

ADDISON WESLEY DATA & ANALYTICS SERIES



MACHINE LEARNING WITH PYTHON FOR EVERYONE



MARK E. FENNER

Machine Learning with Python for Everyone

Machine Learning with Python for Everyone

Table of Contents

Cover

Title Page

Copyright Page

Contents

Foreword

Preface

About the Author

Part I: First Steps

1 Lets Discuss Learning

1.1 Welcome

1.2 Scope, Terminology, Prediction, and Data

1.2.1 Features

1.2.2 Target Values and Predictions

1.3 Putting the Machine in Machine Learning

1.4 Examples of Learning Systems

1.4.1 Predicting Categories: Examples of Classifiers

1.4.2 Predicting Values: Examples of Regressors

1.5 Evaluating Learning Systems

1.5.1 Correctness

1.5.2 Resource Consumption

1.6 A Process for Building Learning Systems

1.7 Assumptions and Reality of Learning

1.8 End-of-Chapter Material

Table of Contents

1.8.1 The Road Ahead

1.8.2 Notes

2 Some Technical Background

2.1 About Our Setup

2.2 The Need for Mathematical Language

2.3 Our Software for Tackling Machine Learning

2.4 Probability

2.4.1 Primitive Events

2.4.2 Independence

2.4.3 Conditional Probability

2.4.4 Distributions

2.5 Linear Combinations, Weighted Sums, and Dot Products

2.5.1 Weighted Average

2.5.2 Sums of Squares

2.5.3 Sum of Squared Errors

2.6 A Geometric View: Points in Space

2.6.1 Lines

2.6.2 Beyond Lines

2.7 Notation and the Plus-One Trick

2.8 Getting Groovy, Breaking the Straight-Jacket, and Nonlinearity

2.9 NumPy versus All the Maths

2.9.1 Back to 1D versus 2D

2.10 Floating-Point Issues

2.11 EOC 53

2.11.1 Summary

2.11.2 Notes

3 Predicting Categories: Getting Started with Classification

3.1 Classification Tasks

3.2 A Simple Classification Dataset

3.3 Training and Testing: Dont Teach to the Test

3.4 Evaluation: Grading the Exam

Table of Contents

3.5 Simple Classifier #1: Nearest Neighbors, Long Distance Relationships, and Assumptions

3.5.1 Defining Similarity

3.5.2 The k in k-NN

3.5.3 Answer Combination

3.5.4 k-NN, Parameters, and Nonparametric Methods

3.5.5 Building a k-NN Classification Model

3.6 Simple Classifier #2: Naive Bayes, Probability, and Broken Promises

3.7 Simplistic Evaluation of Classifiers

3.7.1 Learning Performance

3.7.2 Resource Utilization in Classification

3.7.3 Stand-Alone Resource Evaluation

3.8 EOC

3.8.1 Sophomore Warning: Limitations and Open Issues

3.8.2 Summary

3.8.3 Notes

3.8.4 Exercises

4 Predicting Numerical Values: Getting Started with Regression

4.1 A Simple Regression Dataset

4.2 Nearest-Neighbors Regression and Summary Statistics

4.2.1 Measures of Center: Median and Mean

4.2.2 Building a k-NN Regression Model

4.3 Linear Regression and Errors

4.3.1 No Flat Earth: Why We Need Slope

4.3.2 Tilting the Field

4.3.3 Performing Linear Regression

4.4 Optimization: Picking the Best Answer

4.4.1 Random Guess

4.4.2 Random Step

4.4.3 Smart Step

4.4.4 Calculated Shortcuts

4.4.5 Application to Linear Regression

Table of Contents

4.5 Simple Evaluation and Comparison of Regressors

4.5.1 Root Mean Squared Error

4.5.2 Learning Performance

4.5.3 Resource Utilization in Regression

4.6 EOC

4.6.1 Limitations and Open Issues

4.6.2 Summary

4.6.3 Notes

4.6.4 Exercises

Part II: Evaluation

5 Evaluating and Comparing Learners

5.1 Evaluation and Why Less Is More

5.2 Terminology for Learning Phases

5.2.1 Back to the Machines

5.2.2 More Technically Speaking . . .

5.3 Major Tom, Theres Something Wrong: Overfitting and Underfitting

5.3.1 Synthetic Data and Linear Regression

5.3.2 Manually Manipulating Model Complexity

5.3.3 Goldilocks: Visualizing Overfitting, Underfitting, and Just Right

5.3.4 Simplicity

