Name: Neha Chauhan
College: IIT Kanpur

Solution of Task 1

Supervised learning

Since the dataset was highly imbalanced, I have used Under sampling method to improve the class ratio. Then I have implemented KNN algorithm to learn the model of the dataset. The model gave and f1 score of 0.9 for class 0 (normal) and 0.7 for class 1 (attack). Despite its decent performance the draw back of this model it that it has smaller training dataset than the actual dataset.

Other supervised learning method that can be applied is LSTM as the dataset is based on time series, and LTSM models are good for learning timeseries models.

Unsupervised Learning

I have implemented the approach presented in the reference paper provided.

This approach can be further extended to supervised learning by grouping the data points on the basis of their class (i.e. Normal or Attack) and then learning the class geometric median on the low dimension subspace projection. This will provide the prototype for the Normal system and the idea of the Attack points. At the time of testing the new dataset can be classified based on its distance from the two-class prototype.

1. Why do we need machine learning for intrusion detection?

   Presently there are two kind of approaches in designing an IDS, model-driven approach and data-driven approach. Since the model-driven approach are very complex and modelling the huge number of physical parameters involved in this system is a very difficult task. Hence data driven approach is preferred. Due to the advancement in the field of Machine learning and deep learning many approaches to solve this problem is available and beneficial.

2. Which one works better: supervised, semi-supervised or unsupervised?

   I think semi-supervised learning will be a better approach in this problem as this will learn the underlying model of the datapoints as well as use the labels in the historical dataset.

3. What is your plan during internship?

   I am very fascinated by the various application of the Machine learning and deep learning which has endless possibilities in every field. I have done several course work in Machine learning, Data Analysis and statistics; and now I want to work on cutting edge problems, trying to tackle them with the knowledge I have learned from these courses and also learn new things along the way. I plan on learning and further exploring new techniques in ML field through this internship