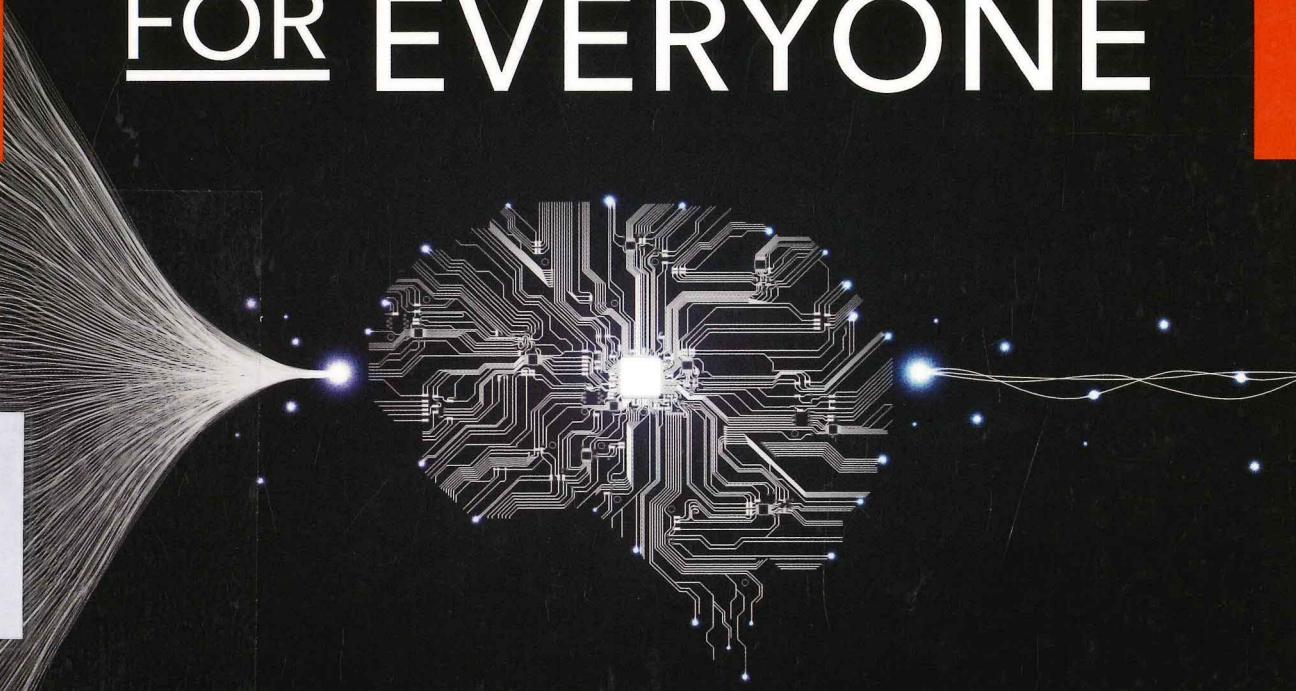


ADDISON WESLEY DATA & ANALYTICS SERIES



# MACHINE LEARNING WITH PYTHON FOR EVERYONE



MARK E. FENNER

b 1670647X  
o 12578095  
i 22686642

สำนักหอสมุด มหาวิทยาลัยเชียงใหม่

# Machine Learning with Python for Everyone

---



Mark E. Fenner

◆ Addison-Wesley

Boston • Columbus • New York • San Francisco • Amsterdam • Cape Town  
Dubai • London • Madrid • Milan • Munich • Paris • Montreal • Toronto • Delhi • Mexico City  
São Paulo • Sydney • Hong Kong • Seoul • Singapore • Taipei • Tokyo

