Homework

Classification

Exercise 1

1. Download and install the framework for machine learning WEKA from:
https://ml.cms.waikato.ac.nz/weka

2. Have a look at the tutorial:
https://waikato.github.io/weka-wiki/primer/

to familiarize with WEKA.

Exercise 2

1. Have a look at the section “Dataset” of the tutorial. Here you will find an explanation about how to prepare the data files containing the training data.

2. Download the following file from Moodle:

`examples.arff`

If you remember from class, in this work we are trying to classify segments composed of a set points in 2D into legs and non-legs of people. The file `examples.arff` contains the features corresponding to each segment together with its classification. The actual label of each segment is defined in the attribute "person," which is 1 for segments corresponding to a "leg of a person”, and -1 for segments corresponding to other objects in the room, i.e “non-leg of a person.” More details can be found in the paper:

> K. O. Arras, O. M. Mozos, and W. Burgard,
> Using boosted features for the detection of people in 2D range data,
> in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 3402-3407, 2007.
> https://doi.org/10.1109/ROBOT.2007.363998

that paper is also available in Moodle:

3. In WEKA, go to the “Explorer” and load the examples.arff file in the “Prepocess” tab. Move to the “Classify” tab. Divide the data into training (70%) and test (30%) using “Percentage split.” Then choose the “bayes/NaiveBayes” classifier and apply it to the data. Using the output from WEKA prepare a confusion matrix for the results using the terms true positives (TP), true negatives (TN), false positives (FP), false negatives (FN). Write the results in percentages.

4. Change the classifier and choose the “meta/Adaboost.M1.” Create the confusion matrix with the results. Do you see any improvement in the results? Please comment the results in terms of TP, TN, FP, and FN, and in terms of global classification results.

5. Now change the classifier to “rules/ZeroR”. This is one of the simplest classifiers. Have a look at the percentage of correct classified and incorrect classified instances. Compare these results with the ones from “bayes/NaiveBayes”. Which classifier is better according to these values?

Despite the results, the “rules/ZeroR” classifier is a bad one, and you cannot trust it. Why? What is this classifier actually doing? (Hint: check the confusion matrix).