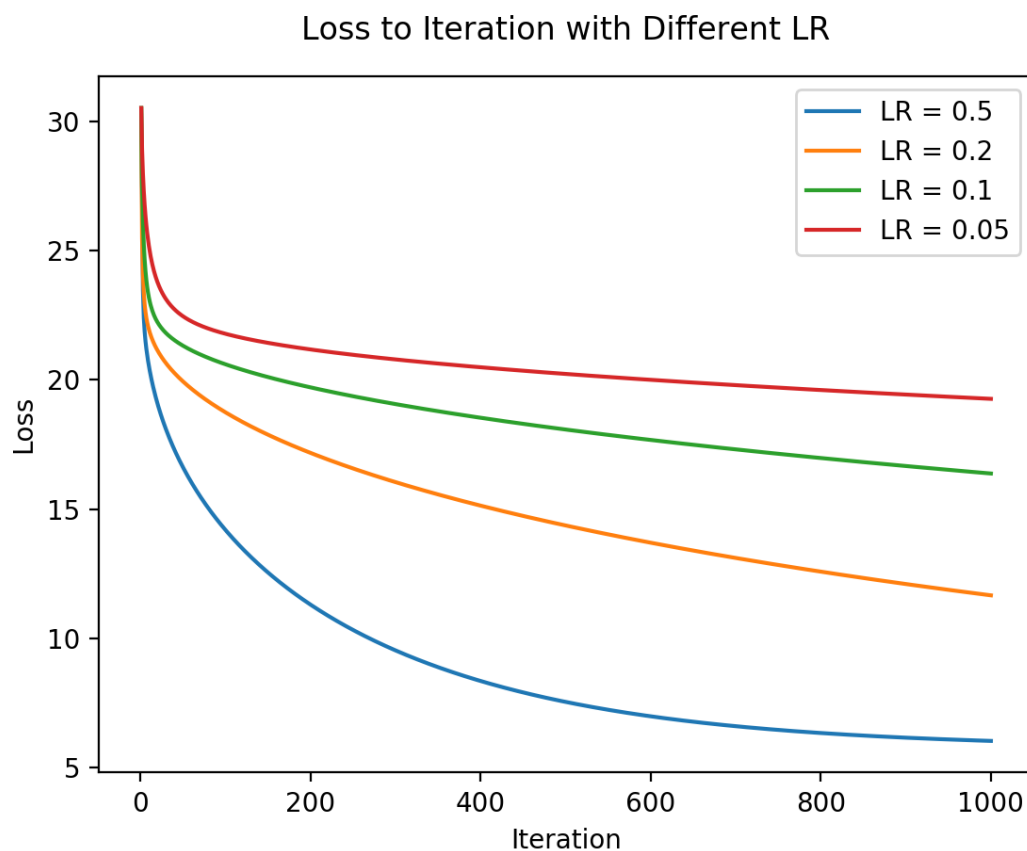


備註：

- 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 ) 。



由圖可知Learning Rate越大，收斂過程前段的斜率絕對值也越大，但可發現LR = 0.5，Iteration 超過200時已逐漸到達漸近線，故斜率趨緩。推測其他三條在Iteration足夠大時也有此現象。

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 (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，至於甚麼是validation set 請見：[https://youtu.be/D\\_S6y0Jm6dQ?t=1949](https://youtu.be/D_S6y0Jm6dQ?t=1949) ) 。

```
[(base) acde-MacBook-Pro:hw1-chinyi0523 chinyi0523$
```

#### Five Hours

Train set loss at Iter 10000	5.86232
Validation set loss	5.67301

```
(base) acde-MacBook-Pro:hw1-chinyi0523 chinyi0523$
```

#### Nine Hours

Train set loss at Iter 10000	5.72019
Validation set loss	5.6653

參考前九小時稍較參考前五小時的 Validation Set Loss 低，推測PM2.5受長期波動的影響仍較大，僅考慮短暫五小時波動容易造成誤差。

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。

```
[(base) acde-MacBook-Pro:hw1-chinyi0523 chinyi0523$
```

#### All Features

Train set loss at Iter 10000	5.72019
Validation set loss	5.6653

#### Only PM 2.5

Train set loss at Iter 10000	6.19403
Validation set loss	5.8611

僅考慮PM2.5時無論Training Loss 或 Validation Loss都較差，可知PM2.5需要倚靠其他Feature來預測，與PM10、風向等等都有極高的相關性，可見下題的附圖說明。

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 )。

(1)將-1都補為0

(2)切除7月資料，容易受到颱風等劇烈天氣影響

(3)使用 sklearn.feature\_selection import SelectKBest, chi2 判斷每個Features的重要程度，如下圖 ( 採用 <https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e> 的方法 )

```
[3.48683919e+02 4.11351927e+00 1.43570717e+02 5.03022937e+01  
3.85411972e+02 5.35416816e+03 4.88739608e+03 9.67834736e+03  
5.65344405e+04 6.97480523e+04 1.72074994e+03 1.39757152e+03  
1.08583783e+03 1.45433003e+01 2.04115619e+04 1.43339133e+04  
6.98036507e+01 7.36654130e+01]
```

其中第10項為PM2.5，其值最大符合常理，但可發現PM10 ( 第9項 )、WD\_HR ( 小時風向值 ) ( 第15項 )、WIND\_DIREC ( 第16項 ) 的數量級皆與PM2.5相同，亦可佐證為何只取PM2.5表現較差。測試刪減各項相關性低feature發現刪除1-5、11、12項效果較佳。

(4)設batch降低overfit的可能性

(5)將風向、小時風向分別取cos和sin取代原本角度，降低角度帶來的誤差，例如359或1度皆指北方，但相差極大。

測試失敗：

Adam without batch momentum，可能需要再仔細思考為何效果不佳