5.3.5 Take-Home Notes on Overfitting

5.4 From Errors to Costs

5.4.1 Loss

5.4.2 Cost

5.4.3 Score

5.5 (Re)Sampling: Making More from Less

5.5.1 Cross-Validation

5.5.2 Stratification

5.5.3 Repeated Train-Test Splits

5.5.4 A Better Way and Shuffling

5.5.5 Leave-One-Out Cross-Validation

5.6 Break-It-Down: Deconstructing Error into Bias and Variance

Table of Contents

5.6.1 Variance of the Data

5.6.2 Variance of the Model

5.6.3 Bias of the Model

5.6.4 All Together Now

5.6.5 Examples of Bias-Variance Tradeoffs

5.7 Graphical Evaluation and Comparison

5.7.1 Learning Curves: How Much Data Do We Need?

5.7.2 Complexity Curves

5.8 Comparing Learners with Cross-Validation

5.9 EOC

5.9.1 Summary

5.9.2 Notes

5.9.3 Exercises

6 Evaluating Classifiers

6.1 Baseline Classifiers

6.2 Beyond Accuracy: Metrics for Classification

6.2.1 Eliminating Confusion from the Confusion Matrix

6.2.2 Ways of Being Wrong

6.2.3 Metrics from the Confusion Matrix

6.2.4 Coding the Confusion Matrix

6.2.5 Dealing with Multiple Classes: Multiclass Averaging

6.2.6 F1

6.3 ROC Curves

6.3.1 Patterns in the ROC

6.3.2 Binary ROC

6.3.3 AUC: Area-Under-the-(ROC)-Curve

6.3.4 Multiclass Learners, One-versus-Rest, and ROC

6.4 Another Take on Multiclass: One-versus-One

6.4.1 Multiclass AUC Part Two: The Quest for a Single Value

6.5 Precision-Recall Curves

6.5.1 A Note on Precision-Recall Tradeoff

6.5.2 Constructing a Precision-Recall Curve

Table of Contents

6.6 Cumulative Response and Lift Curves

6.7 More Sophisticated Evaluation of Classifiers: Take Two

6.7.1 Binary

6.7.2 A Novel Multiclass Problem

6.8 EOC

6.8.1 Summary

6.8.2 Notes

6.8.3 Exercises

7 Evaluating Regressors

7.1 Baseline Regressors

7.2 Additional Measures for Regression

7.2.1 Creating Our Own Evaluation Metric

7.2.2 Other Built-in Regression Metrics

7.2.3 R^2

7.3 Residual Plots

7.3.1 Error Plots

7.3.2 Residual Plots

7.4 A First Look at Standardization

7.5 Evaluating Regressors in a More Sophisticated Way: Take Two

7.5.1 Cross-Validated Results on Multiple Metrics

7.5.2 Summarizing Cross-Validated Results

7.5.3 Residuals

7.6 EOC

7.6.1 Summary

7.6.2 Notes

7.6.3 Exercises

Part III: More Methods and Fundamentals

8 More Classification Methods

8.1 Revisiting Classification

8.2 Decision Trees

8.2.1 Tree-Building Algorithms

Table of Contents

8.2.2 Lets Go: Decision Tree Time

8.2.3 Bias and Variance in Decision Trees

8.3 Support Vector Classifiers

8.3.1 Performing SVC

8.3.2 Bias and Variance in SVCs

8.4 Logistic Regression

8.4.1 Betting Odds

8.4.2 Probabilities, Odds, and Log-Odds

8.4.3 Just Do It: Logistic Regression Edition

8.4.4 A Logistic Regression: A Space Oddity

8.5 Discriminant Analysis

8.5.1 Covariance

8.5.2 The Methods

8.5.3 Performing DA

8.6 Assumptions, Biases, and Classifiers

8.7 Comparison of Classifiers: Take Three

8.7.1 Digits

8.8 EOC

8.8.1 Summary

8.8.2 Notes

8.8.3 Exercises

9 More Regression Methods

9.1 Linear Regression in the Penalty Box: Regularization

9.1.1 Performing Regularized Regression

9.2 Support Vector Regression

9.2.1 Hinge Loss

9.2.2 From Linear Regression to Regularized Regression to Support Vector Regression

9.2.3 Just Do It SVR Style

9.3 Piecewise Constant Regression

9.3.1 Implementing a Piecewise Constant Regressor

9.3.2 General Notes on Implementing Models

9.4 Regression Trees

Table of Contents

9.4.1 Performing Regression with Trees

9.5 Comparison of Regressors: Take Three

9.6 EOC

9.6.1 Summary

9.6.2 Notes

9.6.3 Exercises

10 Manual Feature Engineering: Manipulating Data for Fun and Profit

10.1 Feature Engineering Terminology and Motivation