# Contents

**Foreword xxi**

**Preface xxiii**

**About the Author xxvii**

## **I First Steps 1**

### **1 Let's Discuss Learning 3**

- 1.1 Welcome 3
- 1.2 Scope, Terminology, Prediction, and Data 4
  - 1.2.1 Features 5
  - 1.2.2 Target Values and Predictions 6
- 1.3 Putting the Machine in Machine Learning 7
- 1.4 Examples of Learning Systems 9
  - 1.4.1 Predicting Categories: Examples of Classifiers 9
  - 1.4.2 Predicting Values: Examples of Regressors 10
- 1.5 Evaluating Learning Systems 11
  - 1.5.1 Correctness 11
  - 1.5.2 Resource Consumption 12
- 1.6 A Process for Building Learning Systems 13
- 1.7 Assumptions and Reality of Learning 15
- 1.8 End-of-Chapter Material 17
  - 1.8.1 The Road Ahead 17
  - 1.8.2 Notes 17

### **2 Some Technical Background 19**

- 2.1 About Our Setup 19
- 2.2 The Need for Mathematical Language 19

2.3	Our Software for Tackling Machine Learning	20
2.4	Probability	21
2.4.1	Primitive Events	22
2.4.2	Independence	23
2.4.3	Conditional Probability	24
2.4.4	Distributions	25
2.5	Linear Combinations, Weighted Sums, and Dot Products	28
2.5.1	Weighted Average	30
2.5.2	Sums of Squares	32
2.5.3	Sum of Squared Errors	33
2.6	A Geometric View: Points in Space	34
2.6.1	Lines	34
2.6.2	Beyond Lines	39
2.7	Notation and the Plus-One Trick	43
2.8	Getting Groovy, Breaking the Straight-Jacket, and Nonlinearity	45
2.9	NumPy versus “All the Maths”	47
2.9.1	Back to 1D versus 2D	49
2.10	Floating-Point Issues	52
2.11	EOC	53
2.11.1	Summary	53
2.11.2	Notes	54
<b>3</b>	<b>Predicting Categories: Getting Started with Classification</b>	<b>55</b>
3.1	Classification Tasks	55
3.2	A Simple Classification Dataset	56
3.3	Training and Testing: Don’t Teach to the Test	59
3.4	Evaluation: Grading the Exam	62
3.5	Simple Classifier #1: Nearest Neighbors, Long Distance Relationships, and Assumptions	63
3.5.1	Defining Similarity	63
3.5.2	The $k$ in $k$ -NN	64
3.5.3	Answer Combination	64

3.5.4	<i>k</i> -NN, Parameters, and Nonparametric Methods	65
3.5.5	Building a <i>k</i> -NN Classification Model	66
3.6	Simple Classifier #2: Naive Bayes, Probability, and Broken Promises	68
3.7	Simplistic Evaluation of Classifiers	70
3.7.1	Learning Performance	70
3.7.2	Resource Utilization in Classification	71
3.7.3	Stand-Alone Resource Evaluation	77
3.8	EOC	81
3.8.1	Sophomore Warning: Limitations and Open Issues	81
3.8.2	Summary	82
3.8.3	Notes	82
3.8.4	Exercises	83

#### **4 Predicting Numerical Values: Getting Started with Regression 85**

4.1	A Simple Regression Dataset	85
4.2	Nearest-Neighbors Regression and Summary Statistics	87
4.2.1	Measures of Center: Median and Mean	88
4.2.2	Building a <i>k</i> -NN Regression Model	90
4.3	Linear Regression and Errors	91
4.3.1	No Flat Earth: Why We Need Slope	92
4.3.2	Tilting the Field	94
4.3.3	Performing Linear Regression	97
4.4	Optimization: Picking the Best Answer	98
4.4.1	Random Guess	98
4.4.2	Random Step	99
4.4.3	Smart Step	99
4.4.4	Calculated Shortcuts	100

- 4.4.5 Application to Linear Regression 101
- 4.5 Simple Evaluation and Comparison of Regressors 101
  - 4.5.1 Root Mean Squared Error 101
  - 4.5.2 Learning Performance 102
  - 4.5.3 Resource Utilization in Regression 102
- 4.6 EOC 104
  - 4.6.1 Limitations and Open Issues 104
  - 4.6.2 Summary 105
  - 4.6.3 Notes 105
  - 4.6.4 Exercises 105

