1. Regression:

(a)(b):

Architechure:

Activation function: Leaky Relu (slope=0.1 when x<0)

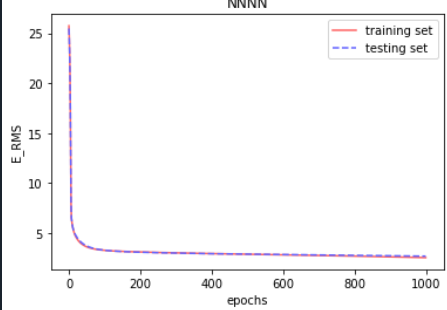
Width: [16,48,24,12,1]

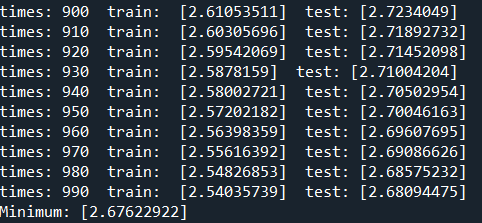
Learning rate:0.002

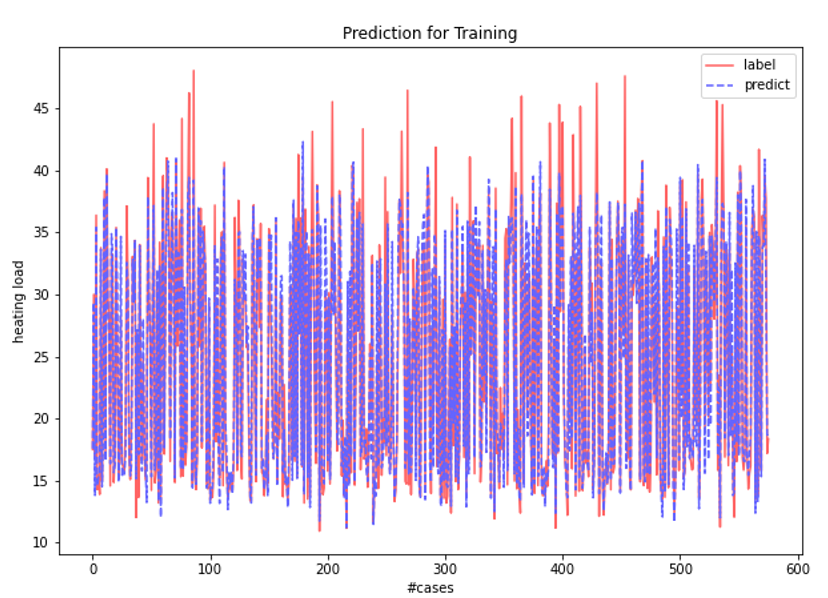
Normalization: standard normalization on input

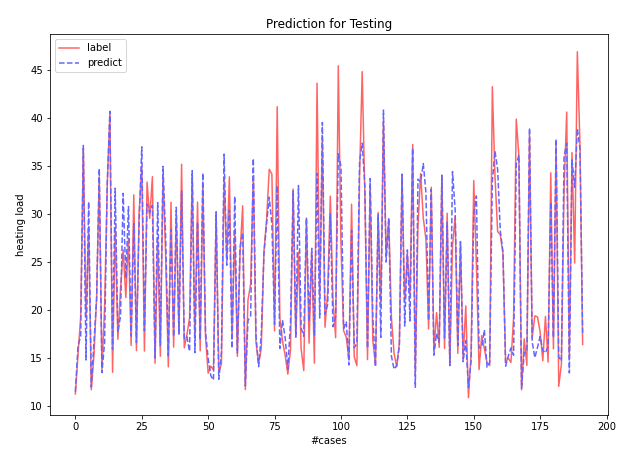
Features:0~7

One hot: on “orientation” and”Glazing area distribution”



