Pattern Recognition Laboratory – Exercise #2
Optimal Bayes Classification

Due date: **26.10.2017 (A), 2.11.2017 (B), 2.11.2017 (C)**

In this exercise, your task will be to prepare Bayesian classifiers, with the different methods of calculating the conditional probability density distributions for each class. You'll be comparing three methods of determining the density :

1. Assuming that the features are independent, and each attribute distribution is normal (in this case the probability density for more than one feature can be calculated as the product of the density for each feature).
2. Assuming that we are dealing with a multi-dimensional normal distribution of the features.
3. Using Parzen window to compute the probability density approximation based on the training set.

Quite common method for approximation of unknown probability distribution is to use the Parzen window. It is based on the fact that an unknown distribution density is "build" on the samples in the training set. Each sample bring small partial density share, placed in the vicinity of sample.
For this exercise, we assume the window function to be multinomial normal distribution. The only parameter we'll have to supply is the window's width h_1 and today you'll check different values in range <0.0001, 0.01>.

Training set for this assignment uses as the features Hu moment invariants (http://en.wikipedia.org/wiki/Image_moment) of the scanned images of cards suits. The first column contains class identifier (4 – spades, 3 – hearts, 2 – diamonds, 1 - clubs). It's worth noting, that suits were printed with different methods: on half of images you can see printing raster and on half there is no such raster. In the training set size of the suit and its rotation angle were changed systematically. In the test set size of the suit can be arbitrary (in the predefined range) and the rotation angle is arbitrary.

Now concrete tasks:

1. 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 sample with known index idx use expression:
   ```
   train(idx, :) = [] ;
   ```
2. 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 0.25.
3. Check how the number of samples in the training set influences the classification quality (you can take for example 1/10, ¼, ½ of the whole training set). You can perform this only for Parzen classifier – for both other classifiers number of samples is not very important as long as the values of mean and covariance matrix does not change substantially.

4. Check how width of the Parzen window h_1 influences the classification quality (again for Parzen classifier only).
5. How will change the classification results if the *a priori* probability will be two times higher for red suits, i.e. (0.17, 0.33, 0.33, 0.17)?
   Note that you should prepare properly **testing set** in this case!
6. What is the classification quality of the 1-NN classifier for these data?
   Don't use in this case leave-one-out method, you have 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.

I expect written report – concise, but containing the important information (I should be able to replicate your results). You should also present me your Octave code used in this exercise.