COMPSCI 589
Lecture 1: Course Overview - Supervised and Unsupervised Learning

Benjamin M. Marlin

College of Information and Computer Sciences
University of Massachusetts Amherst

Slides by Benjamin M. Marlin (marlin@cs.umass.edu).
Created with support from National Science Foundation Award# IIS-1350522.

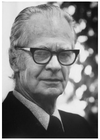
Outline

1. Introduction
2. About the Course
3. Review and Notation

Introduction

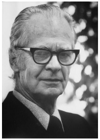
What is Learning?

Definitions of Learning

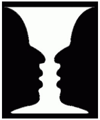


Behaviorism (Skinner, 1900-1950): Learning is a long-term change in behavior due to experience.

Definitions of Learning

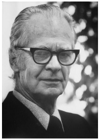


Behaviorism (Skinner, 1900-1950): Learning is a long-term change in behavior due to experience.

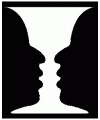


Cognitivism (Gestalt School, 1920-): Learning is an internal mental process that integrates new information into established mental frameworks and updates those frameworks over time.

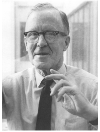
Definitions of Learning



Behaviorism (Skinner, 1900-1950): Learning is a long-term change in behavior due to experience.



Cognitivism (Gestalt School, 1920-): Learning is an internal mental process that integrates new information into established mental frameworks and updates those frameworks over time.



Connectionism (Hebb, 1949): Learning is a physical process in which neurons join by developing the synapses between them.

Introduction

What is Machine Learning?

Views on Machine Learning

Samuel (1959): "Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed."

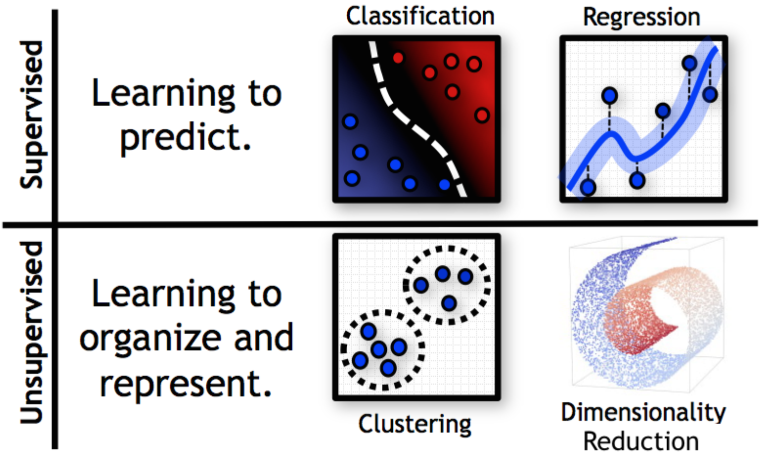
Views on Machine Learning

Samuel (1959): "Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed."

Mitchell (1997): "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Substitute "training data D" for "experience E."

Machine Learning Tasks



Machine Learning Approaches

- **Non-Parametric:** Learning is accomplished by storing the training data (memorization).

Machine Learning Approaches

- **Non-Parametric:** Learning is accomplished by storing the training data (memorization).
- **Parametric:** Learning is accomplished by using an algorithm to adapt the parameters in a mathematical or statistical model given training data. For exmaple:

Machine Learning Approaches

- **Non-Parametric:** Learning is accomplished by storing the training data (memorization).
- **Parametric:** Learning is accomplished by using an algorithm to adapt the parameters in a mathematical or statistical model given training data. For exmaple:

$$f_\theta(\mathbf{x}) = \sum_{d=1}^{D} \theta_d x_d$$

Machine Learning Applications

Machine Learning in Industry

xerox
amazon.com
AdRoll
Baidu Research
Google
Symantec
QUALCOMM
facebook
TOYOTA
BOSCH
Microsoft Research
TWO SIGMA
Booking.com
YAHOO! LABS
D E Shaw & Co
NVIDIA
twitter
IBM Research
Disney Research

