

**STAT 434 – Introduction to Statistical Learning**

Winter 2016

**1. Catalog Description**

**STAT 434 Introduction to Statistical Learning (4 units)**

Modern methods and techniques in predictive modeling and classification including splines, smoothing splines, ridge regression, LASSO, regression and classification trees, generalized additive models, logistic regression, and linear discriminant analysis. Model assessment and selection using cross validation, bootstrapping, AIC, and BIC. Use of statistical software for implementation of methods.

**2. Required Background and/or Experience**

STAT 324, STAT 334 or consent of instructor

**3. Expected Outcomes**

The student should:

- a. Be able to understand the difference between predictive modelling and modelling for inferential purposes
- b. Be able to appropriately implement classification models using logistic regression and discriminant analysis
- c. Be able to assess models using cross validation, bootstrap methods, MSE
- d. Be able to select models using best subsets, AIC and BIC, dimension reduction techniques and shrinkage methods
- e. Know about and how to implement alternatives to linear models, e.g. polynomial regression, splines, and additive models
- f. Know about and how to implement tree-based methods for regression and classification
- g. Be able to compare and contrast different modeling strategies and understand their advantages and limitations
- h. Be able to implement all methods learned in class using statistical software such as R, SAS, or JMP

**4. Suggested Texts and References**

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2015). Introduction to Statistical Learning with Applications in R. New York: Springer.

**5. Minimum Student Materials**

Access to computing facilities

**6. Minimum University Facilities**

Chalkboards for classroom use, overhead projectors, computer facilities for student use in preparing assignments.

**7. Expanded Description of Content and Method**

**CONTENT**

- A. Introduction to statistical learning concepts**
1. Estimating functions
  2. Prediction and inference

**LECTURE HOURS**

**5-6**

	<ul style="list-style-type: none"> <li>3. Parametric versus non-parametric models</li> <li>4. Flexibility versus interpretability of the model</li> <li>5. Supervised versus unsupervised learning</li> <li>6. Bias-variance trade-off; test and training data</li> <li>7. Test error versus training error</li> <li>8. Regression versus classification</li> <li>9. Intro to statistical software (R and/or JMP and/or SAS)</li> </ul>	
<b>B.</b>	<b>Review of linear regression</b>	<b>7-8</b>
	<ul style="list-style-type: none"> <li>1. Least squares method</li> <li>2. Simple linear regression</li> <li>3. Multiple linear regression topics</li> <li>4. Hypothesis testing</li> <li>5. Assessing model accuracy</li> <li>6. Potential model problems</li> </ul>	
<b>C.</b>	<b>Classification models</b>	<b>4</b>
	<ul style="list-style-type: none"> <li>1. Logistic regression</li> <li>2. Discriminant analysis</li> </ul>	
<b>D.</b>	<b>Model Assessment Techniques</b>	<b>4</b>
	<ul style="list-style-type: none"> <li>1. Training and validation data</li> <li>2. Cross validation</li> <li>3. Bootstrap techniques to estimate test error</li> </ul>	
<b>E.</b>	<b>Model Selection Techniques</b>	<b>4</b>
	<ul style="list-style-type: none"> <li>1. Best subsets</li> <li>2. Cross validation</li> <li>3. AIC and BIC</li> <li>4. Ridge regression</li> <li>5. LASSO</li> <li>6. Principal components regression</li> </ul>	
<b>F.</b>	<b>Alternatives to Linear Models</b>	<b>5</b>
	<ul style="list-style-type: none"> <li>1. Polynomial regression</li> <li>2. Basis functions</li> <li>3. Regression splines</li> <li>4. Choosing number and location of knots</li> <li>5. Smoothing splines</li> <li>6. Choosing value of the smoothing parameter</li> <li>7. Local regression</li> <li>8. Generalized additive models</li> </ul>	
<b>G.</b>	<b>Tree-based methods</b>	<b>4</b>
	<ul style="list-style-type: none"> <li>1. Regression trees</li> <li>2. Classification trees</li> <li>3. Tree pruning</li> <li>4. Bagging and boosting</li> <li>5. Random forests</li> </ul>	
<b>H.</b>	<b>Summary of Methods and Techniques</b>	<b>2</b>
	<ul style="list-style-type: none"> <li>1. Contrast various predictive modelling techniques</li> </ul>	

2. Determine appropriateness of each technique

**TOTAL: 35-37**

**METHOD**

Material will be presented in a lecture format. Students will be required to use available computer resources.

**8. Method of Evaluating Outcome**

Homework assignments, computing assignments, exams, and individual or team projects.