學號：r07921001 系級：電機所碩二 姓名：李尚倫

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）。

iters: 34000 , train\_loss: 5.6800454206538085, valid\_loss: 5.42356064369309 min!

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）。

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。

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）。

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6062037/