Relationship to Other Fields

- Machine Learning and Artificial Intelligence

Relationship to Other Fields

- Machine Learning and Artificial Intelligence
- Machine Learning and Probability/Statistics

Relationship to Other Fields

- Machine Learning and Artificial Intelligence
- Machine Learning and Probability/Statistics
- Machine Learning and Numerical Optimization

Relationship to Other Fields

- Machine Learning and Artificial Intelligence
- Machine Learning and Probability/Statistics
- Machine Learning and Numerical Optimization
- Machine Learning and Function Approximation

Relationship to Other Fields

- Machine Learning and Artificial Intelligence
- Machine Learning and Probability/Statistics
- Machine Learning and Numerical Optimization
- Machine Learning and Function Approximation
- Machine Learning and Cognitive Science

Relationship to Other Fields

- Machine Learning and Artificial Intelligence
- Machine Learning and Probability/Statistics
- Machine Learning and Numerical Optimization
- Machine Learning and Function Approximation
- Machine Learning and Cognitive Science
- Machine Learning and Neuroscience

Relationship to Other Fields

- Machine Learning and Artificial Intelligence
- Machine Learning and Probability/Statistics
- Machine Learning and Numerical Optimization
- Machine Learning and Function Approximation
- Machine Learning and Cognitive Science
- Machine Learning and Neuroscience
- Machine Learning and Data Mining

Relationship to Other Fields

- Machine Learning and Artificial Intelligence
- Machine Learning and Probability/Statistics
- Machine Learning and Numerical Optimization
- Machine Learning and Function Approximation
- Machine Learning and Cognitive Science
- Machine Learning and Neuroscience
- Machine Learning and Data Mining
- Machine Learning and Data Science

Relationship to Other Fields

- Machine Learning and Artificial Intelligence
- Machine Learning and Probability/Statistics
- Machine Learning and Numerical Optimization
- Machine Learning and Function Approximation
- Machine Learning and Cognitive Science
- Machine Learning and Neuroscience
- Machine Learning and Data Mining
- Machine Learning and Data Science
- Machine Learning and Big Data

Outline

1. Introduction
2. About the Course
3. Review and Notation

Course Goals

The aim of this course is to develop the knowledge and skills necessary to effectively apply existing machine learning models and algorithms to solve real-world problems. The course will cover:

Course Goals

The aim of this course is to develop the knowledge and skills necessary to effectively apply existing machine learning models and algorithms to solve real-world problems. The course will cover:

- Classification, regression, clustering, dimensionality reduction and representation learning

Course Goals

The aim of this course is to develop the knowledge and skills necessary to effectively apply existing machine learning models and algorithms to solve real-world problems. The course will cover:

- Classification, regression, clustering, dimensionality reduction and representation learning
- Model selection, regularization, design of experiments, model evaluation

Course Goals

The aim of this course is to develop the knowledge and skills necessary to effectively apply existing machine learning models and algorithms to solve real-world problems. The course will cover:

- Classification, regression, clustering, dimensionality reduction and representation learning
- Model selection, regularization, design of experiments, model evaluation
- Use of machine learning across different computing contexts (desktop/cluster/cloud)

Course Goals

The aim of this course is to develop the knowledge and skills necessary to effectively apply existing machine learning models and algorithms to solve real-world problems. The course will cover:

- Classification, regression, clustering, dimensionality reduction and representation learning
- Model selection, regularization, design of experiments, model evaluation
- Use of machine learning across different computing contexts (desktop/cluster/cloud)

This course **will not** teach you how to design new machine learning models and algorithms.

Prerequisites

The course has formal prerequisites as listed below. All students are expected to be familiar with this material or have the ability to make up any gaps in their backgrounds on their own.

- Linear Algebra

Prerequisites

The course has formal prerequisites as listed below. All students are expected to be familiar with this material or have the ability to make up any gaps in their backgrounds on their own.

- Linear Algebra
- Calculus

Prerequisites

The course has formal prerequisites as listed below. All students are expected to be familiar with this material or have the ability to make up any gaps in their backgrounds on their own.

