

SUMMARY OF SECTION 4.5 - COMPARISON OF CLASSIFICATION METHODS

Logistic regression, LDA, QDA, and kNN are all different classification methods that each have their strengths and weaknesses - understanding the scenarios in which they are most useful can help us build more accurate models. Logistic regression and LDA both create linear decision boundaries (only difference between the two is their fitting procedures) and therefore when the true decision boundary is linear in form, these methods perform well. kNN, on the other hand, is completely non-parametric (does not assume shape), and hence performs better when the true boundary is highly-nonlinear. The disadvantage with kNN is that we cannot infer anything about the individual predictors and their impact. QDA falls somewhere in between the two: able to fit a wider range of shapes than linear methods but not as flexible as kNN. By making assumptions about the data, it is able to perform better at lower number of training examples than kNN. There are several in between states as well - such as transformations of the predictors - that can be performed to move between these four main types of classification methods. It is important to realize the benefits and shortcomings of each method so that we can apply the correct one when the problem needs it.