## **II Evaluation 107**

### **5 Evaluating and Comparing Learners 109**

- 5.1 Evaluation and Why Less Is More 109
- 5.2 Terminology for Learning Phases 110
  - 5.2.1 Back to the Machines 110
  - 5.2.2 More Technically Speaking . . . 113
- 5.3 Major Tom, There's Something Wrong: Overfitting and Underfitting 116
  - 5.3.1 Synthetic Data and Linear Regression 117
  - 5.3.2 Manually Manipulating Model Complexity 118
  - 5.3.3 Goldilocks: Visualizing Overfitting, Underfitting, and "Just Right" 120
  - 5.3.4 Simplicity 124
  - 5.3.5 Take-Home Notes on Overfitting 124
- 5.4 From Errors to Costs 125
  - 5.4.1 Loss 125
  - 5.4.2 Cost 126

5.4.3	Score	127
5.5	(Re)Sampling: Making More from Less	128
5.5.1	Cross-Validation	128
5.5.2	Stratification	132
5.5.3	Repeated Train-Test Splits	133
5.5.4	A Better Way and Shuffling	137
5.5.5	Leave-One-Out Cross-Validation	140
5.6	Break-It-Down: Deconstructing Error into Bias and Variance	142
5.6.1	Variance of the Data	143
5.6.2	Variance of the Model	144
5.6.3	Bias of the Model	144
5.6.4	All Together Now	145
5.6.5	Examples of Bias-Variance Tradeoffs	145
5.7	Graphical Evaluation and Comparison	149
5.7.1	Learning Curves: How Much Data Do We Need?	150
5.7.2	Complexity Curves	152
5.8	Comparing Learners with Cross-Validation	154
5.9	EOC	155
5.9.1	Summary	155
5.9.2	Notes	155
5.9.3	Exercises	157
<b>6</b>	<b>Evaluating Classifiers</b>	<b>159</b>
6.1	Baseline Classifiers	159
6.2	Beyond Accuracy: Metrics for Classification	161
6.2.1	Eliminating Confusion from the Confusion Matrix	163
6.2.2	Ways of Being Wrong	164
6.2.3	Metrics from the Confusion Matrix	165
6.2.4	Coding the Confusion Matrix	166
6.2.5	Dealing with Multiple Classes: Multiclass Averaging	168

	6.2.6	$F_1$	170
6.3		ROC Curves	170
	6.3.1	Patterns in the ROC	173
	6.3.2	Binary ROC	174
	6.3.3	AUC: Area-Under-the-(ROC)- Curve	177
	6.3.4	Multiclass Learners, One-versus-Rest, and ROC	179
6.4		Another Take on Multiclass: One-versus-One	181
	6.4.1	Multiclass AUC Part Two: The Quest for a Single Value	182
6.5		Precision-Recall Curves	185
	6.5.1	A Note on Precision-Recall Tradeoff	185
	6.5.2	Constructing a Precision-Recall Curve	186
6.6		Cumulative Response and Lift Curves	187
6.7		More Sophisticated Evaluation of Classifiers: Take Two	190
	6.7.1	Binary	190
	6.7.2	A Novel Multiclass Problem	195
6.8		EOC	201
	6.8.1	Summary	201
	6.8.2	Notes	202
	6.8.3	Exercises	203
<b>7 Evaluating Regressors 205</b>			
7.1		Baseline Regressors	205
7.2		Additional Measures for Regression	207
	7.2.1	Creating Our Own Evaluation Metric	207
	7.2.2	Other Built-in Regression Metrics	208
	7.2.3	$R^2$	209



- 7.3 Residual Plots 214
  - 7.3.1 Error Plots 215
  - 7.3.2 Residual Plots 217
- 7.4 A First Look at Standardization 221
- 7.5 Evaluating Regressors in a More Sophisticated Way: Take Two 225
  - 7.5.1 Cross-Validated Results on Multiple Metrics 226
  - 7.5.2 Summarizing Cross-Validated Results 230
  - 7.5.3 Residuals 230
- 7.6 EOC 232
  - 7.6.1 Summary 232
  - 7.6.2 Notes 232
  - 7.6.3 Exercises 234