- Linear Algebra
- Calculus
- Probability and Statistics

Prerequisites

The course has formal prerequisites as listed below. All students are expected to be familiar with this material or have the ability to make up any gaps in their backgrounds on their own.

- Linear Algebra
- Calculus
- Probability and Statistics
- Algorithms and Data Structures

Prerequisites

The course has formal prerequisites as listed below. All students are expected to be familiar with this material or have the ability to make up any gaps in their backgrounds on their own.

- Linear Algebra
- Calculus
- Probability and Statistics
- Algorithms and Data Structures

The course requires the use of Python for programming. Students are expected to learn Python as we go.

Text Books

The course will use a two textbooks freely available from the authors:

Text Books

The course will use a two textbooks freely available from the authors:

- [ISL]: *An Introduction to Statistical Learning*. James, Witten, Hastie and Tibshirani.

Text Books

The course will use a two textbooks freely available from the authors:

- [ISL]: *An Introduction to Statistical Learning*. James, Witten, Hastie and Tibshirani.
- [ESL]: *The Elements of Statistical Learning, Second Edition*. Hastie, Tibshirani and Friedman.

Text Books

The course will use a two textbooks freely available from the authors:

- [ISL]: *An Introduction to Statistical Learning*. James, Witten, Hastie and Tibshirani.
- [ESL]: *The Elements of Statistical Learning, Second Edition*. Hastie, Tibshirani and Friedman.

Readings are intended to be completed before class.

Programming and Computing

- Students need access to computing to complete regular assignments (any moderately recent laptop/desktop should do).

Programming and Computing

- Students need access to computing to complete regular assignments (any moderately recent laptop/desktop should do).
- Programming assignments will use Python 2.7.

Programming and Computing

- Students need access to computing to complete regular assignments (any moderately recent laptop/desktop should do).
- Programming assignments will use Python 2.7.
- A complete Ubuntu programming environment will be distributed using Vagrant/Virtual Box.

Programming and Computing

- Students need access to computing to complete regular assignments (any moderately recent laptop/desktop should do).
- Programming assignments will use Python 2.7.
- A complete Ubuntu programming environment will be distributed using Vagrant/Virtual Box.
- Access to cloud computing resources is required to complete course components using Apache Spark.

Outline

Linear Algebra

Definition: Vector Space

The real vector space $\mathbb{R}^n$ is a set with elements $\mathbf{x} = [x_1, ..., x_n]$ where each $x_i \in \mathbb{R}$. The elements $\mathbf{x}$ are called vectors, and they satisfy the following properties:

Linear Algebra

Definition: Vector Space

The real vector space $\mathbb{R}^n$ is a set with elements $\mathbf{x} = [x_1, ..., x_n]$ where each $x_i \in \mathbb{R}$. The elements $\mathbf{x}$ are called vectors, and they satisfy the following properties:

- **Addition:** If $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$, then $\mathbf{x} + \mathbf{y} = [x_1 + y_1, ..., x_n + y_n] \in \mathbb{R}^n$.

Linear Algebra

Definition: Vector Space

The real vector space $\mathbb{R}^n$ is a set with elements $\mathbf{x} = [x_1, ..., x_n]$ where each $x_i \in \mathbb{R}$. The elements $\mathbf{x}$ are called vectors, and they satisfy the following properties:

- **Addition:** If $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$, then $\mathbf{x} + \mathbf{y} = [x_1 + y_1, ..., x_n + y_n] \in \mathbb{R}^n$.
- **Scalar Product:** If $\mathbf{x} \in \mathbb{R}^n$ and $a \in \mathbb{R}$, then $a\mathbf{x} = [ax_1, ..., ay_n] \in \mathbb{R}^n$.

