MSPR 6 Classification (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/Hkk38XKvlI` Thanks!

2. **(Learn how to use PRTools)** Study the prtools manual `http://rduin.nl/manual/` and the prtools cheat sheet `http://www.37steps.com/prtools_cheat_sheet.pdf` (0 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 (0,1), white →1, non-white→ 0, male→ 1, female→ 0, US-born→ 1, non-US born → 0, '>50K'→1, <=50K'→ 0. Use the classifiers: minimum-distance classifier, Minimum Mahalanobis classifier, Quadratic classifier. Keep instances 7467, 10326, 1468, 14346, 29201, 22570, 11016, 10148, 31436, 30489 for testing and train the classifiers with the remaining instances. You can use the Matlab function `setdiff` to generate the indices of the training data given the indices of the test data. You may need the Matlab function `double` to convert a binary/boolean variable into a double variable. Don't worry about the warning: `don't worry about the warning: PR_Warning: GETPRIOR: nmsc: No priors found in dataset, class frequencies are used instead`. (80P)

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

3