DataVizA Tutorial: Discriminant Analysis

Department of Econometrics and Business Statistics, Monash University

Tutorial 10

Wine Data

1. Carry out Linear Discriminant Analysis (LDA) using all data in *ExistingWines.rds* in the training set and predict the data in *NewWines.rds*.
2. What are the predictions for the first ten wines in the *NewWines.rds*
3. Repeat the analyis using Quadratic Disciminant Analysis (QDA).
4. What are the predicted probabilities for the first ten wines in the *NewWines.rds* for QDA.
5. Split the data in *ExistingWines.rds* into a training sample (of roughly 70%) and a test sample (of roughly 30%).
6. Is LDA better than QDA for this data?
7. Under what assumptions would QDA theoretically be better than LDA. Investigate whether this assumption holds.
8. What other assumption is required for LDA or QDA to theoretically minimise misclassification rate? Think of a way to do a quick visual check of whether this assumption holds.