

1. Record Nr.	UNINA9910861088903321
Autore	Aggarwal Charu C
Titolo	Probability and Statistics for Machine Learning : A Textbook
Pubbl/distr/stampa	Cham : , : Springer, , 2024 ©2024
ISBN	3-031-53282-1
Edizione	[1st ed.]
Descrizione fisica	1 online resource (530 pages)
Disciplina	006.31
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Contents -- Preface -- Acknowledgments -- Author Biography -- 1 Probability and Statistics: An Introduction -- 1.1 Introduction -- 1.1.1 The Interplay Between Probability, Statistics, and Machine Learning -- 1.1.2 Chapter Organization -- 1.2 Representing Data -- 1.2.1 Numeric Multidimensional Data -- 1.2.2 Categorical and Mixed Attribute Data -- 1.3 Summarizing and Visualizing Data -- 1.4 The Basics of Probability and Probability Distributions -- 1.4.1 Populations versus Samples -- 1.4.2 Modeling Populations with Samples -- 1.4.3 Handling Dependence in Data Samples -- 1.5 Hypothesis Testing -- 1.6 Basic Problems in Machine Learning -- 1.6.1 Clustering -- 1.6.2 Classification and Regression Modeling -- 1.6.2.1 Regression -- 1.6.3 Outlier Detection -- 1.7 Summary -- 1.8 Further Reading -- 1.9 Exercises -- 2 Summarizing and Visualizing Data -- 2.1 Introduction -- 2.1.1 Chapter Organization -- 2.2 Summarizing Data -- 2.2.1 Univariate Summarization -- 2.2.1.1 Measures of Central Tendency -- 2.2.1.2 Measures of Dispersion -- 2.2.2 Multivariate Summarization -- 2.2.2.1 Covariance and Correlation -- 2.2.2.2 Rank Correlation Measures -- 2.2.2.3 Correlations among Multiple Attributes -- 2.2.2.4 Contingency Tables for Categorical Data -- 2.3 Data Visualization -- 2.3.1 Univariate Visualization -- 2.3.1.1 Histogram -- 2.3.1.2 Box Plot -- 2.3.2 Multivariate Visualization -- 2.3.2.1 Line Plot -- 2.3.2.2 Scatter Plot -- 2.3.2.3 Bar Chart -- 2.4 Applications to Data Preprocessing -- 2.4.1 Univariate Preprocessing Methods -- 2.4.2 Whitening: A Multivariate Preprocessing Method -- 2.5 Summary -- 2.6

Further Reading -- 2.7 Exercises -- 3 Probability Basics and Random Variables -- 3.1 Introduction -- 3.1.1 Chapter Organization -- 3.2 Sample Spaces and Events -- 3.3 The Counting Approach to Probabilities -- 3.4 Set-Wise View of Events. 3.5 Conditional Probabilities and Independence -- 3.6 The Bayes Rule -- 3.6.1 The Observability Perspective: Posteriors versus Likelihoods -- 3.7 The Basics of Probability Distributions -- 3.7.1 Closed-Form View of Probability Distributions -- 3.7.2 Continuous Distributions -- 3.7.3 Multivariate Probability Distributions -- 3.8 Distribution Independence and Conditionals -- 3.8.1 Independence of Distributions -- 3.8.2 Conditional Distributions -- 3.8.3 Example: A Simple 1-Dimensional Knowledge-Based Bayes Classifier -- 3.9 Summarizing Distributions -- 3.9.1 Expectation and Variance -- 3.9.2 Distribution Covariance -- 3.9.3 Useful Multivariate Properties Under Independence -- 3.10 Compound Distributions -- 3.10.1 Total Probability Rule in Continuous Hypothesis