

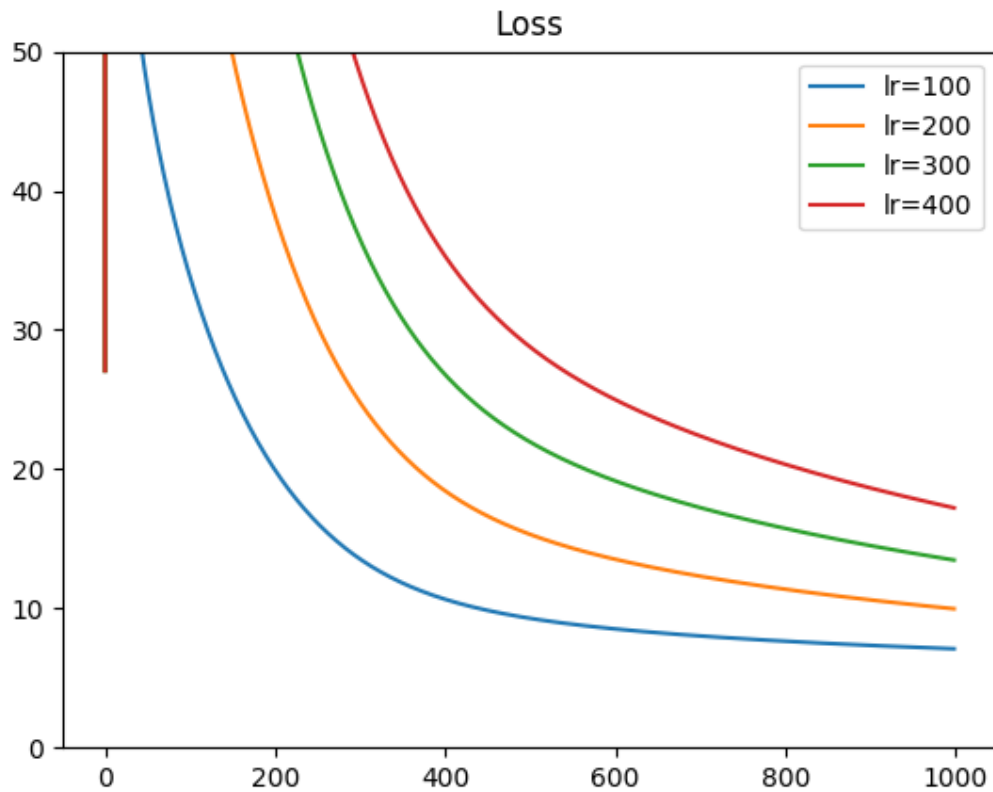
學號：b06901153 系級：電機三 姓名：林瑩昇

Collaborator：b06901180 電機三 鄭謹譯

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 愈大，在 loss 方面收斂的很比較慢，也比較不容易收斂至最佳解

下圖的是用 adagrad、learning\_rate = [100, 200, 300, 400]、iteration = 1000



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。

如果是只用前5 hrs來預測的話，因為參數量不夠多，導致最後loss收斂的地方比較大。可以看出只用5 hrs的方式在validation set上的錯誤也比用9 hrs的大。

下圖是兩種方式做出來的loss以及在validation set上的比較

用的是Adam( $\beta_1 = \beta_2 = 0.99$ )、 $\text{learning\_rate} = 0.005$ 、 $\text{iteration} = 10000$ 、 $\text{split\_ratio} = 0.8$

5 hr cut	9 hr cut
Training loss: 5.86081	Training loss: 5.71675
Error of validation set:5.67633	Error of validation set:5.66857
mean(ans, val)=(22.23032,21.50614)	mean(ans, val)=(22.11504,21.40318)
std(ans, val)=(14.50916,15.61204)	std(ans, val)=(14.32964,15.43373)

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。  
原因與 2 相似，因為這麼做model的複雜度太低，沒有足夠的參數作預測。

下圖是兩種方式做出來的loss以及在validation set上的比較

feature: [9] 代表只取用index = 9的那個參數，也就是PM2.5

feature: [0~17] 代表全部的參數都用了

用的是Adam( $\beta_1 = \beta_2 = 0.99$ )、 $\text{learning\_rate} = 0.005$ 、 $\text{iteration} = 10000$ 、 $\text{split\_ratio} = 0.8$

```
features: [9]
Training loss: 6.19403030
Error of validation set:5.86110
mean(ans, val)=(21.40053,21.40318)
std(ans, val)=(14.37767,15.43373)
features: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]
Training loss: 5.71675755
Error of validation set:5.66857
mean(ans, val)=(22.11504,21.40318)
std(ans, val)=(14.32964,15.43373)
```

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) feature selection :

我挑選的feature是index為 [5, 6, 7, 8, 9, 10, 11, 14, 15, 16, 17] 的幾個feature，我與collaborator先用sklearn找出最相關的幾個feature，再配合一些domain knowledge，選出了這些feature來進行training。結果不管是在validation set或是testing set上都有顯著的進步。

(2)pre-processing :

我之前在看training data的時候，發現有許多奇怪的極值，像是溫度會有不連續的現象，後來去氣象局網站看才知道只要沒有量測到的項目，就會變成-1，其餘也有許多項目的值不正常，像是單位為ppm的項目竟然有負數。我在想這可能是data的一些小錯誤，所以我添加了一個smoothing function會把這些極值用內差的方式消除。雖然在testing set上，加了smoothing的結果會比不加稍微退步，可是在validation set上卻有顯著的進步，所以最後我還是決定把smooth後的版本當final score

(3)learning rate

原本用adagrad時，learning rate調到5就已經收斂很快了，但卻收斂不到最佳解。換成了adam後，learning rate可以調到0.05，為了能夠收斂到更好的解，我每5000個iteration就會把learning rate除二。

(4)Adam

參考作業說明中的影片，我的adam選用的參數 $\beta_1 = \beta_2 = 0.99$ ，相較於adagrad而言，他收斂的更快，而且很快就到最佳解。所以我可以很快知道到底是不是learning rate的問題，或是model本身的問題。

(5)Month selection

與同學討論有發現經過作圖後，7月的data趨勢很不一樣，我們在想可能是因為台灣7月有很多颱風，所以我們嘗試把7月的data去掉進行training，結果在validation set上表現非常好，但在kaggle上卻無法過strong baseline。我們討論的結果是在testing set上也有不正常(有颱風)的data，如果我們把7月去掉，雖然在其他月的預測結果會很好，卻也連帶讓model預測7月PM2.5的能力大幅下降，所以error就會暴增。