Machine Learning Cheatsheet

This **machine learning cheatsheet** serves as a quick reference guide for key concepts and commonly used algorithms in machine learning. It includes essential topics such as supervised learning, unsupervised learning, and reinforcement learning, as well as commonly used algorithms like linear regression and decision trees. This **machine learning (ML) cheatsheet** is valuable for anyone interested in machine learning.

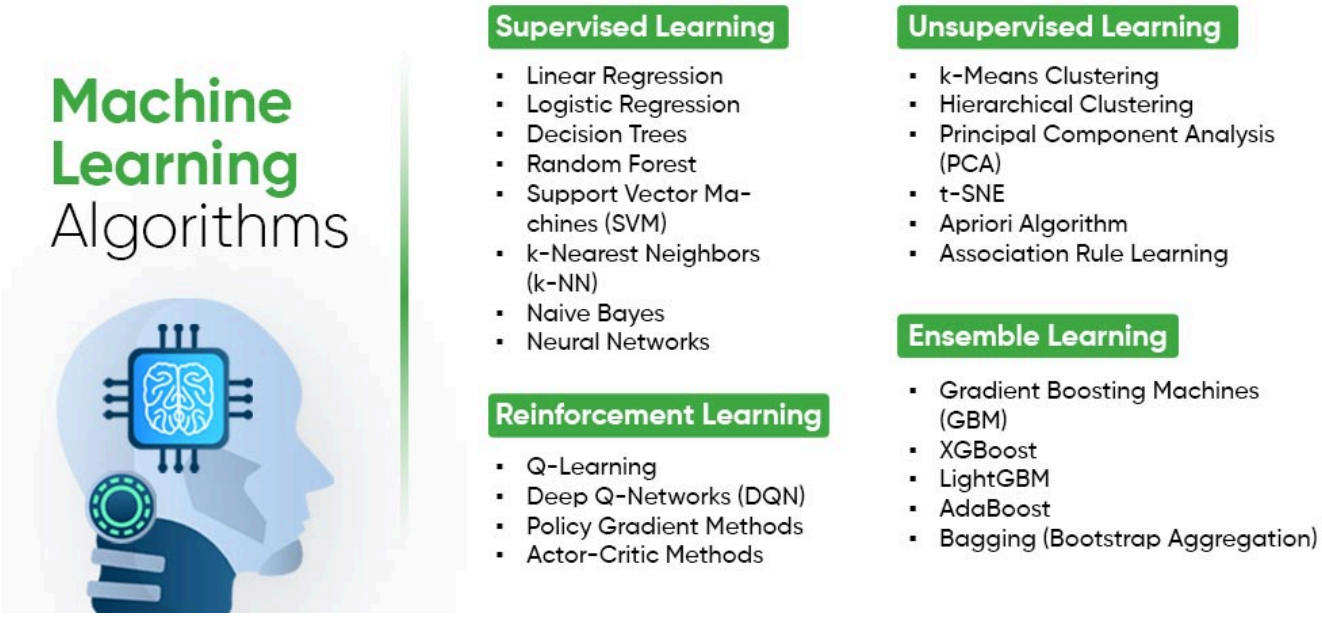
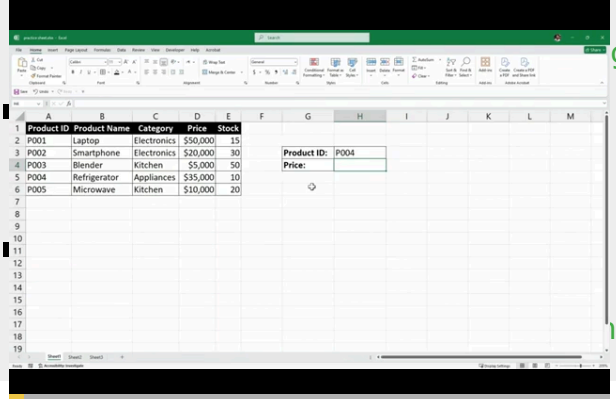


Table of Contents

Advertisement



gorithms

Algorithms

ıms

Supervised Machine Learning

Supervised machine learning is a type of machine learning that trains the algorithms using labeled datasets to predict outcomes.

The main objective of supervised learning is to make algorithms learn an association between input data samples and corresponding outputs after performing multiple training data instances.

Supervised Machine Learning Algorithms

Supervised learning algorithms are categorized into two types of tasks - classification and regression. Below, we have listed commonly used supervised machine learning algorithms, their applications, advantages and disadvantages.

