- It is mainly used for market basket analysis and helps to understand the products that can be bought together.
- It can also be used in the healthcare field to find drug reactions in patients.

10. Principle Component Analysis:

- Principle Component Analysis (PCA) is an unsupervised learning technique, which is used for dimensionality reduction.
- It helps in reducing the dimensionality of the dataset that contains many features correlated with each other.
- It is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation.
- It is one of the popular tools that is used for exploratory data analysis and predictive modelling.
- PCA works by considering the variance of each attribute because the high variance shows the good split between the classes, and hence it reduces the dimensionality.
- Some real-world applications of PCA are image processing, movie recommendation system, optimizing the power allocation in various communication channels.

Probability and Statistics Books for Machine Learning

- Probability and statistics both are the most important concepts for Machine Learning. Probability is about predicting the likelihood of future events, while statistics involves the analysis of the frequency of past events.
- Machine Learning has become one of the first choices for most freshers and IT professionals. But, in order to enter this field, one must have some pre-specified skills and one of those skills in Mathematics.
- Yes, Mathematics is very much important to learn ML technology and develop efficient applications for the business.
- When talking about mathematics for Machine Learning, it especially focuses on Probability and Statistics, which are the essential topics to get started with ML.
- Probability and statistics are considered as the base foundation for ML and data science to develop ML algorithms and build decision-making capabilities.
- Also, Probability and statistics are the primary prerequisites to learn ML.

- In this topic, we will discuss a few important books on Probability and statistics that help you in making the ML process easy and implementing algorithms to business scenarios too.
- Here, we will discuss some of the best books for Probability and Statistics from basic to advanced levels.

Probability in Machine Learning:

- Probability is the bedrock of ML, which tells how likely is the event to occur.
- The value of Probability always lies between 0 to 1.
- It is the core concept as well as a primary prerequisite to understanding the ML models and their applications.
- Probability can be calculated by the number of times the event occurs divided by the total number of possible outcomes.

Let's suppose we tossed a coin, then the probability of getting head as a possible outcome can be calculated as below formula:

P (H) = Number of ways to head occur/ total number of possible outcomes

P (H) = ½

P (H) = 0.5

Where;

P (H) = Probability of occurring Head as outcome while tossing a coin.

Types of Probability:

For better understanding the Probability, it can be categorized further in different types as follows:

Empirical Probability: Empirical Probability can be calculated as the number of times the event occurs divided by the total number of incidents observed.

Theoretical Probability: Theoretical Probability can be calculated as the number of ways the particular event can occur divided by the total number of possible outcomes.

Joint Probability: It tells the Probability of simultaneously occurring two random events.

$P(A \cap B) = P(A).\ P(B)$

Where;

$P(A \cap B)$ = Probability of occurring events A and B both.

P (A) = Probability of event A

P (B) = Probability of event B

Conditional Probability: It is given by the Probability of event A given that event B occurred.

The Probability of an event A conditioned on an event B is denoted and defined as;

$P(A|B) = P(A \cap B)/P(B)$

Similarly, $P(B|A) = P(A \cap B)/ P(A)$. We can write the joint Probability of as A and B as $P(A \cap B)= p(A).P(B|A)$, which means: "The chance of both things happening is the chance that the first one happens, and then the second one is given when the first thing happened."

We have a basic understanding of Probability required to learn Machine Learning. Now, we will discuss the basic introduction of Statistics for ML.

Statistics in Machine Learning

- Statistics is also considered as the base foundation of machine learning which deals with finding answers to the questions that we have about data.

In general, we can define statistics as:

- Statistics is the part of applied Mathematics that deals with studying and developing ways for gathering, analyzing, interpreting and drawing conclusion from empirical data.

- It can be used to perform better-informed business decisions.

- Statistics can be categorized into 2 major parts.

These are as follows:

  - Descriptive Statistics
  - Inferential Statistics

Use of Statistics in ML:

- Statistics methods are used to understand the training data as well as interpret the results of testing different machine learning models.
- Further, Statistics can be used to make better-informed business and investing decisions.

Best Probability and Statistics books for Machine Learning:

- Probability and statistics both are equally important for learning Machine learning technology, but the main question is regarding the best books or sources of learning Probability and statistics for ML.
- Although there are so many books available over the internet as well as offline stores choosing the best appropriate book is the main problem for aspirants.

There are a few best books on Probability and Statistics are given as follows:

1. Probability for Statistics and Machine Learning

Authors of the Book:Anirban DasGupta

Price (Amazon):$118.15

Star Ratings: 3.6/5

Overview: This book is written by Anirban Das Gupta, which includes all fundamental and advanced topics of Probability and Statistics for ML. As per the different reviews, this is one of the best books available in both online and offline modes. This book mainly consists of the

unification of Probability, statistics, and machine learning tools that provides a complete background for self-study and future research in multiple areas.