Linear Algebra

Definition: Vector Space

The real vector space $\mathbb{R}^n$ is a set with elements $\mathbf{x} = [x_1, ..., x_n]$ where each $x_i \in \mathbb{R}$. The elements $\mathbf{x}$ are called vectors, and they satisfy the following properties:

- **Addition:** If $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$, then $\mathbf{x} + \mathbf{y} = [x_1 + y_1, ..., x_n + y_n] \in \mathbb{R}^n$.
- **Scalar Product:** If $\mathbf{x} \in \mathbb{R}^n$ and $a \in \mathbb{R}$, then $a\mathbf{x} = [ax_1, ..., ay_n] \in \mathbb{R}^n$.
- **Inner Product:** If $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$, then $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^n x_i y_i$.

Linear Algebra

Definition: Matrix

A matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ is rectangular array of elements $x_{ij} \in \mathbb{R}$, $1 \leq i \leq n$, $1 \leq j \leq m$:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$$

Linear Algebra

Definition: Matrix

A matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ supports the following operations:

Linear Algebra

Definition: Matrix

A matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ supports the following operations:

- **Addition:** If $\mathbf{X} \in \mathbb{R}^{n \times m}$, $\mathbf{Y} \in \mathbb{R}^{n \times m}$ and $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$, then $\mathbf{Z} \in \mathbb{R}^{n \times m}$ and $Z_{ij} = X_{ij} + Y_{ij}$.

Linear Algebra

Definition: Matrix

A matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ supports the following operations:

- **Addition:** If $\mathbf{X} \in \mathbb{R}^{n \times m}$, $\mathbf{Y} \in \mathbb{R}^{n \times m}$ and $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$, then $\mathbf{Z} \in \mathbb{R}^{n \times m}$ and $Z_{ij} = X_{ij} + Y_{ij}$.
- **Scalar Product:** If $\mathbf{X} \in \mathbb{R}^{n \times m}$, $a \in \mathbb{R}$ and $\mathbf{Z} = a\mathbf{X}$, then $\mathbf{Z} \in \mathbb{R}^{n \times m}$ and $Z_{ij} = aX_{ij}$.

Linear Algebra

Definition: Matrix

A matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ supports the following operations:

- **Addition:** If $\mathbf{X} \in \mathbb{R}^{n \times m}$, $\mathbf{Y} \in \mathbb{R}^{n \times m}$ and $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$, then $\mathbf{Z} \in \mathbb{R}^{n \times m}$ and $Z_{ij} = X_{ij} + Y_{ij}$.
- **Scalar Product:** If $\mathbf{X} \in \mathbb{R}^{n \times m}$, $a \in \mathbb{R}$ and $\mathbf{Z} = a\mathbf{X}$, then $\mathbf{Z} \in \mathbb{R}^{n \times m}$ and $Z_{ij} = aX_{ij}$.
- **Matrix Product:** If $\mathbf{X} \in \mathbb{R}^{n \times m}$, $\mathbf{Y} \in \mathbb{R}^{m \times n}$ and $\mathbf{Z} = \mathbf{XY}$, then $\mathbf{Z} \in \mathbb{R}^{n \times n}$ and $Z_{ij} = \sum_{k=1}^{m} X_{ik} Y_{kj}$.

Linear Algebra

Definition: Matrix

A matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ supports the following operations:

- **Addition:** If $\mathbf{X} \in \mathbb{R}^{n \times m}$, $\mathbf{Y} \in \mathbb{R}^{n \times m}$ and $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$, then $\mathbf{Z} \in \mathbb{R}^{n \times m}$ and $Z_{ij} = X_{ij} + Y_{ij}$.
- **Scalar Product:** If $\mathbf{X} \in \mathbb{R}^{n \times m}$, $a \in \mathbb{R}$ and $\mathbf{Z} = a\mathbf{X}$, then $\mathbf{Z} \in \mathbb{R}^{n \times m}$ and $Z_{ij} = aX_{ij}$.
- **Matrix Product:** If $\mathbf{X} \in \mathbb{R}^{n \times m}$, $\mathbf{Y} \in \mathbb{R}^{m \times n}$ and $\mathbf{Z} = \mathbf{XY}$, then $\mathbf{Z} \in \mathbb{R}^{n \times n}$ and $Z_{ij} = \sum_{k=1}^{m} X_{ik} Y_{kj}$.

