

# Machine Learning - Homework 1

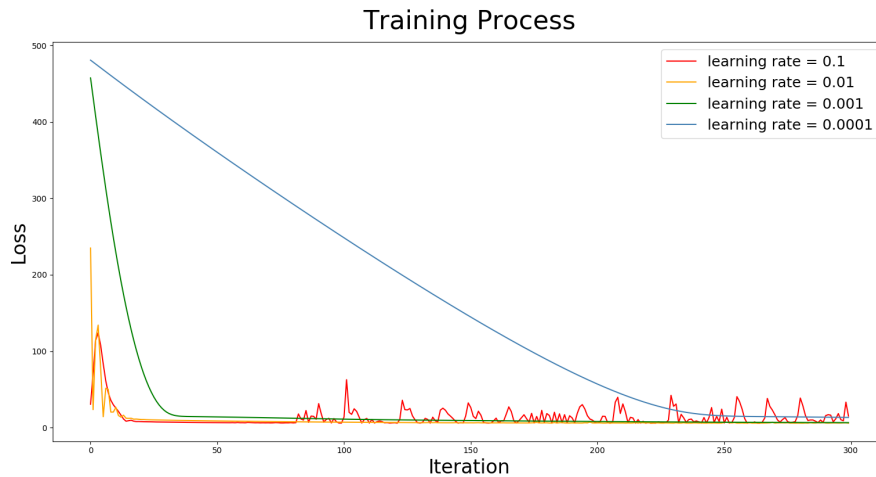
資工四 B05902023 李澤諺

March 26, 2020

備註：

- 1 到 3 題的回答中，NR 請皆設為 0，其它的數值不要做任何更動。
- 可以使用所有 advanced 的 gradient descent 技術（如：Adam、Ada-grad）。
- 1 到 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 的收斂曲線請以不同顏色呈現在一張圖裡做比較）。



在上圖中，我並未進行任何的 data preprocessing 以及 feature transformation，並且，我實作了 Adam 進行 linear regression，其中 batch size 為 1024，learning rate 分別為 0.1、0.01、0.001、0.0001，並且訓練了 300 個 epoch，以此得出上圖的結果。由上圖可以看出，當 learning rate 為 0.1 時，可能是因為 learning rate 太大，導致 loss 的曲線震盪非常大，並且隨著 iteration 的次數越多，loss 並沒有隨之收斂，當 learning rate 為 0.01 時，可能是因為 learning rate 還是偏大，因此 loss 的曲線在 iteration 很少時震盪很大，但是當 iteration 的次數越多時，loss 也開始隨

之收斂，當 learning rate 為 0.001 時，其可能為 learning rate 合適的大小，因此 loss 的曲線會穩定下降並收斂，而當 learning rate 為 0.0001 時，loss 的曲線也會穩定下降並收斂，不過可能是因為 learning rate 偏小，因此此時 loss 下降的速度較 learning rate 為 0.001 時慢了許多。

2. (1%) 比較取前 5 小時和前 9 小時的資料 ( $5 \times 18 + 1$  v.s  $9 \times 18 + 1$ ) 在 validation set 上預測的結果，並說明造成其結果可能的原因 (1. 因為 testing set 預測的結果要上傳到 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，validation set 請參考：[https://youtu.be/D\\_S6y0Jm6dQ?t=1949](https://youtu.be/D_S6y0Jm6dQ?t=1949))(2. 前 9 小時：取前 9 小時預測第 10 小時的 PM2.5；前 5 小時：在前者的 features 中，以第 5 到 9 小時預測第 10 小時的 PM2.5。這樣兩者在相同的 validation set 比例下，會有一樣數量的資料)。

以下的表格中，我選出了 500 筆資料作為 validation dataset，剩餘的資料則作為 training dataset，其中，我並未進行任何的 data preprocessing 以及 feature transformation，並且，我實作了 Adam 進行 linear regression，其中 batch size 為 1024，learning rate 為 0.001，並且訓練了 2000 個 epoch，以此方式重複訓練了 100 次，並將這 100 次訓練所得到在 validation dataset 上的 loss 取平均值，以此得出以下表格的結果：

Feature	Validation Loss
$9 \times 18 + 1$	6.01896
$5 \times 18 + 1$	6.26412

由此可以看出，僅考慮前 5 個小時的資料作為 feature 時，其在 validation dataset 上所得到的 loss 較高，可能是因為每個時間點的 PM2.5 數值會受到不只前 5 個小時的資料所影響，因此僅考慮前 5 個小時的資料作為 feature 可能不足，考慮前 9 個小時的資料作為 feature 時才能提供足夠的 information，使得其 performance 較好。

3. (1%) 比較只取前 9 小時的 PM2.5 和取所有前 9 小時的 features ( $9 \times 1 + 1$  vs.  $9 \times 18 + 1$ ) 在 validation set 上預測的結果，並說明造成其結果可能的原因。

以下的表格中，我選出了 500 筆資料作為 validation dataset，剩餘的資料則作為 training dataset，其中，我並未進行任何的 data preprocessing 以及 feature transformation，並且，我實作了 Adam 進行 linear regression，其中 batch size 為 1024，learning rate 為 0.001，並且訓練了 2000 個 epoch，以此方式重複訓練了 100 次，並將這 100 次訓練所得到在 validation dataset 上的 loss 取平均值，以此得出以下表格的結果：

Feature	Validation Loss
$9 \times 18 + 1$	6.01896
$9 \times 1 + 1$	8.82554

由此可以看出，僅使用 PM2.5 的資料作為 feature 時，其在 validation dataset 上所得到的 loss 較高，可能是因為每個時間點的 PM2.5 數值不是只會受到之前的 PM2.5 數值所影響，因此僅考慮 PM2.5 的資料作為 feature 可能不足，考慮前 9 個

小時的所有資料作為 feature 時才能提供足夠的 information，使得其 performance 較好。

4. (2%) 請說明你超越 baseline 的 model (最後在 Kaggle 上選擇提交的 model) 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)。

我選出了 500 筆資料作為 validation dataset，剩餘的資料則作為 training dataset，其中，我並未進行任何的 data preprocessing 以及 feature selection/transformation，並且，我實作了 Adam 進行 linear regression，其中 batch size 為 1024，learning rate 為 0.001，並且訓練了 2000 個 epoch，以此得出最終的 model，其所得到的 loss 如下：

Train	Validation	Test	
		Public	Private
5.81514	5.75666	5.39004	7.41801