STAT665 Machine Learning

Bo Song

I. Width of the Network

#nodes	2	8	32	128	512
Accuracy	0.333	0.711	0.694	0.757	0.786

The accuracy increases when nodes increase. It makes sense since more nodes prevent under fitting. In the case with 32 nodes, the accuracy fluctuates and drops a little. It may because it falls into a local optimal.

II. Depth of the Network

512 nodes in each layer

#layers	1	2	3	4	5
Accuracy	0.786	0.761	0.752	0.774	0.746

The accuracy fluctuates and drops down when the depth of layer increases. It's because the model is overfited. The case with 4 layers is abnormal. It may because of randomness.

III. Freezing layer

#layer	1	2	3	4	5
Accuracy	0.784	0.802	0.797	0.783	0.788

The rates are better than that in part II in every stage. It resembles hierarchical model. We could use stacking to combine the model in III and II together

IV. Autoencoder

#nodes	32	128	512	1024
MSE	0.025	0.024	0.024	0.017

The MSE decreases when the number of nodes increases. The pattern in this part is quite similar to that in part I. Model performs better when number of nodes is increased. With more nodes, our model is more complex and contains more information.

V. Autoencoder as pretraining

#layer	2	3	4	5
Accuracy	0.802	0.738	0.689	0.611

This model performs better when there are 2 layers (1024, 512). The model resembles PCA and hierarchical modeling.

VI. Experiment

For this question, we construct a 2 hidden layers neural network. First layer contains 1024 nodes trained by autoencoder. Second layer contains 512 nodes trained normally.

The classification accuracy is 0.806.

Actually, the model we designed here is first two layers of the model we used in part V. The performance prevails in part I to part V. So we simply reran this model and got a better result.