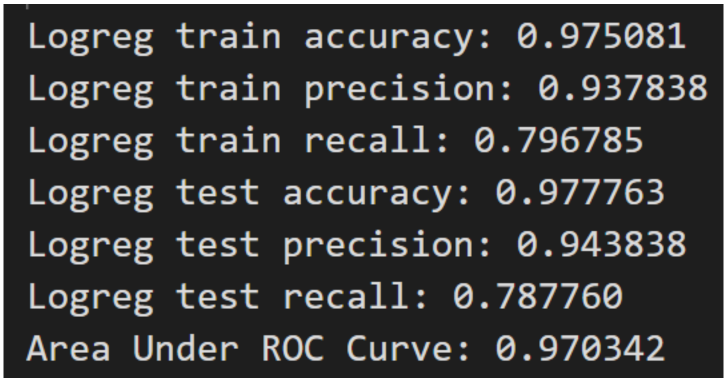
Log_Regression_SGD

110505259 陳柏燊

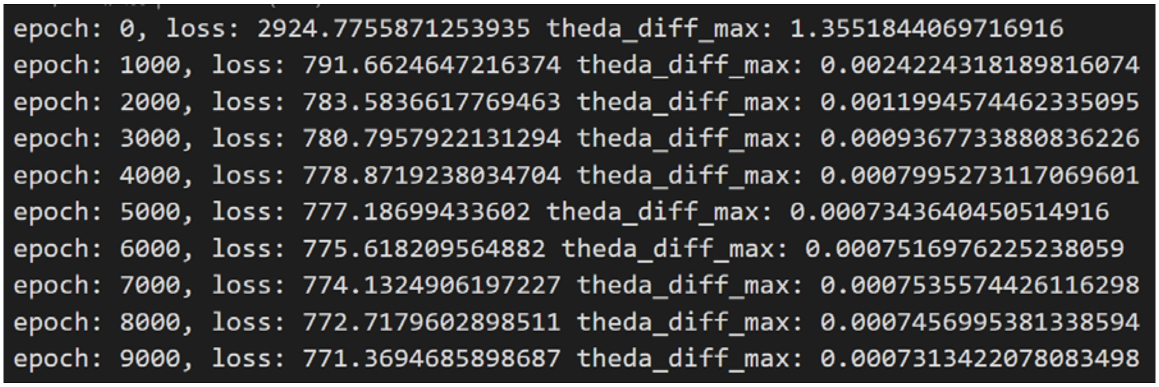
Accuracy and precision



```
Logreg train accuracy: 0.975081
Logreg train precision: 0.937838
Logreg train recall: 0.796785
Logreg test accuracy: 0.977763
Logreg test precision: 0.943838
Logreg test recall: 0.787760
Area Under ROC Curve: 0.970342
```

The effect of different parameters

Alpha = 0.01, Max_epochs = 10000, eps = 1e-4

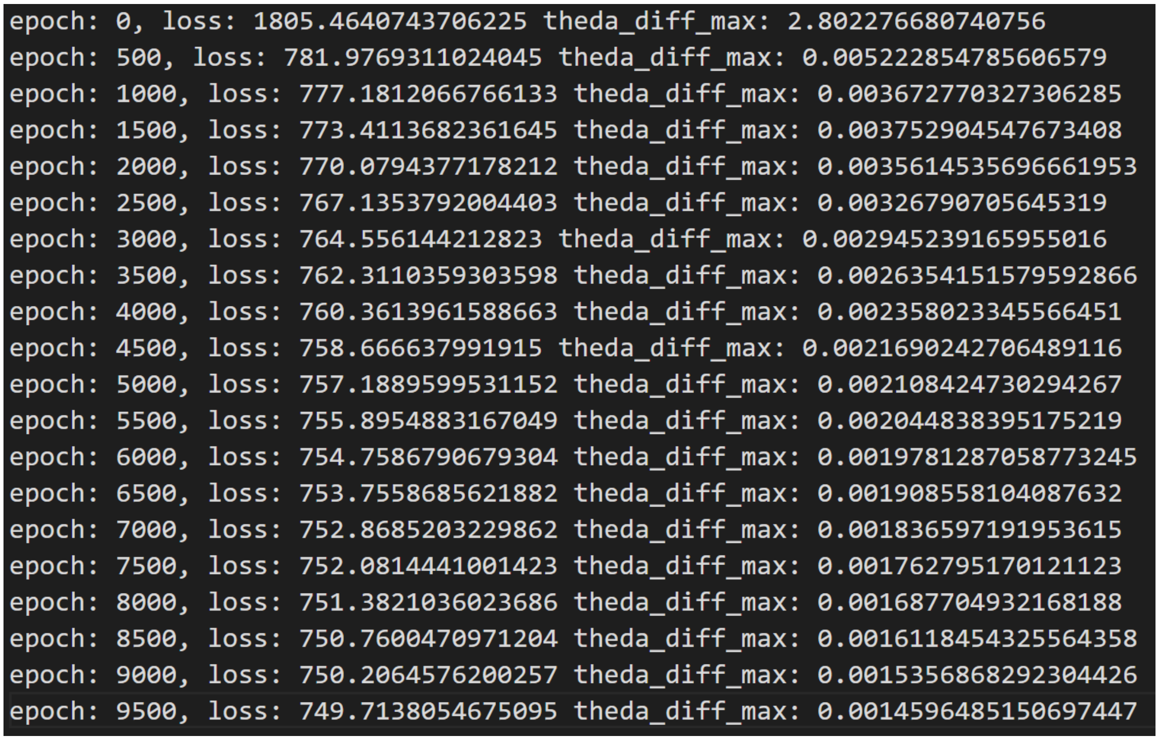


```
epoch: 0, loss: 2924.7755871253935 theda_diff_max: 1.3551844069716916
epoch: 1000, loss: 791.6624647216374 theda_diff_max: 0.0024224318189816074
epoch: 2000, loss: 783.5836617769463 theda_diff_max: 0.0011994574462335095
epoch: 3000, loss: 780.7957922131294 theda_diff_max: 0.0009367733880836226
epoch: 4000, loss: 778.8719238034704 theda_diff_max: 0.0007995273117069601
epoch: 5000, loss: 777.18699433602 theda_diff_max: 0.0007343640450514916
epoch: 6000, loss: 775.618209564882 theda_diff_max: 0.0007516976225238059
epoch: 7000, loss: 774.1324906197227 theda_diff_max: 0.0007535574426116298
epoch: 8000, loss: 772.7179602898511 theda_diff_max: 0.0007456995381338594
epoch: 9000, loss: 771.3694685898687 theda_diff_max: 0.0007313422078083498
```

