Practical Statistical Learning

John Marten and Feng Liang

2022-09-22

Contents

Preface

These notes are based on a course in statistical learning developed by John Marten and Feng Liang at UIUC using the text **The Elements of Statistical Learning** by Hastie, Tibshirani and Friedman https://hastie.su.domains/Elem StatLearn/.

Chapter 1

Introduction

Notes [lec_W1.1_Introduction_Statistical_Learning.pdf] [lec_W1.2_kNN_vs_LinearRegression.pdf] [lec_W1.2_kNN_vs_LinearRegression_figs.pdf] [lec_W1.3_Introduction_LearningTheory.pdf]

R/Python Code: [Rcode_W1_SimulationStudy.html] [Rcode_W1_Examples_from_ESL.html] [Python_W1_SimulationStudy.html]

1.1 Introduction to Statistical Learning

1.1.1 Types of learning problems

1.1.2 Challenge of supervised learning

1.1.3 Curse of dimensionality

1.1.4 A glimpse of learning theory

1.1.5 1.5 Bias and variance tradeoff

1.2 Least squares vs. nearest neighbors

1.2.1 Introduction to LS and kNN

1.2.2 2.2 Simulation study with R

1.2.3 2.3 Compute Bayes rule

1.2.4 Discussion

Chapter 2

Linear Regression

Notes [lec_W2.1_LinearRegression_MLR.pdf] [lec_W2.2_LinearRegression_Geometry.pdf] [lec_W2.3_LinearRegression_Practice.pdf] [lec_W2_appendix_SLR.pdf]

R/Python Code [Rcode_W2_LinearRegression.html] [Python_W2_LinearRegression_1.html] [Python_W2_LinearRegression_2.html]

2.1 Good predictions: Squared error loss and in-sample error

2.2 Matrices and least-squares estimates

2.3 Regression inference

2.4 Prediction using least-squares

2.5 Geometric interpretation

2.5.1 Basic concepts in vector spaces

2.5.2 LS and projection

2.6 Practical issues

2.6.1 Using R

2.6.2 Interpret LS coefficients

2.6.3 Handle categorical variables

2.6.4 Outliers and rank deficiency

Chapter 3

Variable Selection and Regularization

Notes [lec_W3_VariableSelection.pdf] [lec_W3_PCR.pdf] [lec_W3_More_on_Ridge_Lasso.pdf]
R/Python Code [Rcode_W3_VarSel_SubsetSelection] [Rcode_W3_VarSel_RidgeLasso]

3.1 Subset selection

3.2 Ridge regression

3.3 Lasso regression

Chapter 4

Regression Trees

Chapter 5

Nonlinear Regression

5.1 Polynomials

5.2 Regression Splines

5.3 Smoothing Splines

5.4 Sinces and cosines

5.5 A glimpse of wavelets

Chapter 6

Clustering Analysis

Chapter 7

Latent Structure Models

Chapter 8

More on Clustering

Chapter 9

Distriminant Analysis

Chapter 10

Logistic Regression

Chapter 11

Support Vector Machines

Chapter 12

Classification Trees and Boosting

Chapter 13

Recommender System

Chapter 14

Introduction to Deep Learning