Linear Algebra

Definition: Matrix

A matrix $\mathbf{X} \in \mathbb{R}^{n\times m}$ supports the following operations:

- **Addition:** If $\mathbf{X} \in \mathbb{R}^{n\times m}$, $\mathbf{Y} \in \mathbb{R}^{n\times m}$ and $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$, then $\mathbf{Z} \in \mathbb{R}^{n\times m}$ and $Z_{ij} = X_{ij} + Y_{ij}$.
- **Scalar Product:** If $\mathbf{X} \in \mathbb{R}^{n\times m}$, $a \in \mathbb{R}$ and $\mathbf{Z} = a\mathbf{X}$, then $\mathbf{Z} \in \mathbb{R}^{n\times m}$ and $Z_{ij} = aX_{ij}$.
- **Matrix Product:** If $\mathbf{X} \in \mathbb{R}^{n\times m}$, $\mathbf{Y} \in \mathbb{R}^{m\times n}$ and $\mathbf{Z} = \mathbf{XY}$, then $\mathbf{Z} \in \mathbb{R}^{n\times n}$ and $Z_{ij} = \sum_{k=1}^{m} X_{ik}Y_{kj}$.

Yous should be familiar with basic matrix types (square, diagonal, identity), basic matrix operations (transpose, inverse, trace, determinant, etc.), and matrix concepts (eigenvalues, eigenvectors, orthogonality, etc.).

Probability Distributions

Definition: Probability Distribution

A probability distribution P over a sample space Ω is a mapping from subsets of Ω to the real numbers that satisfies the following conditions:

Probability Distributions

Definition: Probability Distribution

A probability distribution P over a sample space Ω is a mapping from subsets of Ω to the real numbers that satisfies the following conditions:

- Non-negativity: $P(\alpha) \geq 0$ for all $\alpha \subseteq \Omega$

Probability Distributions

Definition: Probability Distribution

A probability distribution P over a sample space Ω is a mapping from subsets of Ω to the real numbers that satisfies the following conditions:

- Non-negativity: $P(\alpha) \geq 0$ for all $\alpha \subseteq \Omega$
- Normalization: $P(\Omega) = 1$

Probability Distributions

Definition: Probability Distribution

A probability distribution P over a sample space Ω is a mapping from subsets of Ω to the real numbers that satisfies the following conditions:

- Non-negativity: $P(\alpha) \geq 0$ for all $\alpha \subseteq \Omega$
- Normalization: $P(\Omega) = 1$
- Additivity: For all $\alpha, \beta \subseteq \Omega$ that are disjoint sets, $P(\alpha \bigcup \beta) = P(\alpha) + P(\beta)$

Random Variables

Definition: Random Variable

A random variable X is defined by a function f_X that maps each element ω of the sample space Ω to a value $f_X(\omega)$ in a set $\mathcal{X}$ called the *range* of the random variable.

Random Variables

Definition: Random Variable

A random variable X is defined by a function f_X that maps each element ω of the sample space Ω to a value $f_X(\omega)$ in a set $\mathcal{X}$ called the *range* of the random variable.

For each $x \in \mathcal{X}$ the event $\{X = x\}$ refers to the subset of the sample space $\{\omega | \omega \in \Omega, f_X(\omega) = x\}$.

Random Variables

Definition: Random Variable

A random variable X is defined by a function f_X that maps each element ω of the sample space Ω to a value $f_X(\omega)$ in a set $\mathcal{X}$ called the *range* of the random variable.

For each $x \in \mathcal{X}$ the event $\{X = x\}$ refers to the subset of the sample space $\{\omega | \omega \in \Omega, f_X(\omega) = x\}$.

For each $x \in \mathcal{X}$ the probability $P(X = x) = P(\{\omega | \omega \in \Omega, f_X(\omega) = x\})$.

Probability and Random Variables

We can also specify a probability distribution for a random variable X with range $\mathcal{X}$ directly instead of via an underlying sample space Ω. The following conditions must hold:

- **Discrete PMF:** $P(X = x) \geq 0 \;\; \forall x \in \mathcal{X}$ and $\sum_{x \in \mathcal{X}} P(X = x) = 1$.