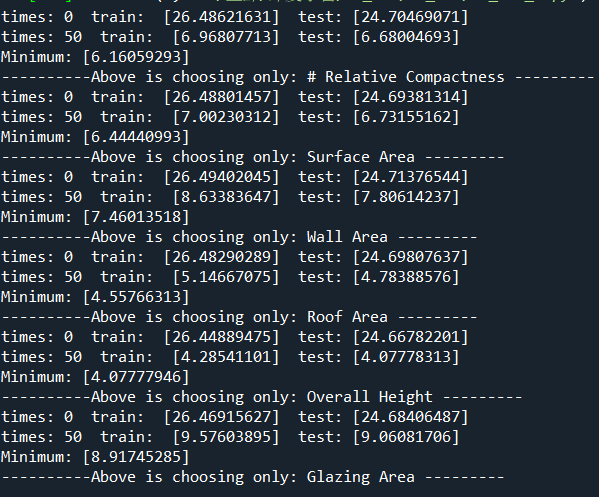


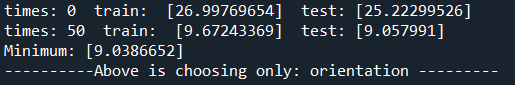


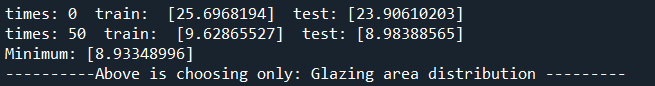


(c )feature selection

Using same data sequence, and using same initial weights and bias

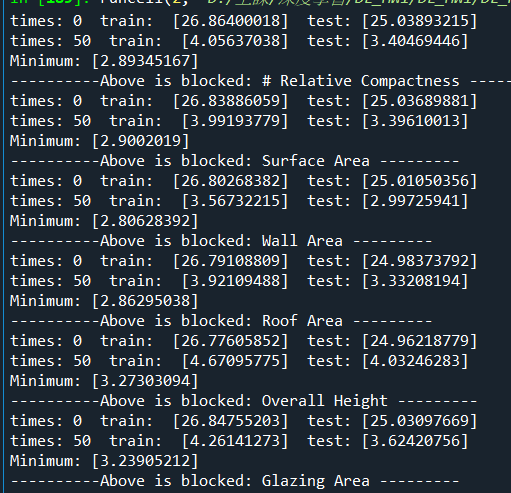


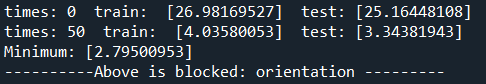


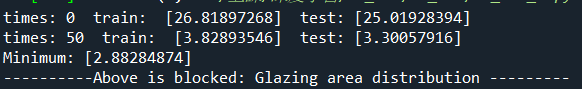


If choosing only one features, choosing “Overall Height” is a good choice

But “orientation” might not be a good choice







Thus, if it’s needed to eliminate one feature. prefered choosing “Surface Area”or”orientaion”

But eliminating”Overall Height” or “Glazing Area” might not be a good choice

By evaluating from two aspects, we found “Overall Height” is more related to target and having more effect on minimizing loss function

1. Classification

(a)(b):

Architecture:

Activation function: Leaky Relu (slope=0.1 when x<0), softmax at output layer

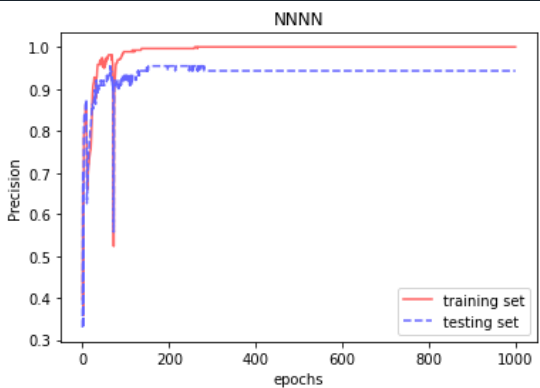
Width: [33,132,3,2]

