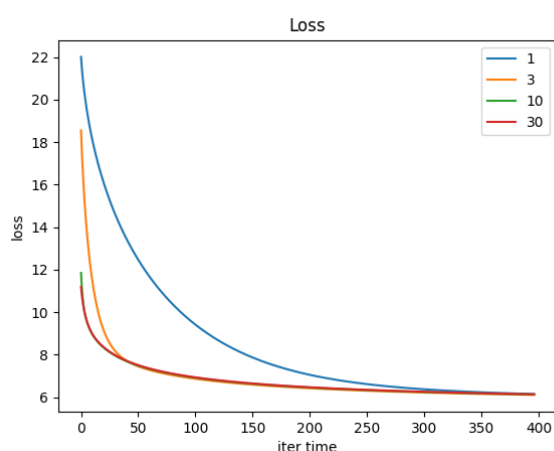


學號：B07901029 系級：電機二 姓名：陳君輔

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）。



learning rate 越大，越快衰減到極值(loss \approx 5.8)

*learning rate 太大時前幾個 iteration 的 loss 會太大，故從 iter time=3 開始作圖。

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)。

隨機取 10%的資料當 validation，結果：

9hr：loss = 6.131

5hr：loss = 6.168

另外在 kaggle 上測試的結果是

9hr：loss = 5.45

5hr：loss = 5.66

只取 5 小時預測的效果較差，或許是因為 input 太少導致 underfitting 的情況。

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。

只取 PM2.5 : loss = 6.20

取 7 個最相關的 feature : loss = 6.09

取所有 feature : loss = 6.18

兩種情況都不太優，取所有 feature 會參雜太多不相關的參數，稀釋掉其他相對重要的參數，僅參考 PM2.5 則會捨棄掉一些重要的參數。

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)。

以 corrccoef 選擇與 PM2.5 較相關的參數進行 training。並加入 regularization 抑制 W 的過度成長。