

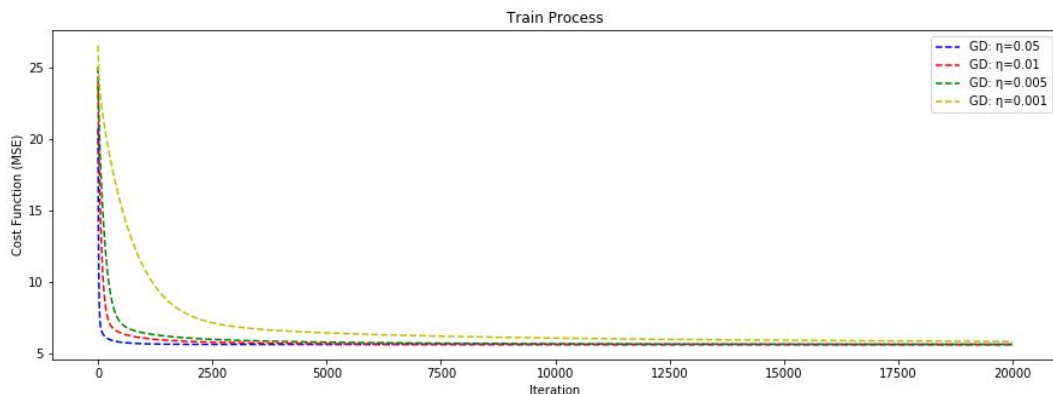
備註：

- 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。

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）。

Ans:

這邊嘗試的 learning rate 分別是 0.05/0.01/0.005/0.001, 結果看起來初始的 loss 都差不多，但 learning rate 越大的收斂越快，但 iterations 數過多的話這次看起來都會收斂到相近的地方



2. (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 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:

固定條件 GD 搭配 learning rate: 0.05, iterations: 2000
結果是只用 5 小時的訓練，在 Validation set 上表現會比較差（rmse 高），最終使用 9 小時會收斂在 5.830199，使用 5 小時會收斂在 6.223241
可能原因是考慮的時間段變少對於整體預測能力下降

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:

固定條件 GD 搭配 learning rate: 0.05, iterations: 2000
結果是只用 5 小時的訓練，在 Validation set 上表現會比較差 (rmse 高)，最終使用 9 小時會收斂在 5.830199，使用 5 小時會收斂在 6.123273
當只用 PM 2.5 下去訓練的話會很快就收斂，但最終的 loss 會比較高，feature 太少可能有 underfitting 狀況發生

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:

1. 將所有負的數據設為 0，因為所有數據在合理情況都應該是正的
2. 使用 Adagrad, Learning rate 設為 1.27, Iteration 設為 20000
3. 以上只能 Test error 達到 5.41 ~ 5.42
4. 最後在 Test 時將結果四捨五入，可以達到 5.38 ~ 5.39