Spaces -- 3.10.2 Bayes Rule in Continuous Hypothesis Spaces -- 3.11 Functions of Random Variables (*) -- 3.11.1 Distribution of the Function of a Single Random Variable -- 3.11.2 Distribution of the Sum of Random Variables -- 3.11.3 Geometric Derivation of Distributions of Functions -- 3.12 Summary -- 3.13 Further Reading -- 3.14 Exercises -- 4 Probability Distributions -- 4.1 Introduction -- 4.1.1 Chapter Organization -- 4.2 The Uniform Distribution -- 4.3 The Bernoulli Distribution -- 4.4 The Categorical Distribution -- 4.5 The Geometric Distribution -- 4.6 The Binomial Distribution -- 4.7 The Multinomial Distribution -- 4.8 The Exponential Distribution -- 4.9 The Poisson Distribution -- 4.10 The Normal Distribution -- 4.10.0.1 Closure Properties of the Normal Distribution Family -- 4.10.1 Multivariate Normal Distribution: Independent Attributes -- 4.10.2 Multivariate Normal Distribution: Dependent Attributes -- 4.11 The Student's t-Distribution -- 4.12 The 2-Distribution -- 4.12.1 Application: Mahalanobis Method for Outlier Detection -- 4.13 Mixture Distributions: The Realistic View. 4.13.1 Why Mixtures are Ubiquitous: A Motivating Example -- 4.13.2 The Basic Generative Process of a Mixture Model -- 4.13.3 Some Useful Results for Prediction -- 4.13.4 The Conditional Independence Assumption -- 4.14 Moments of Random Variables (*) -- 4.14.1 Central and Standardized Moments -- 4.14.2 Moment Generating Functions -- 4.15 Summary -- 4.16 Further Reading -- 4.17 Exercises -- 5 Hypothesis Testing and Confidence Intervals -- 5.1 Introduction -- 5.1.1 Chapter Organization -- 5.2 The Central Limit Theorem -- 5.3 Sampling Distribution and Standard Error -- 5.4 The Basics of Hypothesis Testing -- 5.4.1 Confidence Intervals -- 5.4.2 When Population Standard Deviations Are Not Available -- 5.4.3 The One-Tailed Hypothesis Test -- 5.5 Hypothesis Tests For Differences in Means -- 5.5.1 Unequal Variance t-Test -- 5.5.1.1 Tightening the Degrees of Freedom -- 5.5.2 Equal Variance t-Test -- 5.5.3 Paired t-Test -- 5.6 2-Hypothesis Tests -- 5.6.1 Standard Deviation Hypothesis Test -- 5.6.2 2-Goodness-of-Fit Test -- 5.6.3 Independence Tests -- 5.7 Analysis of Variance (ANOVA) -- 5.8 Machine Learning Applications of Hypothesis Testing -- 5.8.1 Evaluating the Performance of a Single Classifier -- 5.8.2 Comparing Two Classifiers -- 5.8.3 2-Statistic for Feature Selection in Text -- 5.8.4 Fisher Discriminant Index for Feature Selection -- 5.8.5 Fisher Discriminant Index for Classification (*) -- 5.8.5.1 Most Discriminating Direction for the Two-Class Case -- 5.8.5.2 Most Discriminating Direction for Multiple Classes -- 5.9 Summary -- 5.10 Further Reading -- 5.11 Exercises -- 6 Reconstructing Probability Distributions -- 6.1 Introduction -- 6.1.1 Chapter Organization -- 6.2 Maximum Likelihood

Estimation -- 6.2.1 Comparing Likelihoods with Posteriors -- 6.3
Reconstructing Common Distributions from Data -- 6.3.1 The Uniform
Distribution.
