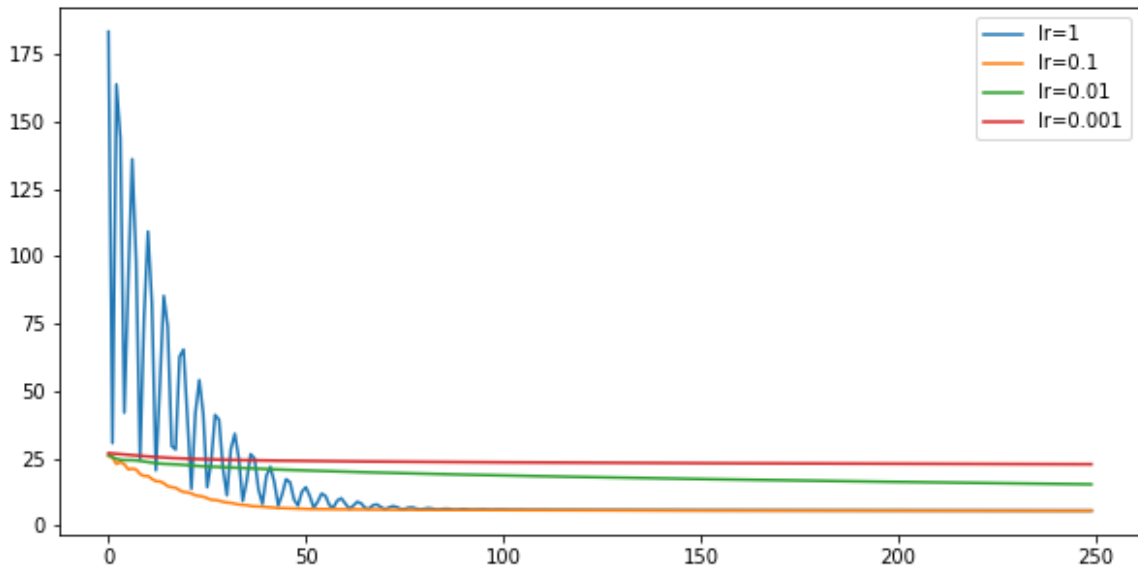


備註：

- a. 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）。
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）。



大的 learning rate 快速下降但出現震盪的情形，小的 learning rate 則是穩定下降，但速度較慢。

2. (1%) 比較取前 5 hrs 和前 9 hrs 的資料 ($5 \times 18 + 1$ v.s $9 \times 18 + 1$) 作預測的結果 (將資料分成 training 和 validation 兩組)，並說明造成的可能原因。

(Adam, 10000 rounds, lr=1)	5hr	9hr
Train	5.984722753038824	5.415194776163903
Validation (30%)	5.848697597685268	6.715638740121492

可以觀察到 5hrs 有比較大的 train loss，可是有比較小的 validation loss。原因可能是 5hrs 的模型 Bias 比較大但 Variance 較小，而 9hrs 則相反，是比較強（複雜）的模型，但 Variance 較高，出現了一點點 Overfitting 的現象。

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. $9 \times 18 + 1$) (請將 training 資料分成 training 和 validation 兩組)，並說明造成的可能原因。

(Adam, 10000 rounds, lr=1)	PM2.5	All
Train	6.127778709570307	5.694086132758031
Validation (30%)	6.399143109820692	5.927839194954463

可以觀察到 PM2.5 那一組的 train 和 validation loss 皆高於 All 那組的兩個 loss。原因可能是因為 All 那組的模型 Bias 遠小於 PM2.5 那組，也就是說只用 PM2.5 訓練出來的模型有一點點 Underfitting 的現象。換言之，只用 PM2.5 這個 feature 不太足以預測之後的 PM2.5。

4. (2%) 請說明你超越 baseline 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) 。

- 使用原本 feature 的子集 : ['CH4', 'CO', 'NO', 'NO2', 'O3', 'PM2.5', 'RAINFALL', 'RH', 'SO2', 'WIND_SPEED', 'WS_HR', 'PM10']
- 沒有做 normalization
- 使用 gradient descent with adagrad as optimizer, lr=0.01
- Public test set : 5.39831

Reference: b07902064