

NETW504 Project Task 1

The objective of this task is to discover the data that is to be used for intrusion prediction. The discovery aims at gaining an understanding of the data and gaining a visual insight into the data. The objective is to understand which fields are more relevant in the prediction and to understand the relation between these fields. In this task you are required to do the following

1. Build a python code that would load the data into python and do the following
 - a. List the data fields
 - b. List the types of the data fields (columns)
 - c. Check if there is any missing data or inf data
 - d. Check print the number of categories in each field (column)
 - e. Check and print the Maximum, Minimum , average and variance of each data field
 - f. Divide each data field into 4 quarters (in terms of value) and check and print the Maximum, Maximum, Minimum , average and variance of each data field quarter.
2. Expand the attach type into a number of fields equal to the number of attacks where each field will indicate that this attach happened or not.
3. For each field use python to calculate the PMF or the pdf (depending on the type of data) and plot this pdf.
4. Calculate and draw the CDF for this field
5. For each field use python to calculate the PMF or the pdf (depending on the type of data) Given the attach type and print the original PMF/pdf in blue and the conditional pdf in another color on the same figure.
6. Print scatter plot that indicate the relation between any two data fields
7. Calculate the joint PMF/pdf of any two different fields
8. Calculate the joint PMF/pdf of any two different fields conditioned on the type of attach and find a way to visualize the difference
9. Calculate the correlation between different data fields
10. Find which field are dependent on the type of attack