

Week 9: Practical Data Science with R, Chapter 5

1. Models are used to solve concrete business problems such as detecting fraudulent transactions, identifying spam, predicting loan default, making purchasing recommendations, measuring the effects of a price change, and grouping customers for targeted advertising.
2. Determining the right combination of features, algorithms, and validation standards is not an exact science, and may require some trial and error. In general, feature sets should be kept small to avoid over-fitting. As for algorithms: Naïve Bayes works best for classification with large numbers of features; decision trees are best if features interact, or if you want to create a visualization of the learning process, and logistic regression is best if you're trying to produce a probability.
3. Performance measures for classification includes accuracy (total correct/total), precision (true positive/total identified as positive), recall (true positive/total that are actually positive), sensitivity (same as precision) and specificity (opposite of recall: true negative/total actually negative).