

Breakout room exercise Session S12 on Discriminant Analysis (10 minutes)

1. Load the R script **BreakoutS12.R** posted in the Slack channel of the course. This file contains an instruction to load a data file containing the genotype of a SNP and its two corresponding allele intensities, A and B.
2. Load the data in the R environment.
3. Make a scatterplot of the allele intensities, colouring individuals according to their genotype (0, 1, 2 or NA).
4. Run a linear discriminant analysis with genotype as the categorical variable, and the allele intensities as predictors, using the complete data as the training data set. Calculate the confusion matrix. Calculate the classification rate.
5. Plot the training data in discriminant space. Is there a good separation?
6. Predict the genotype of the individuals whose genotype is NA (the test data)
7. Add the test data to your plot of the discriminant space, marking individuals with a maximum posterior probability below 0.95. What do you observe?
8. Save your results and compare your results with those discussed in the online session after the breakout room closes.