

# Machine Learning HW1 Report

學號：B06901045 系級：電機三 姓名：曹林熹

備註：

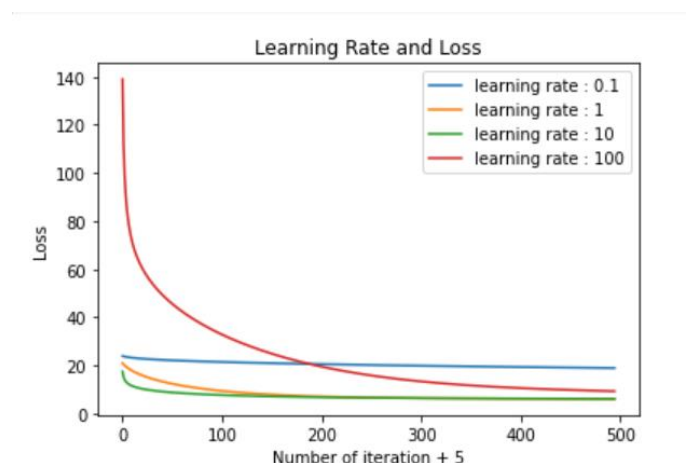
- 1~3 題的回答中，NR 請皆設為 0，其他的數值不要做任何更動。
- 可以使用所有 advanced 的 gradient descent 技術（如 Adam、Adagrad）。
- 1~3 題請用 **linear regression** 的方法進行討論作答。

- (2%) 使用四種不同的 learning rate 進行 training（其他參數需一致），作圖並討論其收斂過程（橫軸為 iteration 次數，縱軸為 loss 的大小，四種 learning rate 的收斂線請以不同顏色呈現在一張圖裡做比較）。

ANS:

在這裡我們給出了四種不同的 learning rate 對 training loss 的比較，分別為 0.1、1、10、100，而每一條線都有對應的顏色顯示在圖中。可以看到一開始  $lr = 100$  時的 loss 最高，但是經過一次次的 iter 訓練後，loss 漸漸小於  $lr = 0.1$  的，而  $lr = 0.1$  雖然一開始 loss 就很小，但是因為  $lr$  太小，因此在後期也無法學習使得 loss 下降。 $lr = 10$  與 1 在最後的結果其實是差不多好的，可以得出結論是  $lr$  不能夠太大或太小，才能使模型進入比較好的 minimum。

（圖中真實的 iter 次數代表 x-axis 上再加 5，目的是使作圖較好觀察比較，ex. 圖中  $iter = 0$  實際上是  $iter = 0 + 5 = 5$ ）



- (1%) 比較取前 5 hrs 和前 9 hrs 的資料 ( $5 \times 18 + 1$  v.s  $9 \times 18 + 1$ ) 在 validation set 上預測的結果，並說明造成的可能原因

(1. 因為 testing set 預測結果要上傳 Kaggle 後才能得知，所以在報告中並不要求同學們呈現 testing set 的結果，至於什麼是 validation set 請參考：

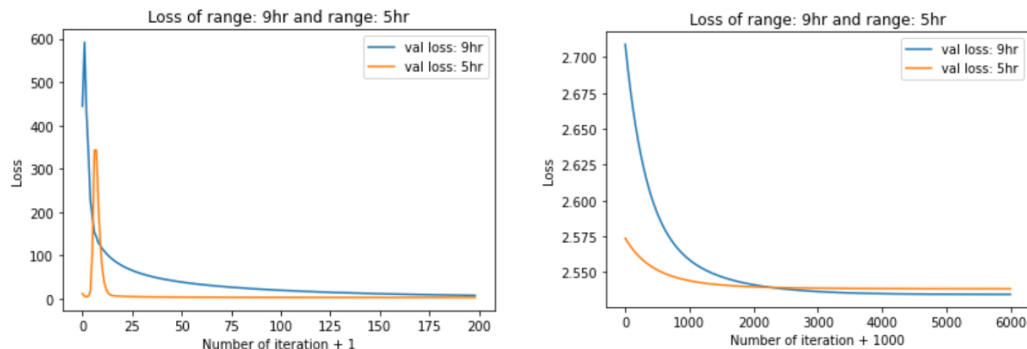
[https://youtu.be/D\\_S6y0Jm6dQ?t=1949](https://youtu.be/D_S6y0Jm6dQ?t=1949))