6.3.2 The Bernoulli Distribution -- 6.3.3 The Geometric Distribution --
6.3.4 The Binomial Distribution -- 6.3.5 The Multinomial Distribution
-- 6.3.6 The Exponential Distribution -- 6.3.7 The Poisson Distribution
-- 6.3.8 The Normal Distribution -- 6.3.9 Multivariate Distributions
with Dimension Independence -- 6.3.10 Gaussian Distribution with
Dimension Dependence -- 6.4 Mixture of Distributions: The EM
Algorithm -- 6.5 Kernel Density Estimation -- 6.6 Reducing
Reconstruction Variance -- 6.6.1 Variance in Maximum Likelihood
Estimation -- 6.6.2 Prior Beliefs with Maximum A Posteriori (MAP)
Estimation -- 6.6.2.1 Example: Laplacian Smoothing -- 6.6.3 Kernel
Density Estimation: Role of Bandwidth -- 6.7 The Bias-Variance Trade-
Off -- 6.8 Popular Distributions Used as Conjugate Priors (*) -- 6.8.1
Gamma Distribution -- 6.8.2 Beta Distribution -- 6.8.3 Dirichlet
Distribution -- 6.9 Summary -- 6.10 Further Reading -- 6.11 Exercises
-- 7 Regression -- 7.1 Introduction -- 7.1.1 Chapter Organization --
7.2 The Basics of Regression -- 7.2.1 Interpreting the Coefficients --
7.2.2 Feature Engineering Trick for Dropping Bias -- 7.2.3 Regression:
A Central Problem in Statistics and Linear Algebra -- 7.3 Two
Perspectives on Linear Regression -- 7.3.1 The Linear Algebra
Perspective -- 7.3.2 The Probabilistic Perspective -- 7.3.2.1 Example:
Regression with L1-Loss -- 7.4 Solutions to Linear Regression -- 7.4.1
Closed-Form Solution to Squared-Loss Regression -- 7.4.2 The Case of
One Non-Trivial Predictor Variable -- 7.4.3 Solution with Gradient
Descent for Squared Loss -- 7.4.3.1 Stochastic Gradient Descent --
7.4.4 Gradient Descent For L1-Loss Regression -- 7.5 Handling
Categorical Predictors -- 7.6 Overfitting and Regularization -- 7.6.1
Closed-Form Solution for Regularized Formulation -- 7.6.2 Solution
Based on Gradient Descent -- 7.6.3 LASSO Regularization.
7.7 A Probabilistic View of Regularization -- 7.8 Evaluating Linear
Regression -- 7.8.1 Evaluating In-Sample Properties of Regression --
7.8.1.1 Correlation Versus R²-Statistic -- 7.8.2 Out-of-Sample
Evaluation -- 7.9 Nonlinear Regression -- 7.9.1 Interpretable Feature
Engineering -- 7.9.2 Explicit Feature Engineering with Similarity
Matrices -- 7.9.3 Implicit Feature Engineering with Similarity Matrices
-- 7.10 Summary -- 7.11 Further Reading -- 7.12 Exercises -- 8
Classification: A Probabilistic View -- 8.1 Introduction -- 8.1.1 Chapter
Organization -- 8.2 Generative Probabilistic Models -- 8.2.1
Continuous Numeric Data: The Gaussian Distribution -- 8.2.1.1
Prediction -- 8.2.1.2 Handling Overfitting -- 8.2.2 Binary Data: The
Bernoulli Distribution -- 8.2.2.1 Prediction -- 8.2.2.2 Handling
Overfitting -- 8.2.3 Sparse Numeric Data: The Multinomial Distribution
-- 8.2.3.1 Prediction -- 8.2.3.2 Handling Overfitting -- 8.2.3.3
Extending Multinomial Distributions to Real-Valued Data -- 8.2.4 Plate
Diagrams for Generative Processes -- 8.3 Loss-Based Formulations: A
Probabilistic View -- 8.3.1 Least-Squares Classification -- 8.3.1.1 The
Probabilistic Interpretation and Its Problems -- 8.3.1.2 Practical Issues
with Least Squares Classification -- 8.3.2 Logistic Regression --
8.3.2.1 Maximum Likelihood Estimation for Logistic Regression --
8.3.2.2 Gradient Descent and Stochastic Gradient Descent -- 8.3.2.3
Interpreting Updates in Terms of Error Probabilities -- 8.3.3
Multinomial Logistic Regression -- 8.3.3.1 The Probabilistic Model --
8.3.3.2 Maximum Likelihood Estimation -- 8.3.3.3 Gradient Descent
and Stochastic Gradient Descent -- 8.3.3.4 Probabilistic Interpretation
of Gradient Descent Updates -- 8.4 Beyond Classification: Ordered
Logit Model -- 8.4.1 Maximum Likelihood Estimation for Ordered Logit

-- 8.5 Summary -- 8.6 Further Reading -- 8.7 Exercises.
9 Unsupervised Learning: A Probabilistic View.
