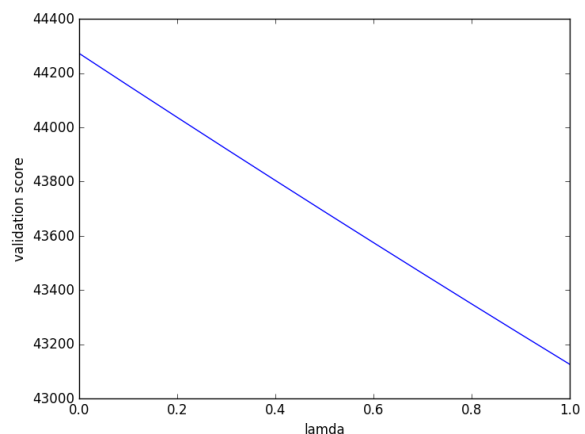


4. 請討論正規化(regularization)對於 PM2.5 預測準確率的影響

答：這裡是利用 second order model 做實驗，預留 500 筆資料作為 validation data，從 0-2.9 每隔 0.1 利用 8000 次 iteration 產生一個 model，分別畫出。發現的問題就是。在沒有 overfitting 的情況下，加正規化是沒有意義的，反而會讓整個模型越來越差。



下圖用的是 third order 模型。是利用一樣的測試資料，但是 lamda 最多只到 1，間隔 0.1（由於跑一個 model 需要很多時間）由於這每個 model 只跑了 4000 次，所以分數比較差，但是可以看到 lamda 在這裡的作用就體現出來了。我承認這線好直，但是做出來的結果就是這樣，我也不知道要怎麼解釋。



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：

$$\nabla F = 2(X^T X w - X^T y) = 0$$

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