

# Contents

|                                                                                            |           |
|--------------------------------------------------------------------------------------------|-----------|
| List of Figures                                                                            | ix        |
| Preface                                                                                    | xiii      |
| <b>1 Introduction to Machine Learning</b>                                                  | <b>1</b>  |
| 1.1 A simple supervised model: Nearest Neighbor . . . . .                                  | 2         |
| 1.1.1 Tuning Hyperparameters with Cross-Validation . . . . .                               | 9         |
| 1.2 Preprocessing . . . . .                                                                | 14        |
| 1.2.1 Scaling Data . . . . .                                                               | 15        |
| 1.2.2 Forcing Data to be Gaussian: an Introduction to Power Trans-<br>formations . . . . . | 19        |
| 1.2.3 Dealing with Categorical Variables . . . . .                                         | 21        |
| 1.2.4 Handling with Missing Values . . . . .                                               | 24        |
| 1.3 Methods for Dealing with Imbalanced Data . . . . .                                     | 26        |
| 1.3.1 Random Oversampling of the Majority Class . . . . .                                  | 28        |
| 1.3.2 Random Undersampling of the Majority Class . . . . .                                 | 29        |
| 1.3.3 Oversampling using Synthetic Data: SMOTE . . . . .                                   | 30        |
| 1.4 Reducing Dimensionality: Principal Component Analysis . . . . .                        | 31        |
| 1.4.1 PCA as dimensionality reduction . . . . .                                            | 32        |
| 1.4.2 Feature extraction . . . . .                                                         | 36        |
| 1.4.3 Nonlinear Manifold Algorithm: t-SNE . . . . .                                        | 37        |
| <b>2 Linear Models for Machine Learning</b>                                                | <b>41</b> |
| 2.1 Linear Regression . . . . .                                                            | 42        |
| 2.2 Shrinkage Methods . . . . .                                                            | 44        |
| 2.2.1 Ridge Regression . . . . .                                                           | 44        |
| 2.2.2 Lasso Regression . . . . .                                                           | 49        |
| 2.2.3 Elastic Net . . . . .                                                                | 51        |

|          |                                                                                                |            |
|----------|------------------------------------------------------------------------------------------------|------------|
| 2.3      | Robust Regression . . . . .                                                                    | 52         |
| 2.3.1    | Huber Regression . . . . .                                                                     | 53         |
| 2.3.2    | RANSAC . . . . .                                                                               | 56         |
| 2.4      | Logistic Regression . . . . .                                                                  | 58         |
| 2.4.1    | Why Logistic Regression is Linear? . . . . .                                                   | 59         |
| 2.4.2    | Logistic Regression Predictions (Raw Model Output) vs Probabilities (Sigmoid Output) . . . . . | 60         |
| 2.4.3    | Logistic Regression in Python . . . . .                                                        | 61         |
| 2.4.4    | Model Performance Evaluation . . . . .                                                         | 62         |
| 2.4.5    | Regularization . . . . .                                                                       | 66         |
| 2.5      | Linear Support Vector Machine . . . . .                                                        | 68         |
| 2.6      | Beyond Linearity: Kernelized Models . . . . .                                                  | 73         |
| 2.6.1    | Into the Hood of the Kernel Trick . . . . .                                                    | 76         |
| 2.6.2    | Practical Classification Example: Face Recognition . . . . .                                   | 77         |
| <b>3</b> | <b>Beyond Linearity: Ensemble Methods for ML</b>                                               | <b>83</b>  |
| 3.1      | Introduction . . . . .                                                                         | 83         |
| 3.2      | Ensemble Methods . . . . .                                                                     | 84         |
| 3.2.1    | Bootstrap Aggregation . . . . .                                                                | 88         |
| 3.2.2    | Out-of-Bag Estimation . . . . .                                                                | 90         |
| 3.3      | Random Forests . . . . .                                                                       | 91         |
| 3.3.1    | Random Forests Classifier . . . . .                                                            | 91         |
| 3.3.2    | Random Forests Regressor . . . . .                                                             | 94         |
| 3.4      | Boosting Methods . . . . .                                                                     | 95         |
| 3.4.1    | AdaBoost . . . . .                                                                             | 95         |
| 3.4.2    | Gradient Boosting . . . . .                                                                    | 96         |
| 3.4.3    | Extreme Gradient Boosting (XGBoost) . . . . .                                                  | 99         |
| 3.4.4    | CatBoost . . . . .                                                                             | 106        |
| <b>4</b> | <b>An Introduction to Modern ML Techniques</b>                                                 | <b>115</b> |
| 4.1      | Introduction to Natural language Processing . . . . .                                          | 115        |
| 4.1.1    | Preprocessing with Text Data . . . . .                                                         | 116        |
| 4.1.2    | Numerical Representation of Documents: the Bag-of-Words . . . . .                              | 121        |
| 4.1.3    | Practical Example: Sentiment Analysis with IMDb Reviews Dataset . . . . .                      | 124        |
| 4.1.4    | Term Frequency-Inverse Document Frequency . . . . .                                            | 126        |
| 4.1.5    | Bag-of-Words with More Than One Word (n-Grams) . . . . .                                       | 127        |

