Wen Sheng
STAT 665
April 6, 2016

Problem Set 06

Data Mining and Machine Learning

I.   Width of the Model

The results shows that the classification rate will increase when the hidden nodes number is increasing. This phenomenon is logical because with more nodes in the hidden layer, the output of hidden layer has more units and can contain more information of different input. Also it can be predicted that when the layer node number increased to an extent, the increasing rate will decrease or even the classification rate be decreased as a result of over fit.

Hidden Nodes Number	2	8	32	128	512
Classification Rate	0.333	0.563	0.658	0.762	0.785

II.   Depth of the model

The result shows that layers number of 1 has the best classification rate and increasing hidden layers won't help the classification result. One possible reason is: when having more layers, the information contained in one image will be transformed and distored.

Hidden Layers Number	1	2	3	4	5
Classification Rate	0.778	0.776	0.763	0.776	0.763

III.   Freezing Layers

Compare to question 2's keep decreasing rate, the classification rate here will first increase then decrease when the freezing layers number keep increasing. Because with more layers to an extent (here is 3-4 freezing layers), the architecture of the whole model is changed and previous weight is more and more not applicable to later layers which decrease the classification rate. This approach resembles additive model as one output of the previous layer is treated as the input of the next hidden layer. We can implement a model which have 4 hidden layers with first 3 as freezing layers to achieve the highest classification rate.

Freezing Layers Number	1	2	3	4	5
Classification Rate	0.752	0.793	0.794	0.795	0.785

IV.   Autoencoder

The pattern is the mean squared error will decrease with the increasing hidden nodes number. Because with more nodes in the hidden layer, more information can be represented in the output of this layer and achieve a higher representation of the individual image. Compared to part I, the pattern is reversed as in part I the classification rate is increasing. This opposite result shows that classification rate and mean squared error have a negative correlation relationship.

Hidden Nodes Number	32	128	512	1024
Mean Squared Error	0.0375	0.0345	0.0341	0.0325

V. Autoencoder as pertaining

The result is pretty surprising because with more layers number, the classification rate is decreased which is conversed with my expectation. This resembles the additive layer as the output of the previous layer is treated the input of the next layer.

Freezing Layers Number	1	2	3	4	5
Classification Rate	0.811	0.757	0.608	0.528	0.427

VI. Experiment

Optimization	ADM	Adgrad
Classification Rate	0.791	0.792

My model is composed of one pre-training autoencoder and one freezing layer with parameters tuned as Part V shows this combination can have the highest classification rate. I experienced all the Keras optimization methods and it shows Stochastic gradient descent (SGD) and Adamax will get the low classification rate and Adam/ Adgrad optimization method will get a pretty high classification rate of 0.791 and 0.792