10.1.1 Why Engineer Features?

10.1.2 When Does Engineering Happen?

10.1.3 How Does Feature Engineering Occur?

10.2 Feature Selection and Data Reduction: Taking out the Trash

10.3 Feature Scaling

10.4 Discretization

10.5 Categorical Coding

10.5.1 Another Way to Code and the Curious Case of the Missing Intercept

10.6 Relationships and Interactions

10.6.1 Manual Feature Construction

10.6.2 Interactions

10.6.3 Adding Features with Transformers

10.7 Target Manipulations

10.7.1 Manipulating the Input Space

10.7.2 Manipulating the Target

10.8 EOC

10.8.1 Summary

10.8.2 Notes

10.8.3 Exercises

11 Tuning Hyperparameters and Pipelines

11.1 Models, Parameters, Hyperparameters

11.2 Tuning Hyperparameters

11.2.1 A Note on Computer Science and Learning Terminology

11.2.2 An Example of Complete Search

Table of Contents

- 11.2.3 Using Randomness to Search for a Needle in a Haystack
- 11.3 Down the Recursive Rabbit Hole: Nested Cross-Validation
 - 11.3.1 Cross-Validation, Redux
 - 11.3.2 GridSearch as a Model
 - 11.3.3 Cross-Validation Nested within Cross-Validation
 - 11.3.4 Comments on Nested CV
- 11.4 Pipelines
 - 11.4.1 A Simple Pipeline
 - 11.4.2 A More Complex Pipeline
- 11.5 Pipelines and Tuning Together
- 11.6 EOC
 - 11.6.1 Summary
 - 11.6.2 Notes
 - 11.6.3 Exercises

Part IV: Adding Complexity

12 Combining Learners

- 12.1 Ensembles
- 12.2 Voting Ensembles
- 12.3 Bagging and Random Forests
 - 12.3.1 Bootstrapping
 - 12.3.2 From Bootstrapping to Bagging
 - 12.3.3 Through the Random Forest
- 12.4 Boosting
 - 12.4.1 Boosting Details
- 12.5 Comparing the Tree-Ensemble Methods
- 12.6 EOC
 - 12.6.1 Summary
 - 12.6.2 Notes
 - 12.6.3 Exercises

13 Models That Engineer Features for Us

- 13.1 Feature Selection

Table of Contents

13.1.1 Single-Step Filtering with Metric-Based Feature Selection

13.1.2 Model-Based Feature Selection

13.1.3 Integrating Feature Selection with a Learning Pipeline

13.2 Feature Construction with Kernels

13.2.1 A Kernel Motivator

13.2.2 Manual Kernel Methods

13.2.3 Kernel Methods and Kernel Options

13.2.4 Kernelized SVCs: SVMs

13.2.5 Take-Home Notes on SVM and an Example

13.3 Principal Components Analysis: An Unsupervised Technique

13.3.1 A Warm Up: Centering

13.3.2 Finding a Different Best Line

13.3.3 A First PCA

13.3.4 Under the Hood of PCA

13.3.5 A Finale: Comments on General PCA

13.3.6 Kernel PCA and Manifold Methods

13.4 EOC

13.4.1 Summary

13.4.2 Notes

13.4.3 Exercises

14 Feature Engineering for Domains: Domain-Specific Learning

14.1 Working with Text

14.1.1 Encoding Text

14.1.2 Example of Text Learning

14.2 Clustering

14.2.1 k-Means Clustering

14.3 Working with Images

14.3.1 Bag of Visual Words

14.3.2 Our Image Data

14.3.3 An End-to-End System

14.3.4 Complete Code of BoVW Transformer

14.4 EOC

Table of Contents

14.4.1 Summary

14.4.2 Notes

14.4.3 Exercises

15 Connections, Extensions, and Further Directions

15.1 Optimization

15.2 Linear Regression from Raw Materials

15.2.1 A Graphical View of Linear Regression

15.3 Building Logistic Regression from Raw Materials

15.3.1 Logistic Regression with Zero-One Coding

15.3.2 Logistic Regression with Plus-One Minus-One Coding

15.3.3 A Graphical View of Logistic Regression

15.4 SVM from Raw Materials

15.5 Neural Networks

15.5.1 A NN View of Linear Regression

15.5.2 A NN View of Logistic Regression

15.5.3 Beyond Basic Neural Networks

15.6 Probabilistic Graphical Models

15.6.1 Sampling

15.6.2 A PGM View of Linear Regression

15.6.3 A PGM View of Logistic Regression

15.7 EOC

15.7.1 Summary

15.7.2 Notes

15.7.3 Exercises

Appendix A: mlwpy.py Listing

Index