Photometric Identification of Infrared Sources

Machine Learning Project* Phase 3: **Neural Network**

June 19, 2021

1 Introduction

Our data consists of infrared sources spectrums that we were supposed to classify into five classes —YSO, O-rich, C-rich, S-rich, post-AGB. Up to this point, we prepared, processed, and analyzed our data; and then applied different classification models to it. At the current stage, we can build a neural network to model our data.

2 Building the NN Model

After cleaning, encoding, scaling, and performing PCA on our data, we split into a test/train ratio of 20/80. In our dataset, the input is of 22 values —different wavelengths in which intensity is measured. And the output is of 5 values —different categories to which each source may belong. So the input and output layer is of 22 and 5 dimensions, respectively. In our neural network, we are using two hidden layers of 10 and 20 dimension. Then we went through the training step in **keras** using **model.fit**.

(batch size = 800, number of epochs=70)

After 70 epochs the neural network was trained in 14.3 seconds, and the training accuracy reached 99.5 %. You can see the activation functions and other information about each layer of this neural network in Table 1.

*Group members: Mobin Saidy, Reyhane Javadi, Amirhosein Masoudnezhad, Negin Khosravaninezhad

Layer	Nodes	Parameters	Activation Function
1	10	230	Relu
2	20	220	Tanh
3	5	105	Softmax

Table 1: Information about each layer of the NN

3 Measuring the NN Model Performance

Now we will visualize training and validation losses and accuracies.

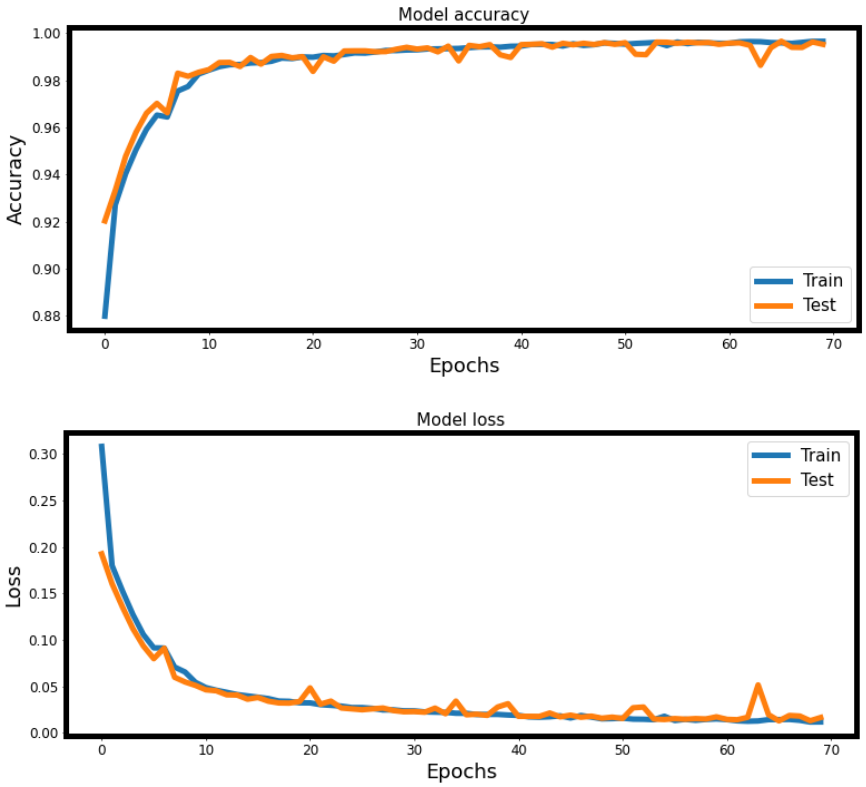

Figure 1: Diagrams of Accuracy and Loss

As for any other machine learning classification problem, we can measure our model performance using Confusion Matrix.

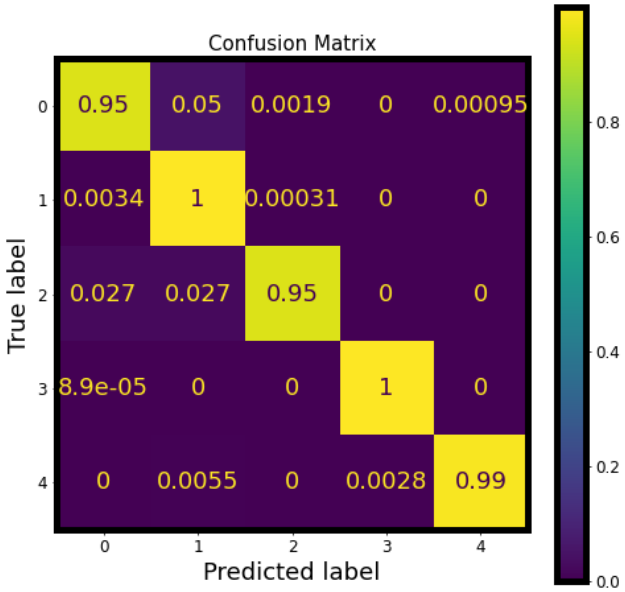

Figure 2: Confusion Matrix

The precision[1], recall[2], and F_1-score[3] calculated for each category from the Confusion Matrix above are as shown in Table 2.

Category	**Precision**	**Recall**	**F_1-score**
C-rich AGB	0.99	0.95	0.97
O-rich AGB	0.98	1.00	0.99
S-rich AGB	0.96	0.95	0.95
YSO	1.00	1.00	1.00
post-AGB	1.00	0.99	0.99

Table 2: Parameters to measure the performance of the NN

This results in an averaged accuracy of 1.00 which means that our neural network has a very good performance in modeling the categories of our infrared sources.

[1]From all the classes we have predicted as positive, how many are actually positive.
[2]From all the positive classes, how many have been predicted correctly.
[3]A measurement of recall and precision at the same time.