### **III More Methods and Fundamentals 235**

#### **8 More Classification Methods 237**

- 8.1 Revisiting Classification 237
- 8.2 Decision Trees 239
  - 8.2.1 Tree-Building Algorithms 242
  - 8.2.2 Let's Go: Decision Tree Time 245
  - 8.2.3 Bias and Variance in Decision Trees 249
- 8.3 Support Vector Classifiers 249
  - 8.3.1 Performing SVC 253
  - 8.3.2 Bias and Variance in SVCs 256
- 8.4 Logistic Regression 259
  - 8.4.1 Betting Odds 259
  - 8.4.2 Probabilities, Odds, and Log-Odds 262
  - 8.4.3 Just Do It: Logistic Regression Edition 267
  - 8.4.4 A Logistic Regression: A Space Oddity 268

- 8.5 Discriminant Analysis 269
  - 8.5.1 Covariance 270
  - 8.5.2 The Methods 282
  - 8.5.3 Performing DA 283
- 8.6 Assumptions, Biases, and Classifiers 285
- 8.7 Comparison of Classifiers: Take Three 287
  - 8.7.1 Digits 287
- 8.8 EOC 290
  - 8.8.1 Summary 290
  - 8.8.2 Notes 290
  - 8.8.3 Exercises 293
- 9 More Regression Methods 295**
  - 9.1 Linear Regression in the Penalty Box: Regularization 295
    - 9.1.1 Performing Regularized Regression 300
  - 9.2 Support Vector Regression 301
    - 9.2.1 Hinge Loss 301
    - 9.2.2 From Linear Regression to Regularized Regression to Support Vector Regression 305
    - 9.2.3 Just Do It—SVR Style 307
  - 9.3 Piecewise Constant Regression 308
    - 9.3.1 Implementing a Piecewise Constant Regressor 310
    - 9.3.2 General Notes on Implementing Models 311
  - 9.4 Regression Trees 313
    - 9.4.1 Performing Regression with Trees 313
  - 9.5 Comparison of Regressors: Take Three 314
  - 9.6 EOC 318
    - 9.6.1 Summary 318
    - 9.6.2 Notes 318
    - 9.6.3 Exercises 319

## **10 Manual Feature Engineering: Manipulating Data for Fun and Profit 321**

- 10.1 Feature Engineering Terminology and Motivation 321
  - 10.1.1 Why Engineer Features? 322
  - 10.1.2 When Does Engineering Happen? 323
  - 10.1.3 How Does Feature Engineering Occur? 324
- 10.2 Feature Selection and Data Reduction: Taking out the Trash 324
- 10.3 Feature Scaling 325
- 10.4 Discretization 329
- 10.5 Categorical Coding 332
  - 10.5.1 Another Way to Code and the Curious Case of the Missing Intercept 334
- 10.6 Relationships and Interactions 341
  - 10.6.1 Manual Feature Construction 341
  - 10.6.2 Interactions 343
  - 10.6.3 Adding Features with Transformers 348
- 10.7 Target Manipulations 350
  - 10.7.1 Manipulating the Input Space 351
  - 10.7.2 Manipulating the Target 353
- 10.8 EOC 356
  - 10.8.1 Summary 356
  - 10.8.2 Notes 356
  - 10.8.3 Exercises 357

## **11 Tuning Hyperparameters and Pipelines 359**

- 11.1 Models, Parameters, Hyperparameters 360
- 11.2 Tuning Hyperparameters 362
  - 11.2.1 A Note on Computer Science and Learning Terminology 362
  - 11.2.2 An Example of Complete Search 362
  - 11.2.3 Using Randomness to Search for a Needle in a Haystack 368

- 11.3 Down the Recursive Rabbit Hole:  
Nested Cross-Validation 370
  - 11.3.1 Cross-Validation, Redux 370
  - 11.3.2 GridSearch as a Model 371
  - 11.3.3 Cross-Validation Nested  
within Cross-Validation 372
  - 11.3.4 Comments on Nested  
CV 375
- 11.4 Pipelines 377
  - 11.4.1 A Simple Pipeline 378
  - 11.4.2 A More Complex  
Pipeline 379
- 11.5 Pipelines and Tuning Together 380
- 11.6 EOC 382
  - 11.6.1 Summary 382
  - 11.6.2 Notes 382
  - 11.6.3 Exercises 383

