



STEPHEN F. AUSTIN STATE UNIVERSITY

Department of Mathematics and Statistics

STAT 3348–Predictive Modeling Departmental Course Syllabus

Course description: A broad overview of various statistical methods with an emphasis in predictive modeling and R programming. Topics can include statistical learning, classification, advanced feature selection methods, bootstrap, cross validation, tree algorithms, unsupervised learning, and Bayesian concepts.

Credit hours: 3

Contact hours: 3

Course Prerequisites and Corequisites: STA 3342

Course outline

Course Topics:

Approximate time spent

- | | |
|--|-----|
| • What is statistical learning? <ul style="list-style-type: none">○ Exploratory Analysis○ Key Definitions○ Need for Validation○ The Bias/Variance Trade Off | 15% |
| • Multiple Linear Regression <ul style="list-style-type: none">○ Estimation and assumptions○ Pitfalls and Short comings○ K-Nearest Neighbors | 15% |
| • Classification <ul style="list-style-type: none">○ Logistic Regression○ Linear Discriminate Analysis○ Intro to ROC Curves | 25% |
| • Resampling Methods <ul style="list-style-type: none">○ Cross Validation○ Bootstrap○ Feature Selection Using LASSO | 15% |
| • Tree Based Methods <ul style="list-style-type: none">○ CART models○ CART vs Regression○ Random Forrest | 15% |
| • Unsupervised Learning and Data Visualization <ul style="list-style-type: none">○ Clustering methods○ Intro to PCA○ Visualizations for Unsupervised learning | 15% |

Student Learning Outcomes (SLO): At the end of STAT 3348, a student who has studied and learned the material should be able to:

1. Exhibit a firm understanding of how to enter, manipulate, analyze, and visualize data within R statistical software.
2. Identify and apply correct modeling procedures and strategies to different data types depending on the study design
3. Assess model and prediction performance through resampling techniques and correct interpretation of ROC curves

4. Explain the general concept of the trade offs between model complexity and interpretability
5. Explain the difference between how resampling can play a role in feature selection and how that potentially differs from estimating prediction performance
6. Apply appropriate unsupervised analysis techniques to aid in data descriptions, reduction, and visualization
7. Use machine learning algorithms such as random forest or support vector machines as an alternative to traditional predictive modeling techniques.

There are no specific program learning outcomes for this major addressed in this course. It is a general education core curriculum course and/or a service course.

Date of document: 9/22/2026