|                   |                                                            |            |
|-------------------|------------------------------------------------------------|------------|
| 4.1.6             | Beyond Bag-of-words: Word Embeddings . . . . .             | 132        |
| 4.2               | Introduction to Deep Learning . . . . .                    | 140        |
| 4.2.1             | Dealing with Complex Data into a Neural Network . . . . .  | 143        |
| 4.2.2             | Multiclass classification . . . . .                        | 147        |
| <b>Appendices</b> |                                                            | <b>151</b> |
| <b>A</b>          | <b>A crash course in Python</b>                            | <b>153</b> |
| A.1               | Building Blocks in Python . . . . .                        | 153        |
| A.1.1             | Variables . . . . .                                        | 153        |
| A.1.2             | Methods . . . . .                                          | 155        |
| A.2               | Data Structure in Python . . . . .                         | 156        |
| A.2.1             | List and Tuples . . . . .                                  | 156        |
| A.2.2             | Sets . . . . .                                             | 158        |
| A.2.3             | Dictionaries . . . . .                                     | 158        |
| A.3               | Loops in Python . . . . .                                  | 159        |
| A.3.1             | The For Loop . . . . .                                     | 159        |
| A.3.2             | The While Loop . . . . .                                   | 160        |
| A.4               | Advanced Data Structure in Python . . . . .                | 161        |
| A.4.1             | List comprehensions . . . . .                              | 161        |
| A.4.2             | Lambda Functions . . . . .                                 | 163        |
| A.5               | Advanced Concepts on Functions . . . . .                   | 164        |
| A.5.1             | The magic of Wildcards into Function's arguments . . . . . | 164        |
| A.5.2             | Local vs Global Scope in Functions . . . . .               | 168        |
| A.6               | Introduction to Object-Oriented Programming . . . . .      | 169        |
| A.6.1             | Objects, Classes and Attributes . . . . .                  | 170        |
| A.6.2             | Subclasses and Inheritance . . . . .                       | 172        |
| <b>B</b>          | <b>Mathematics behind the skip-gram model</b>              | <b>175</b> |

# Contents

|                                                                                            |           |
|--------------------------------------------------------------------------------------------|-----------|
| List of Figures                                                                            | ix        |
| Preface                                                                                    | xiii      |
| <b>1 Introduction to Machine Learning</b>                                                  | <b>1</b>  |
| 1.1 A simple supervised model: Nearest Neighbor . . . . .                                  | 2         |
| 1.1.1 Tuning Hyperparameters with Cross-Validation . . . . .                               | 9         |
| 1.2 Preprocessing . . . . .                                                                | 14        |
| 1.2.1 Scaling Data . . . . .                                                               | 15        |
| 1.2.2 Forcing Data to be Gaussian: an Introduction to Power Trans-<br>formations . . . . . | 19        |
| 1.2.3 Dealing with Categorical Variables . . . . .                                         | 21        |
| 1.2.4 Handling with Missing Values . . . . .                                               | 24        |
| 1.3 Methods for Dealing with Imbalanced Data . . . . .                                     | 26        |
| 1.3.1 Random Oversampling of the Majority Class . . . . .                                  | 28        |
| 1.3.2 Random Undersampling of the Majority Class . . . . .                                 | 29        |
| 1.3.3 Oversampling using Synthetic Data: SMOTE . . . . .                                   | 30        |
| 1.4 Reducing Dimensionality: Principal Component Analysis . . . . .                        | 31        |
| 1.4.1 PCA as dimensionality reduction . . . . .                                            | 32        |
| 1.4.2 Feature extraction . . . . .                                                         | 36        |
| 1.4.3 Nonlinear Manifold Algorithm: t-SNE . . . . .                                        | 37        |
| <b>2 Linear Models for Machine Learning</b>                                                | <b>41</b> |
| 2.1 Linear Regression . . . . .                                                            | 42        |
| 2.2 Shrinkage Methods . . . . .                                                            | 44        |
| 2.2.1 Ridge Regression . . . . .                                                           | 44        |
| 2.2.2 Lasso Regression . . . . .                                                           | 49        |
| 2.2.3 Elastic Net . . . . .                                                                | 51        |