(2. 9hr:取前 9 小時預測第 10 小時的 PM2.5；5hr:在前面的那些 features 中，以 5~9hr 預測第 10 小時的 PM2.5。這樣兩者在相同的 validation set 比例下，會有一樣筆數的資料)

ANS:

首先可以看到左圖，選取 9hr 的 loss 會比 5hr 的 loss 還要大，但是經過一次次 iter 後，可以看到趨勢慢慢接近。接著看到右圖，可以看到 9hr 大約在 iter = 3200 後的 loss 表現比 5hr 更好，可以推論出因為 9hr 能夠考慮到更多的時間範圍，雖然一開始 loss 比較大，但在長期下的表現是比較好的。最後補充說明，在左圖中可以看到一個明顯的峰值，我推論是 model 不小心跳出了 local minimum 而造成的 loss 區域極大值，但是經過一次次的梯度下降後，又使 loss 變小了。

(左圖真實的 iter 次數代表 x-axis 上再加 1，右圖真實的 iter 次數代表 x-axis 上再加 1000，目的是使作圖較好觀察比較)

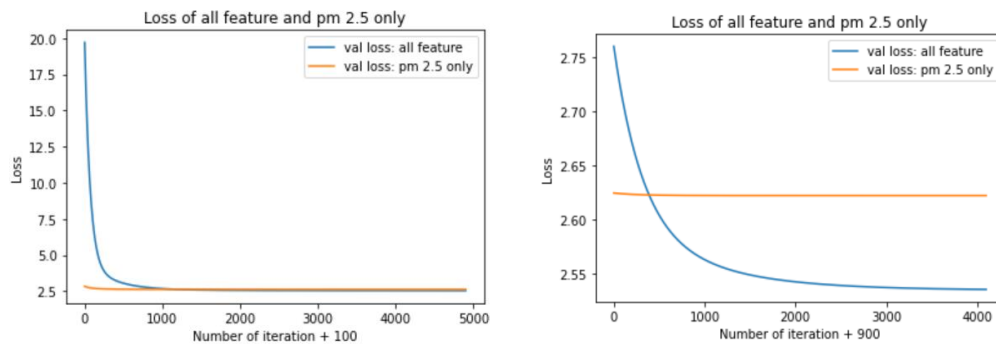


3. (1%) 比較只取前 9 hrs 的 PM2.5 和取所有前 9 hrs 的 features ( $9 \times 1 + 1$  vs.  $9 \times 18 + 1$ ) 在 validation set 上預測的結果，並說明造成的可能原因。

ANS:

首先看到左圖，可以看到從一開始只使用 pm 2.5 作為 feature 的 model 相當穩定，而使用所有 feature 是慢慢地 loss 變小。接著藉由右圖觀察細部，可以看到我們使用所有 feature 成功的比 pm 2.5 only 的模型 loss 較小。這個例子跟上述差不多，都是探討 feature 數目與模型表現的關聯，在這個例子如果單純只用 pm 2.5 訓練會使得很多變因無法去微調 ex. 某種空氣的含量等等，因此還是 feature 較多會比單一 feature 來得好。

(左圖真實的 iter 次數代表 x-axis 上再加 100，右圖真實的 iter 次數代表 x-axis 上再加 900，目的是使作圖較好觀察比較)



4. (2%) 請說明你超越 baseline 的 model(最後選擇在 Kaggle 上提交的) 是如何實作的 (例如：怎麼進行 feature selection, 有沒有做 pre-processing、learning rate 的調整、advanced gradient descent 技術、不同的 model 等等)。

ANS:

實際上，因為此次任務比較簡單，因此我們可以根據第一題調整 lr 或是更改較大的 iter 使我們的模型學習較好的參數而過 strong baseline。除此之外，經過第二題與第三題的探討，可以得知我們可以藉由除去多餘的 fearutes 使模型找到影響預測結果的關鍵因子，讀者可以多試試不同的方法而改善模型學習。