Topic covered in this book:

- Review of Univariate Probability
- Multivariate Discrete Distributions
- Multidimensional Densities
- Advanced Distribution Theory
- Multivariate Normal and Related Distributions
- Finite Sample Theory of Order Statistics and Extremes
- Essential Asymptotics and Applications
- Characteristic Functions and Applications
- Asymptotic of Extremes and Order Statistics
- Markov Chains and Applications
- Random Walks
- Brownian Motion and Gaussian Processes
- Poisson Processes and Applications
- Discrete-Time Martingales and Concentration Inequalities
- Probability Metrics
- Empirical Processes and VC Theory
- Large Deviations
- The Exponential Family and Statistical Applications
- Simulation and Markov Chain Monte Carlo
- Useful Tools for Statistics and Machine Learning

2. Python for Probability, Statistics, and Machine Learning

Authors of the Book:José Unpingco

Price (Amazon):$ 82.36

Star Ratings:4.4/5

This book is available with the latest Python version 3.6+, which includes all essential areas of Probability, Statistics, and ML illustrated using Python. This book gives you exposure to various machine learning methods and examples using different analytical methods and Python codes which help you in deploying your theoretical concepts into real-time scenarios. It also provides detailed descriptions of various important results using modern Python libraries such as Pandas, Scikit-learn, TensorFlow, and Keras. Many abstract mathematical ideas, such as convergence in probability theory, are developed and illustrated with numerical examples.

Topics covered in this book: This book is divided into 5 chapters as follows:

- Getting Started with Scientific Python
- Probability
- Statistics
- Machine Learning
- Correction to: Probability

3. An Introduction to Statistical Learning

Authors of the Book:Gareth James, Daniela Witten, Trevor Hastie and Rob Tibshirani

Price (Amazon):$29.22

Star Ratings: 4.5/5

Overview: An Introduction to Statistical Learning with application in R is offered by Springer in two editions. Statistics is one of the main toolkits for Machine learning and data scientists' aspirants. This book provides a broad and less technical treatment of key topics in statistical learning with the help of R. This book is suitable for all users who want good exposure to data analysis with statistics learning. This book is available in various languages such as Chinese, Italian, Japanese, Korean, Mongolian, Russian and Vietnamese. The authors of this book Gareth James, Daniela Witten, Trevor Hastie and Rob Tibshirani, have divided this book into two editions.

Topics covered in this book:

1st Edition of this book covers the following topics:

- Sparse methods for classification and regression
- Decision trees
- Boosting
- Support vector machines
- Clustering

2nd edition of this book covers the following topics:

- Deep learning
- Survival analysis
- Multiple testing
- Naive Bayes and generalized linear models
- Bayesian additive regression trees
- Matrix completion

4. The Elements of Statistical Learning:

Authors of Book: Jerome Friedman, Trevor Hastie, and Robert Tibshirani

Price:$84.95 (Amazon)

Star Ratings:4.6/5

Overview: The books illustrate important ideas in different fields such as medical, finance, marketing, etc., which is a reference of a common framework.

As this book shows the statistical approach, hence it mainly focuses on explaining the concepts rather than mathematics. It contains different examples of each topic with different colour graphics.

This book is one of the best resources for Machine Learning professionals and one who is interested in data mining concepts. The various concepts of the book range from supervised to unsupervised learning.

It includes different important topics such as neural network, support vector machine, Classification trees and boosting. This book also contains a chapter on methods for "wide" data (p bigger than n) along with multiple testing and false discovery rates.

5. Probability and Statistical Inference

Author: Robert V. Hogg, Elliot Tanis, and Dale Zimmerman

Price on Amazon: $181.99

Star Rating: 4.9/5

Overview: This book is written and designed by three popular statisticians named Robert V. Hogg, Elliot Tanis, and Dale Zimmerman. The latest edition of this book is the tenth edition, which focuses on the existence of variation in each process, and also helps readers to understand this variation with the help of Probability and Statistics.

The book includes the applied introduction to Probability and statistics that reinforces the mathematical concepts with different real-world examples and applications. These examples also illustrate relevance to the key concepts of statistics. The book's syllabus is designed for two-semester courses, but it can be completed in a one-semester course only.

There is no requirement to have knowledge of Probability and statistics to read this book, but sound knowledge of calculus is required.

This book includes popular concepts of Probability and statistics such as Probability, Conditional Probability, Bayes' Theorem, statistical hypotheses, standard chi-square tests, analysis of variance including general factorial designs, and some procedures associated with regression, correlation, and statistical quality control, etc.