We compare eps and epochs together.

We can find that eps and epochs both decide how many rounds should we update our Theda. We can see if we set eps to 8e-4, we will stop at near epoch 4000, and if we set epochs to 1000, we will stop at epochs 1000 stop. IF we stop earlier, we can see we will have higher los on train data, but if we encounter overfitting, stop earlier will help to solved it.

Alpha = 0.05, Max_epochs = 10000, eps = 1e-4



```
epoch: 0, loss: 1805.4640743706225 theda_diff_max: 2.802276680740756
epoch: 500, loss: 781.9769311024045 theda_diff_max: 0.005222854785606579
epoch: 1000, loss: 777.1812066766133 theda_diff_max: 0.003672770327306285
epoch: 1500, loss: 773.4113682361645 theda_diff_max: 0.003752904547673408
epoch: 2000, loss: 770.0794377178212 theda_diff_max: 0.0035614535696661953
epoch: 2500, loss: 767.1353792004403 theda_diff_max: 0.003267907056453l9
epoch: 3000, loss: 764.556144212823 theda_diff_max: 0.002945239165955016
epoch: 3500, loss: 762.3110359303598 theda_diff_max: 0.0026354151579592866
epoch: 4000, loss: 760.3613961588663 theda_diff_max: 0.002358023345566451
epoch: 4500, loss: 758.666637991915 theda_diff_max: 0.002169024270648911б
epoch: 5000, loss: 757.1889599531152 theda_diff_max: 0.002108424730294267
epoch: 5500, loss: 755.8954883167049 theda_diff_max: 0.002044838395175219
epoch: 6000, loss: 754.7586790679304 theda_diff_max: 0.0019781287058773245
epoch: 6500, loss: 753.7558685621882 theda_diff_max: 0.001908558104087632
epoch: 7000, loss: 752.8685203229862 theda_diff_max: 0.001836597191953615
epoch: 7500, loss: 752.0814441001423 theda_diff_max: 0.001762795170121123
epoch: 8000, loss: 751.3821036023686 theda_diff_max: 0.001687704932168188
epoch: 8500, loss: 750.7600470971204 theda_diff_max: 0.0016118454325564358
epoch: 9000, loss: 750.2064576200257 theda_diff_max: 0.0015356868292304426
epoch: 9500, loss: 749.7138054675095 theda_diff_max: 0.0014596485150697447
```

We compare the 2 pictures above. There are same in Max_epochs and eps but different in alpha. We first look at epoch 0. It shows that the loss with high alpha is lower than which with smaller alpha. I consider that alpha decide the size of steps. So if we move in a bigger step, we will got lower loss quicker than small steps.

Discussion of the results

We can find that the dataset is very incline to 0. We find that label with 1 with just 1639(9%) instances when the all data set is 17898 instances. So it more useful to use ROC curve that PR curve when PR curve would have lots of can’t reach area. I also find another excited thing. We find that since the value set of logReg is from 0 to 1, so the bias of the feature will always be zero, so I think it will cause it is useless to add 1 to all instances. Finally I don’t want to calculate AUC by myself so I all sklearn.mtrics.AUC to help me calculate that. I found that we should do some cleaning o our tpr and fpr to use the function. (sklearn AUC require for one fpr should only have one tpr and fpr should be increasing). That all my discussion.