

## **Stat 202: Data Mining**

### **Professor: Art Owen**

This course covers Data Mining, with an applied orientation. We will cover a selection of topics, such as: Association rules, Clustering, Decision Trees, Neural networks, and Nearest Neighbors.

#### **Prerequisites**

Students must be able to use R or Splus. It is not necessary to know one of those languages already, if one has prior experience programming. Some elementary probability is required: students should be familiar with random variables, probabilities of events, means, variances, correlations, and similar notions.

#### **Texts**

Venables and Ripley “Modern Applied Statistics with S”  
Hand, Mannila, and Smyth “Principles of Data Mining”

#### **Evaluation**

There are no exams. There will be four to six problem sets to implement, extend and apply the methods taught in class.

#### **Times**

Skilling 193, MWF, 1:05-2:05  
URL: [www-stat.stanford.edu/~owen/courses/362](http://www-stat.stanford.edu/~owen/courses/362)