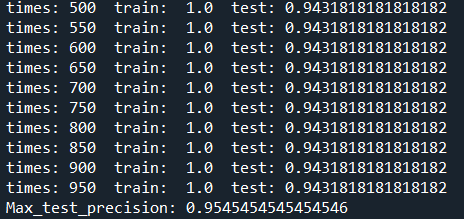
Learning rate:0.23

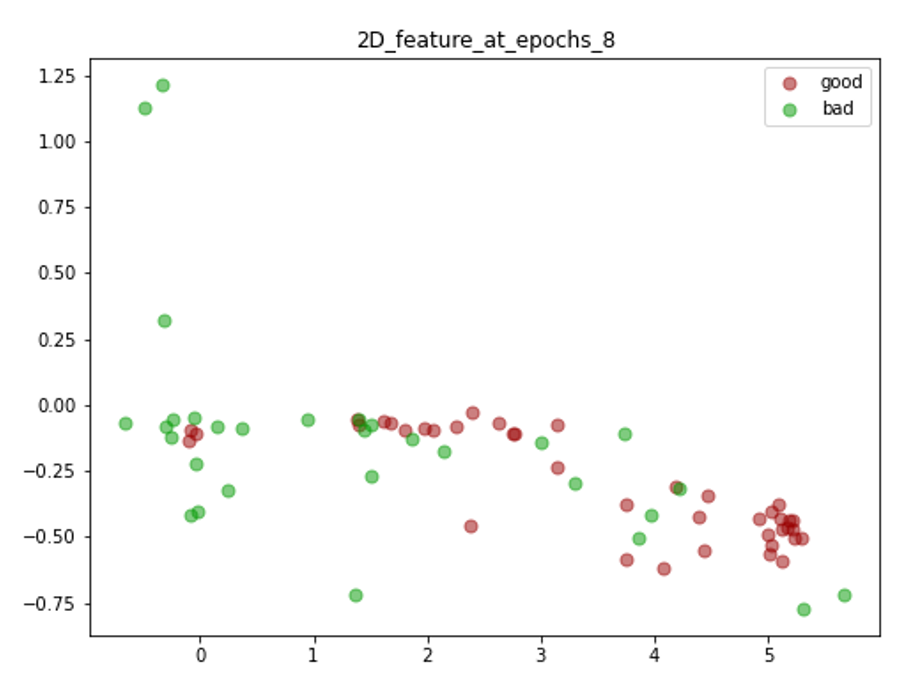
Normalization: delete second feature(all zero)

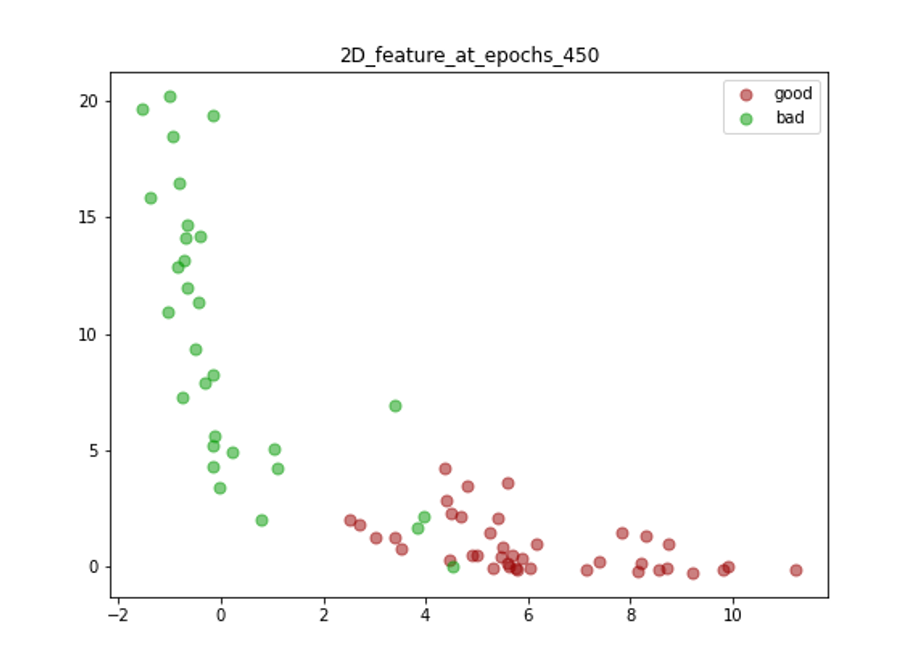
Features:0~34

One hot: on target good=[1,0] bad=[0,1]









心得:在開始自己implement一個MLP網路之前都沒有意識到會遇

到的問題有多少，而且感覺architecture方面很多的問題都是case by case網路上寫的很多優化不一定適合用到個別的問題上。

**以下是我遇到的問題:**

1. normalize data:如果沒有的話，會產生太大的gradient導致所有weight bias都變得過大(正或負)導致下個epoch產生更大的gradient惡性循環。而且除了避免上述情況之外，合理的normalized還能讓各個feature比較公平競爭。然而我在做regression的時候把target也normalized卻產生很大的RMS，所以target可能還是不要處裡。
2. learning curve:自己在操作learning rate,跟layers，常會出現各種形狀的learning curve。通常learning curve很大一部分的程度展現了learning rate的情況，像是抖動的形狀就是lr太高，還有比較少見的就是像打勾一樣的形狀，大概也是lr太高準備要抖動，但epoch不夠長沒看出來，只是我還是不太懂classification幾乎每次plot都會出現的突然下墜是為什麼。
3. architecture: activation方面，leaky relu的效果相較sigmoid好，在Regression的時候還感覺差不多，但在進行classification的時候就能明顯感覺到sigmoid收斂慢，而且precision也較低。

initial weight跟bias取值在-0.5~0.5之間，體感比0~1之間好，除了一開始就有正有負外，也比較可以避免所有weight跟bias累加到output的時候出現過大的數值。