Algorithm	Description	Applications	Advantages	Disadvantages
Linear Regression	Predicts a continuous numerical value based on a linear relationship between input and output variables.	Predicting house prices, stock prices, sales figures.	Simple to implement, interpretable, efficient.	Sensitive to outliers, assumes linearity.
Logistic Regression	Predicts a categorical value (e.g., binary classification) using a logistic function.	Classifying email as spam or not spam, predicting customer churn.	Interpretable, efficient, can handle categorical features.	Prone to overfitting, limited to linear relationships.

Advertisement

Ridge Regression	Regularized linear regression that adds a penalty term to the loss function to prevent overfitting.	Regression tasks, feature selection.	Can handle multicollinearity, improves model generalization.	Requires tuning the regularization parameter.
Lasso Regression	Regularized linear regression that adds a penalty term to the loss function to encourage sparsity (feature selection).	Regression tasks, feature selection.	Can handle multicollinearity, performs feature selection.	May introduce bias in feature selection.
K-Nearest Neighbors (KNN)	Classifies or predicts the value of a new data point based on the majority class or average value of its k nearest neighbors in the training dataset.	Classification, regression, recommendation systems.	Simple to implement, no training phase required, can handle non-linear relationships.	Can be computationally expensive for large datasets, sensitive to the choice of distance metric and the value of k.
Support Vector Machines (SVMs)	optimal hyperplane to separate data points into different classes.	classification, text classification, anomaly detection.	Effective for high-dimensional data, handles non-linear relationships with kernels.	Can be computationally expensive for large datasets, sensitive to outliers.

Advertisement

Decision Tree	Creates a tree-like model to make decisions based on a series of rules.	Classification, regression, predictive modeling.	Easy to understand and interpret, can handle both numerical and categorical features.	Prone to overfitting, can be sensitive to small changes in data.
Random Forests	An ensemble of decision trees, combining multiple models to improve accuracy and reduce overfitting.	Classification, regression, predictive modeling.	More accurate than individual decision trees, robust to noise and outliers.	Can be computationally expensive for large datasets.
Naive Bayes	A probabilistic classifier based on Bayes' theorem, assuming independence of features.	Text classification, spam filtering, sentiment analysis.	Simple to implement, efficient, can handle categorical and numerical features.	Assumes independence of features, which may not always hold true.
Gradient Boosting Regression	An ensemble method that iteratively trains weak models to improve accuracy.	Regression, classification, predictive modeling.	Highly accurate, can handle complex relationships.	Can be computationally expensive, requires careful tuning of hyperparameters.
XGBoost	A scalable and efficient gradient boosting framework.	Regression, classification, ranking.	Highly accurate, efficient, can handle large datasets.	Can be complex to configure.
LightGBM Regressor	A gradient boosting framework	Regression, classification, ranking.	Faster than XGBoost,	May have slightly lower accuracy

Advertisement

	that uses histograms and gradient boosting for efficient training.		efficient for large datasets.	than XGBoost in some cases.
Neural Networks (Deep Learning)	Complex models with multiple layers, capable of learning complex patterns and relationships.	Image classification, natural language processing, speech recognition.	Highly accurate, can handle complex tasks.	Can be computationally expensive, requires careful tuning of hyperparameters.

Unsupervised Machine Learning

Un[Advertisement]f machine learning that learns patterns and structures within the data without human supervision. Unsupervised learning uses machine learning algorithms to analyze the data and discover underlying patterns within unlabeled data sets.

Unsupervised Machine Learning Algorithms

Unsupervised learning algorithms are categorised into three categories – clustering, association, and dimensionality reduction. Below, we have listed commonly used

unsupervised machine learning algorithms, their applications, advantages and disadvantages.

Algorithm	Description	Applications	Advantages	Disadvantages
K-Means Clustering	Partitions data into K clusters based on similarity.	Customer segmentation, image segmentation, anomaly detection.	Simple to implement, efficient, can handle large datasets.	Requires specifying the number of clusters, sensitive to initialization.
Hierarchical Clustering	Creates a hierarchy of clusters, either agglomerative (bottom-up) or divisive (top-down).	Customer segmentation, image segmentation, outlier detection.	Can reveal hierarchical structures, doesn't require specifying the number of clusters.	Can be computationally expensive for large datasets, sensitive to distance metrics.
Principal Component Analysis (PCA)	Reduces the dimensionality of data while preserving the most important features.	Data visualization, feature engineering, noise reduction.	Efficient, can reveal underlying patterns in data.	May lose some information in the dimensionality reduction process.
Singular Value Decomposition (SVD)	Decomposes a matrix into its singular values and vectors.	Data analysis, recommendation systems, image compression.	Can be used for dimensionality reduction and feature extraction.	Can be computationally expensive for large matrices.
Independent Component Analysis (ICA)	Identifies independent sources of signals from mixed observations.	Blind source separation, signal processing.	Can separate mixed signals, useful in applications like speech recognition.	Can be sensitive to initialization and assumptions about the independence of sources.