Probability and Random Variables

We can also specify a probability distribution for a random variable X with range $\mathcal{X}$ directly instead of via an underlying sample space Ω. The following conditions must hold:

- **Discrete PMF:** $P(X = x) \geq 0 \ \ \forall x \in \mathcal{X}$ and $\sum_{x \in \mathcal{X}} P(X = x) = 1$.
- **Continuous PDF:** $p(X = x) \geq 0 \ \ \forall x \in \mathcal{X}$ and $\int_{\mathcal{X}} p(X = x) dx = 1$.

Random Variables and Data Sets

In machine learning and statistics, probability distributions are defined over data cases described by multiple attributes that are identified with random variables.

Random Variables and Data Sets

In machine learning and statistics, probability distributions are defined over data cases described by multiple attributes that are identified with random variables.

Example: Heart Disease Dataset

Gender	Blood Pressure	Cholesterol	Heart Disease
Male	Med	Low	No
Male	Hi	Hi	Yes
Male	Med	Med	Yes
Male	Med	Hi	No
Female	Med	Low	No
Male	Low	Med	No

Joint Probability Distributions

- A *joint probability distribution* is a probability distribution defined over a collection of random variables $(X_1, ..., X_m)$ with ranges $\mathcal{X}_1, ..., \mathcal{X}_m$: $P(X_1 = x_1, ..., X_m = x_m)$.

Joint Probability Distributions

- A *joint probability distribution* is a probability distribution defined over a collection of random variables $(X_1, ..., X_m)$ with ranges $\mathcal{X}_1, ..., \mathcal{X}_m$: $P(X_1 = x_1, ..., X_m = x_m)$.
- A joint distribution defined over random variables $X_1, ..., X_m$ must satisfy normalization and non-negativity with respect to the Cartesian product of their ranges $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$.

Joint Probability Distributions

- A *joint probability distribution* is a probability distribution defined over a collection of random variables $(X_1, ..., X_m)$ with ranges $\mathcal{X}_1, ..., \mathcal{X}_m$: $P(X_1 = x_1, ..., X_m = x_m)$.
- A joint distribution defined over random variables $X_1, ..., X_m$ must satisfy normalization and non-negativity with respect to the Cartesian product of their ranges $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$.
- Alternatively, a joint distribution can be viewed as a probability distribution over a single vector-valued random variable $\mathbf{X} = [X_1, ..., X_m]$ whose range is $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$.

Joint Distributions: Heart Disease Example

Consider the heart disease example. The joint distribution over the random variables *Gender*, *BloodPressure*, *Cholesterol* and *HeartDisease* is just a big table:

Joint Distributions: Heart Disease Example

Consider the heart disease example. The joint distribution over the random variables *Gender*, *BloodPressure*, *Cholesterol* and *HeartDisease* is just a big table:

Gender	BloodPressure	Cholesterol	HeartDisease	P
F	L	L	N	0.0127
F	L	L	Y	0.0007
F	L	M	N	0.0098
F	L	M	Y	0.0009
F	L	H	N	0.0087
F	L	H	Y	0.0010
⋮	⋮	⋮	⋮	⋮

Important Probability Concepts

You should be familiar with the following fundamental concepts from probability theory

Important Probability Concepts

You should be familiar with the following fundamental concepts from probability theory

- Marginalization

Important Probability Concepts

You should be familiar with the following fundamental concepts from probability theory

- Marginalization
- Conditioning

Important Probability Concepts

You should be familiar with the following fundamental concepts from probability theory

- Marginalization
- Conditioning
- Bayes Rules

Important Probability Concepts

You should be familiar with the following fundamental concepts from probability theory

- Marginalization
- Conditioning
- Bayes Rules
- Expectations

Important Probability Concepts

You should be familiar with the following fundamental concepts from probability theory

- Marginalization
- Conditioning
- Bayes Rules
- Expectations
- Classical Distributions (Bernoulli, Binomial, Multinomial, Gaussian)