

# The Elements of Statistical Learning

Hastie, Tibshirani, and Friedman (2009). The Elements of Statistical Learning. Second Edition. Springer.

## 1 Introduction

Statistical learning plays a key role in science, finance, industry, and many more areas. This book is about learning from data: supervised learning (presence of outcome variable for learning) and unsupervised learning (outcome variable is not observed).

Running examples:

- classification of spam emails
- explaining prostate specific antigen from clinical measurements via regression
- classification of handwritten digits
- clustering of DNA microarray data for cancer diagnostic

## 2 Overview of Supervised Learning

Two simple but powerful prediction methods are least squares linear models and k-nearest neighbor prediction rule. The former makes huge assumptions about structure and yields stable but possibly inaccurate predictions (low variance, high bias), the latter makes very mild assumptions with often accurate but unstable predictions (low bias, high variance).

Local methods like k-nearest neighbor prediction suffer from the curse of dimensionality: in high dimensions, samples only sparsely populate the input space and are close to an edge. By imposing restrictions on the model class (e.g., linear models), this can be avoided. Many models have been proposed that lie in the spectrum between rigid model assumptions and flexibility.

## 3 Linear Methods for Regression

...