MSPR 7 Evaluation (Due: 18.10.2015, 12 p.m. (noon))

Dr. Hendrik Purwins, Assistant Professor
Jan Stian Banas, TA
Dept. Architecture, Design and Media Technology, Aalborg University Copenhagen
A.C. Meyers Vænge 15, DK-2450 Copenhagen SV, Denmark

1. (Feedback) Please give us feedback on the last lecture and homework: `http://goo.gl/forms/bQqfdxOXWh` Thanks!

2. From the iris dataset take instances 51-75, 101-125 as the training data and instances 76-100, 126-150 as test data for classes *versicolor* and *virginica*. Train a linear discriminant analysis (=minimum Mahalanobis classifier) with the training data set und use the trained classifier to classify the test data.

   (a) Use Matlab and PRTools to calculate the confusion matrix. (20 P)

   (b) What are true positive, false negative, false positive, true negative? (10 P)

   (c) Calculate precision, precision,recall, f-measure, false alarm, accuracy. (10 P)

   (d) Plot the performance graph in the ROC space. Calculate the AUC. (10 P)

3. Based on the `adult` data set from the UCL machine learning repository, predict whether someone is rich ('>50K') or not, based on the following features: `age, fnlwgt, education-num, race, sex, capital-gain, capital-loss, hours-per-week, native country`. Convert the string atrributes race, sex, native-country into a number, white →1, non-white→ 0, male→ 1, female→ 0, US-born→ 1, non-US born → 0, '>50K'→2, <=50K'→ 1. (*not 0,1*as in the previous exercise) Perform 10-fold cross-validation with the classifiers: minimum-distance classifier, minimum Mahalanobis classifier (=linear discriminant analysis), quadratic classifier. Calculate the f-measure. Compare the results of the 3 classifiers to random guess using prior knowledge on the different sizes of the rich and poor class. Comment on the results! (50 P)

4. Self Assessment: Check the exercises that you have seriously worked on.

2 a	2b	2c	2d	3