

METU Informatics Institute

Min720 Pattern Classification with Bio-Medical Applications

Spring 2011

Homework 2

Due Date: 20.4.2011

This homework will allow you to download a prepared medical data set and apply the classification techniques that are already studied in class.

'UCI Machine Learning Repository' is a website with free data to work for pattern classification problems. You will be doing some experiments with 'Breast Cancer Dataset' to result with a binary decision 'malignant' or 'benign'. The data contains features computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image.

Follow the steps below to perform several classification tasks using Matlab or similar software.

1. Find and download the dataset described above '**Breast Cancer Wisconsin (Diagnostic) Data Set**'. Study the features.
2. Assuming that 2 different categories (B: Benign, M: Malignant) are distributed with multivariate gaussian, estimate all parameters using MLE. Use randomly selected 50 samples from each category.
3. Use minimum risk Bayes Classification rule for classification *with* $\lambda_{12}=0.3$, $\lambda_{21}=0.7$. Use 50 samples different than the ones used for learning . Document your results with a confusion matrix.
4. Reduce your data size by using PCA to obtain 10 features and repeat 3 above.
5. Comment on the overall results.