

EPART Lab 2 Report

Rustenis Tolpeznikas

November 4, 2024

Point 1

Task:

Check the data, esp. the training set. Outliers can change significantly computed distribution parameters, which can dramatically reduce recognition quality. You can try here to compare *mean* and *median* values, plot histogram of individual features (*hist* function) ...

To remove a sample with known index *idx* use expression:

$$\text{train}(\text{idx}, :) = [] ;$$

Results:

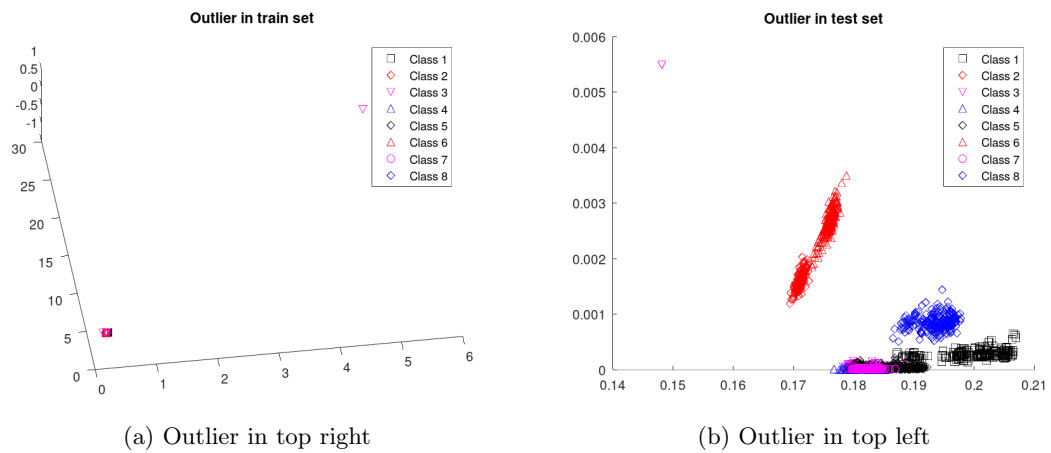


Figure 1: Outliers found

Point 2

Task:

Select two features (note that you have *plot2features* function supplied) and build three Bayes classifiers with different probability density computations (according to points 1-3 above). You should use equal *a priori* probabilities of 0.125.

Results:

Point 3

Task:

Check how the number of samples in the training set influences the classification quality (you can take for example 10%, 25%, 50% of the whole training set).

Note: an appropriate part of the samples from the training set should be drawn independently from each class; because we introduce a random element, the experiment must be repeated (minimum 5 times) and report should contain averaged results (good practice is to include not only mean value but also a standard deviation).

Here you should implement *reduce* function, which leaves the appropriate part of each class. At this point, the reduction applies only to the training set.

Results:

Point 4

Task:

Check how width of the Parzen window h_1 influences the classification quality (note that this point has sense for Parzen classifier only).

Results:

Point 5

Task:

How will the classification results change if the *a priori* probability will be two times higher for black suits, i.e. (0.165, 0.085, 0.085, 0.165, 0.165, 0.085, 0.085, 0.165)?

Note that in this case you should reduce number of red suits in the **testing set** only!

Results:

Point 6

What is the classification quality of the 1-NN classifier (`cls1nn.m`) for these data? Don't use in this case leave-one-out method, you have large enough testing set at your disposal. Think about data normalization. If there is big difference in standard deviations between features you should normalize data before classification.

Task:

Results: