

An Introduction to Statistical Learning

1. Introduction

2. Statistical Learning

- What is statistical learning
- Assessing model accuracy
- Lab: introduction to R

3. Linear Regression

- Simple linear regression
- Multiple linear regression
- Other considerations in the regression model
- The marketing plan
- Comparison of linear regression with k-nearest neighbors
- Lab: Linear regression

4. Classification

- An overview of classification
- Why not linear regression
- Logistic regression
- Linear discriminant analysis
- A comparison of classification methods
- Lab: logistic regression, LDA, QDA and KNN

5. Resampling Methods

- Cross-Validation
- The bootstrap
- Lab: Cross-Validation and the bootstrap

6. Linear Model Selection and Regularization

- Subset selection
- Shrinkage methods
- Dimension reduction methods
- Considerations in high dimensions
- Lab1: Subset selection methods
- Lab2: Ridge regression and the lasso
- Lab3: PCR and PLS regression

7. Moving Beyond Linearity

- Polynomial regression
- Step functions
- Basis Functions
- Regression splines
- Smoothing splines
- Local regression
- Generalized additive models
- Lab: Non-linear modeling

8. Tree-Based Methods

- The basics of decision trees
- Bagging, random forest, boosting
- Lab: Decision trees

9. Support Vector Machines

- Maximal margin classifier
- Support vector classifiers
- Support vector machines
- SVMs with more than two classes
- Relationship to logistic regression
- Lab: Support vector machines

10. Unsupervised Learning

- The challenge of unsupervised learning
- Principal components analysis
- Clustering methods
- Lab1: Principal components analysis
- Lab2: Clustering
- Lab3: NCI60 data example