For this homework you again work in groups of maximum 6 people. All the rules of previous group homework apply. Please refer to the paper "Open Set Intrusion recognition for fine-grained Attack Categorization" if you need to understand how the data is collected, and what each data vector stands for. The paper is under Files --> HW5 on canvas. However, you do not necessarily use the SVM classifier -- and you can use any classification method at your disposal as long as you can classify between the 4 different categories of attacks.

To understand what you have to do go to http://kdd.ics.uci.edu/databases/kddcup99/task.html

This explains what you need to do.

The dataset is at http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

Note that there is also a 10% subset of the dataset but you are required to work with the full dataset. If you work with 10% dataset, you will be graded for 50% of total marks. The full dataset might take longer to run. You might even need access to departmental servers -- if that is the case, team lead must contact me.

The due date for this will be October 31, 2019 before 11:55 pm as a gzip or zip archive with all the code (only with test datafile, and no other part of the data), and a report with all the classification results, misclassification results etc.

You must take out 25% data before you beginning from the labelled dataset -- and leave it alone while doing your training. You must do your training on 50% data and 25% for validation. If you do cross validation, you will work with 75% of the data -- but please DO NOT use the test data you separate in the beginning for cross validation.

You must submit the test data file (only the 25%) along with your homework inside the zip/gzip file.