學號：b05901168　系級：電機四　姓名：陳冠豪

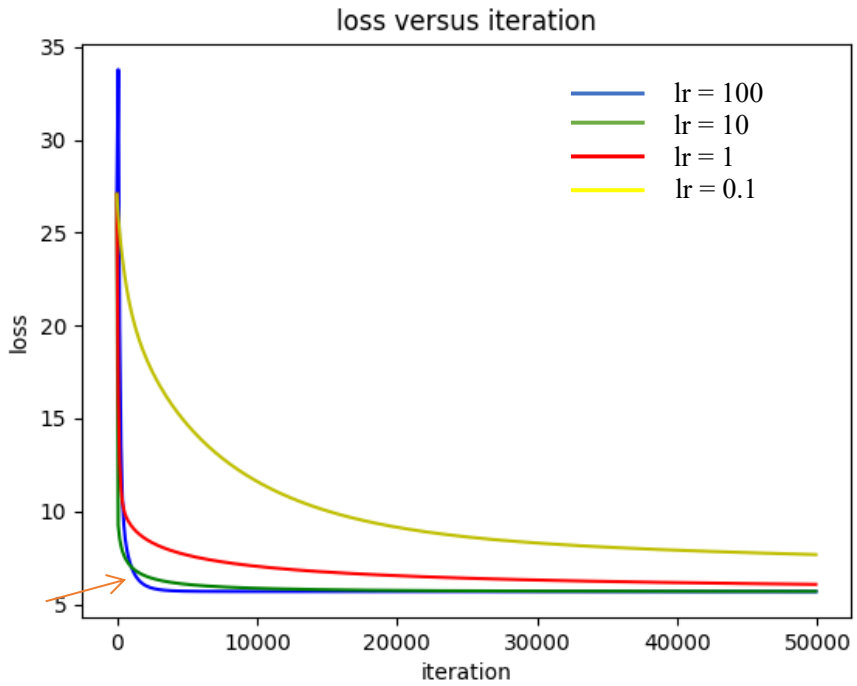
備註：

a. 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。

b. 可以使用所有 advanced 的 gradient descent 技術（如 Adam、Adagrad）。

c. 1~3 題請用 linear regression 的方法進行討論作答。

1. (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）。



由圖可以知道整體來說 learning rate 值較大，收斂會比較快，但也不是把 learning rate 設越大越好，如果所示，learning rate 100 和 10 的圖形有交點

2. (1%) 比較取前 5 hrs 和前 9 hrs 的資料（5*18 + 1 v.s 9*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 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）。

	Training set	Validation set
5 hrs	5.860806	5.676258
9 hrs	5.699620	5.652984

可以看出在 training set 和 validation set 裡,，用 9 hours 得到的結果是比較好的，但是在 validation set 兩者的差距沒有 training set 明顯，可能的原因是因為用 9 hours feature 結果所在的 function set 是比較大的，因此可能有 overfitting 的可能

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

	Training set	Validation set
All features	5.699620	5.652984
PM 2.5	6.193260	5.862648

可以看出在 training set 和 validation set 裡,，用 all feature 得到的結果是比較好的，但因為目前使用的模型是線性的，因此如果用了某些不相關的 feature 反而可能造成結果變爛，所以選擇時可以嘗試用部分的 feature。另外在 validation set 兩者的差距也沒有 training set 明顯，可能的原因是因為用 all feature 結果所在的 function set 是比較大的，因此可能有 overfitting 的可能。

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）。

在 pre-processing 的部分，我將每個 feature 各自做 normalize，並且作 normalize 後每個 feature 的圖，圖中藍色是 pm2.5 的曲線，其餘為各 feature 的曲線，可以依照圖片來做 feature 選擇，另外因為有些數值會很明顯地不一樣，因此我將每個資料跟前一個資料做平均，讓曲線更平滑。
Learning rate 因為很快就收斂了，所以選多少都差不多，我自己是用 100，gradient descent 的部分則是採用 Adagrad