|          |                                                                                                |            |
|----------|------------------------------------------------------------------------------------------------|------------|
| 2.3      | Robust Regression . . . . .                                                                    | 52         |
| 2.3.1    | Huber Regression . . . . .                                                                     | 53         |
| 2.3.2    | RANSAC . . . . .                                                                               | 56         |
| 2.4      | Logistic Regression . . . . .                                                                  | 58         |
| 2.4.1    | Why Logistic Regression is Linear? . . . . .                                                   | 59         |
| 2.4.2    | Logistic Regression Predictions (Raw Model Output) vs Probabilities (Sigmoid Output) . . . . . | 60         |
| 2.4.3    | Logistic Regression in Python . . . . .                                                        | 61         |
| 2.4.4    | Model Performance Evaluation . . . . .                                                         | 62         |
| 2.4.5    | Regularization . . . . .                                                                       | 66         |
| 2.5      | Linear Support Vector Machine . . . . .                                                        | 68         |
| 2.6      | Beyond Linearity: Kernelized Models . . . . .                                                  | 73         |
| 2.6.1    | Into the Hood of the Kernel Trick . . . . .                                                    | 76         |
| 2.6.2    | Practical Classification Example: Face Recognition . . . . .                                   | 77         |
| <b>3</b> | <b>Beyond Linearity: Ensemble Methods for ML</b>                                               | <b>83</b>  |
| 3.1      | Introduction . . . . .                                                                         | 83         |
| 3.2      | Ensemble Methods . . . . .                                                                     | 84         |
| 3.2.1    | Bootstrap Aggregation . . . . .                                                                | 88         |
| 3.2.2    | Out-of-Bag Estimation . . . . .                                                                | 90         |
| 3.3      | Random Forests . . . . .                                                                       | 91         |
| 3.3.1    | Random Forests Classifier . . . . .                                                            | 91         |
| 3.3.2    | Random Forests Regressor . . . . .                                                             | 94         |
| 3.4      | Boosting Methods . . . . .                                                                     | 95         |
| 3.4.1    | AdaBoost . . . . .                                                                             | 95         |
| 3.4.2    | Gradient Boosting . . . . .                                                                    | 96         |
| 3.4.3    | Extreme Gradient Boosting (XGBoost) . . . . .                                                  | 99         |
| 3.4.4    | CatBoost . . . . .                                                                             | 106        |
| <b>4</b> | <b>An Introduction to Modern ML Techniques</b>                                                 | <b>115</b> |
| 4.1      | Introduction to Natural language Processing . . . . .                                          | 115        |
| 4.1.1    | Preprocessing with Text Data . . . . .                                                         | 116        |
| 4.1.2    | Numerical Representation of Documents: the Bag-of-Words . . . . .                              | 121        |
| 4.1.3    | Practical Example: Sentiment Analysis with IMDb Reviews Dataset . . . . .                      | 124        |
| 4.1.4    | Term Frequency-Inverse Document Frequency . . . . .                                            | 126        |
| 4.1.5    | Bag-of-Words with More Than One Word (n-Grams) . . . . .                                       | 127        |

|                   |                                                            |            |
|-------------------|------------------------------------------------------------|------------|
| 4.1.6             | Beyond Bag-of-words: Word Embeddings . . . . .             | 132        |
| 4.2               | Introduction to Deep Learning . . . . .                    | 140        |
| 4.2.1             | Dealing with Complex Data into a Neural Network . . . . .  | 143        |
| 4.2.2             | Multiclass classification . . . . .                        | 147        |
| <b>Appendices</b> |                                                            | <b>151</b> |
| <b>A</b>          | <b>A crash course in Python</b>                            | <b>153</b> |
| A.1               | Building Blocks in Python . . . . .                        | 153        |
| A.1.1             | Variables . . . . .                                        | 153        |
| A.1.2             | Methods . . . . .                                          | 155        |
| A.2               | Data Structure in Python . . . . .                         | 156        |
| A.2.1             | List and Tuples . . . . .                                  | 156        |
| A.2.2             | Sets . . . . .                                             | 158        |
| A.2.3             | Dictionaries . . . . .                                     | 158        |
| A.3               | Loops in Python . . . . .                                  | 159        |
| A.3.1             | The For Loop . . . . .                                     | 159        |
| A.3.2             | The While Loop . . . . .                                   | 160        |
| A.4               | Advanced Data Structure in Python . . . . .                | 161        |
| A.4.1             | List comprehensions . . . . .                              | 161        |
| A.4.2             | Lambda Functions . . . . .                                 | 163        |
| A.5               | Advanced Concepts on Functions . . . . .                   | 164        |
| A.5.1             | The magic of Wildcards into Function's arguments . . . . . | 164        |
| A.5.2             | Local vs Global Scope in Functions . . . . .               | 168        |
| A.6               | Introduction to Object-Oriented Programming . . . . .      | 169        |
| A.6.1             | Objects, Classes and Attributes . . . . .                  | 170        |
| A.6.2             | Subclasses and Inheritance . . . . .                       | 172        |
| <b>B</b>          | <b>Mathematics behind the skip-gram model</b>              | <b>175</b> |