

Praktikum zu Moderne Methoden der Datenanalyse

Exercise 9: Data Mining Cup: Neural Network

In this exercise we continue with the Data Mining Cup task which was also the subject of the previous two exercises. After testing a cut based approach in exercise 7 and a likelihood based approach in exercise 8 we will use a Neural Network for the classification task in this exercise.

- **Exercise 9.1:**

Take the variables, which you used for the likelihood based approach in exercise 8 and train a Neural Network using the `TMultiLayerPerceptron` class from ROOT (one has to load the `libMLP.so` library before using it). After training, check on the training sample, that the two classes (good and bad customers) are well separated.

When you have trained the Neural Network, use the training sample to find out the best working point (cut on the Neural Network output) for our classification task.

- **Exercise 9.2:**

Compare the classification power between your Neural Network and the likelihood based approach. As there is no a priori reason, why a Neural Network approach should be better than a likelihood based approach, it can happen, that your Neural Network will be worse than your likelihood (or even cut) based approach. In that case, think what could be a reason for that and consult it with the tutor.

- **Exercise 9.3:**

Include the probability that a customer is a good customer from exercise 7 as one of the inputs to the Neural Network and train it again. Again compare its performance with previous solutions on the training sample.

- **Exercise 9.4:**

Try to add more input variables or remove some, which could be insignificant. Also play with the network topology (number of hidden layers, number of nodes in the hidden layers) and check if you can improve the power of your Neural Network.

- **Exercise 9.5:**

Classify the test dataset `class.root` with the Neural Network which you think is your best. Give the classification file to your tutor who will calculate a score for you, so that you can compare the Neural Network based approach with your likelihood or cut based approach.