Advertisement

Gaussian Mixture Model (GMM)	Models data as a mixture of Gaussian distributions, assuming each cluster is generated from a Gaussian distribution.	Clustering, density estimation, anomaly detection.	Can handle complex data distributions, flexible.	Can be computationally expensive, sensitive to initialization.
Apriori Algorithm	A frequent itemset mining algorithm used to discover associations between items in a dataset.	Market basket analysis, recommendation systems.	Efficient for finding frequent itemsets, can be used for association rule mining.	May not be suitable for large datasets with many items.
t-SNE	Non-linear dimensionality reduction technique that preserves local structure.	Data visualization, clustering, anomaly detection.	Effective for visualizing high-dimensional data in low-dimensional space.	Can be computationally expensive, sensitive to parameters.
UMAP	Another non-linear dimensionality reduction technique that preserves global structure and local	Data visualization, clustering, anomaly detection.	Often faster and more scalable than t-SNE, preserves global structure well.	May require careful parameter tuning.

Advertisement

Reinforcement Learning

Reinforcement learning is a type of machine learning where an agent (generally a software entity) is trained to interpret the environment by performing actions and monitoring the results. For every good action, the agent gets positive feedback and for every bad action the agent gets negative feedback. It's inspired by how animals learn from their experiences, making decisions based on the consequences of their actions.

Reinforcement Learning Algorithms

In this section, we have listed some well known reinforcement learning algorithms, their applications, advantages and disadvantages.

Algorithm	Description	Applications	Advantages	Disadvantages
Q-Learning	Off-policy learning algorithm that learns the optimal action-value function.	Game playing, robotics, control systems.	Simple to implement, can handle complex environments.	Can be computationally expensive for large state spaces.
SARSA	On-policy learning algorithm that updates the action-value function based on the current policy.	Game playing, robotics, control systems.	Can handle continuous action spaces, suitable for online learning.	Can be sensitive to exploration-exploitation trade-off.
Deep Q-Networks (DQN)	Combines deep learning with Q-learning, using a neural network to approximate the action-value function.	Atari game playing, robotics, self-driving cars.	Can handle complex environments with large state and action spaces.	Requires careful tuning of hyperparameters, can be computationally expensive.
Policy Gradients	Directly optimizes the policy to maximize rewards.	Robotics, game playing, language processing.	Can handle continuous action spaces, can be more sample efficient than value-based methods.	Can be sensitive to noise and instability.
Actor-Critic	Combines policy-based and value-based	Robotics, game playing, natural	Can be more stable and efficient than pure policy-	Requires careful balancing of exploration and exploitation.

Advertisement

	methods, using both a policy function and a value function.	language processing.	based or value-based methods.	
Asynchronous Advantage Actor-Critic (A3C)	A parallel version of actor-critic that can handle complex environments with large state spaces.	Robotics, game playing, natural language processing.	Can be more efficient than traditional actor-critic methods, suitable for distributed training.	Can be complex to implement.

TOP TUTORIALS

Python Tutorial

Java Tutorial

C++ Tutorial

C Programming Tutorial

C# Tutorial

PHP Tutorial

R Tutorial

HTML Tutorial

CSS Tutorial

JavaScript Tutorial

SQL Tutorial

TRENDING TECHNOLOGIES

Advertisement

Cloud Computing Tutorial

Amazon Web Services Tutorial

Microsoft Azure Tutorial

Git Tutorial

Ethical Hacking Tutorial

Docker Tutorial

Kubernetes Tutorial

DSA Tutorial

Spring Boot Tutorial

SDLC Tutorial

Unix Tutorial

CERTIFICATIONS

Business Analytics Certification

Java & Spring Boot Advanced Certification

Data Science Advanced Certification

Cloud Computing And DevOps

Advanced Certification In Business Analytics

Artificial Intelligence And Machine Learning

DevOps Certification

Game Development Certification

Front-End Developer Certification

AWS Certification Training

Python Programming Certification

COMPILERS & EDITORS

Online Java Compiler

Online Python Compiler

Online Go Compiler

Online C Compiler

Online C++ Compiler

Online C# Compiler

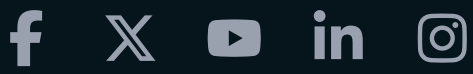
Online PHP Compiler

Online MATLAB Compiler

Online Bash Terminal

Advertisement

Online Html Editor



Tutorials Point is a leading Ed Tech company striving to provide the best learning material on technical and non-technical subjects.

© Copyright 2025. All Rights Reserved.

Advertisement