## **IV Adding Complexity 385**

### **12 Combining Learners 387**

- 12.1 Ensembles 387
- 12.2 Voting Ensembles 389
- 12.3 Bagging and Random Forests 390
  - 12.3.1 Bootstrapping 390
  - 12.3.2 From Bootstrapping to  
Bagging 394
  - 12.3.3 Through the Random  
Forest 396
- 12.4 Boosting 398
  - 12.4.1 Boosting Details 399
- 12.5 Comparing the Tree-Ensemble  
Methods 401
- 12.6 EOC 405
  - 12.6.1 Summary 405
  - 12.6.2 Notes 405
  - 12.6.3 Exercises 406

**13 Models That Engineer Features for Us 409**

- 13.1 Feature Selection 411
  - 13.1.1 Single-Step Filtering with Metric-Based Feature Selection 412
  - 13.1.2 Model-Based Feature Selection 423
  - 13.1.3 Integrating Feature Selection with a Learning Pipeline 426
- 13.2 Feature Construction with Kernels 428
  - 13.2.1 A Kernel Motivator 428
  - 13.2.2 Manual Kernel Methods 433
  - 13.2.3 Kernel Methods and Kernel Options 438
  - 13.2.4 Kernelized SVCs: SVMs 442
  - 13.2.5 Take-Home Notes on SVM and an Example 443
- 13.3 Principal Components Analysis: An Unsupervised Technique 445
  - 13.3.1 A Warm Up: Centering 445
  - 13.3.2 Finding a Different Best Line 448
  - 13.3.3 A First PCA 449
  - 13.3.4 Under the Hood of PCA 452
  - 13.3.5 A Finale: Comments on General PCA 457
  - 13.3.6 Kernel PCA and Manifold Methods 458
- 13.4 EOC 462
  - 13.4.1 Summary 462
  - 13.4.2 Notes 462
  - 13.4.3 Exercises 467

**14 Feature Engineering for Domains: Domain-Specific Learning 469**

- 14.1 Working with Text 470
  - 14.1.1 Encoding Text 471
  - 14.1.2 Example of Text Learning 476
- 14.2 Clustering 479
  - 14.2.1 *k*-Means Clustering 479

- 14.3 Working with Images 481
  - 14.3.1 Bag of Visual Words 481
  - 14.3.2 Our Image Data 482
  - 14.3.3 An End-to-End System 483
  - 14.3.4 Complete Code of BoVW Transformer 491
- 14.4 EOC 493
  - 14.4.1 Summary 493
  - 14.4.2 Notes 494
  - 14.4.3 Exercises 495

## **15 Connections, Extensions, and Further Directions 497**

- 15.1 Optimization 497
- 15.2 Linear Regression from Raw Materials 500
  - 15.2.1 A Graphical View of Linear Regression 504
- 15.3 Building Logistic Regression from Raw Materials 504
  - 15.3.1 Logistic Regression with Zero-One Coding 506
  - 15.3.2 Logistic Regression with Plus-One Minus-One Coding 508
  - 15.3.3 A Graphical View of Logistic Regression 509
- 15.4 SVM from Raw Materials 510
- 15.5 Neural Networks 512
  - 15.5.1 A NN View of Linear Regression 512
  - 15.5.2 A NN View of Logistic Regression 515
  - 15.5.3 Beyond Basic Neural Networks 516
- 15.6 Probabilistic Graphical Models 516
  - 15.6.1 Sampling 518
  - 15.6.2 A PGM View of Linear Regression 519

	15.6.3	A PGM View of Logistic Regression	523
15.7	EOC		525
	15.7.1	Summary	525
	15.7.2	Notes	526
	15.7.3	Exercises	527

## **A mlwpy.py Listing 529**

## **Index 537**