machine learning terminology

divided into categories. I have grouped the terms based on their relevance and similarity. I hope you find this list useful and informative.

**Machine Learning Basics**

* Machine learning: A branch of artificial intelligence that enables systems to learn from data and improve their performance without explicit programming.
* Data: The raw information or facts that are used as input or output for machine learning algorithms.
* Label: The value or category that we want to predict or explain using machine learning models.
* Feature: An attribute or characteristic of the data that is used as input for machine learning models.
* Model: A mathematical representation of the relationship between the features and the label, or between the features themselves.
* Training: The process of adjusting the parameters of a model to fit the data and minimize the error.
* Inference: The process of using a trained model to make predictions or decisions based on new or unseen data.
* Evaluation: The process of measuring the performance and accuracy of a model using metrics and tests.
* Validation: The process of checking whether a model meets the requirements and expectations of the problem domain and the stakeholders.
* Deployment: The process of making a model available for use in a production environment or a real-world scenario.

**Machine Learning Types**

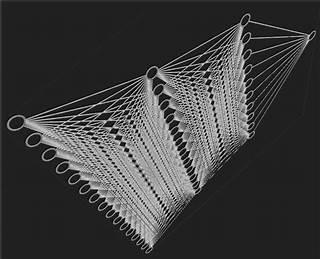
* Supervised learning: A type of machine learning that uses labeled data (data with known answers) to train algorithms to classify data or predict outcomes.
* Unsupervised learning: A type of machine learning that uses unlabeled data (data without known answers) to discover patterns, structures, or hidden information in the data.
* Semi-supervised learning: A type of machine learning that uses a combination of labeled and unlabeled data to train algorithms that can benefit from both types of data.
* Reinforcement learning: A type of machine learning that uses feedback from the environment to train algorithms to learn optimal actions or policies for a given situation.
* Self-supervised learning: A type of machine learning that uses self-generated labels or objectives to train algorithms to learn from unlabeled data.

**Machine Learning Tasks**

* Classification: A machine learning task that involves assigning a label or category to an input based on predefined classes or criteria.
* Regression: A machine learning task that involves predicting a continuous value or outcome for an input based on numerical features or variables.
* Clustering: A machine learning task that involves grouping similar inputs into clusters or segments based on their features or characteristics.
* Dimensionality reduction: A machine learning task that involves reducing the number of features or variables in the data by preserving the most relevant or important information.
* Anomaly detection: A machine learning task that involves identifying abnormal or unusual inputs that deviate from the normal or expected pattern in the data.
* Association rule mining: A machine learning task that involves finding rules or patterns that describe how different items or variables are related or co-occur in the data.
* Recommendation system: A machine learning task that involves suggesting items or products that are likely to be of interest or preference to a user based on their past behavior or profile.

**Machine Learning Algorithms**

* Linear regression: A supervised learning algorithm that models the relationship between one or more features and a continuous label using a linear function.
* Logistic regression: A supervised learning algorithm that models the probability of an input belonging to a binary class using a logistic function.
* Decision tree: A supervised learning algorithm that builds a tree-like structure where each node represents a feature, each branch represents a decision, and each leaf represents a label.



Explore

Here are some more terminologies related to machine learning that you might find useful:

* Bias: The difference between the expected value and the true value of a target variable. Bias can also refer to the tendency of a model to learn incorrect patterns or assumptions from the data.
* Variance: The variability or spread of the predicted values around the expected value. Variance can also refer to the sensitivity of a model to small changes or noise in the data.
* Underfitting: A problem that occurs when a model is too simple or has too few parameters to capture the complexity or variability of the data. Underfitting results in high bias and low variance, and poor performance on both training and test data.
* Overfitting: A problem that occurs when a model is too complex or has too many parameters to fit the data too well. Overfitting results in low bias and high variance, and poor generalization to new or unseen data.
* Regularization: A technique that can help to prevent overfitting by adding a penalty term to the cost function that depends on the values or complexity of the model parameters. Regularization can help to reduce the variance and noise of the model by shrinking or pruning the parameters.
* Cross-validation: A technique that can help to evaluate and compare the performance of different models or hyperparameters by splitting the data into multiple subsets and using some subsets for training and some subsets for testing. Cross-validation can help to reduce the overfitting and bias of the model by using different combinations of data for training and testing.
* Hyperparameter: A parameter that controls the behavior or configuration of a machine learning algorithm, such as the learning rate, the number of iterations, or the number of hidden layers. Hyperparameters are not learned from the data, but are set before training.
* Optimization: The process of finding the optimal values or settings for the model parameters or hyperparameters that minimize the cost function or maximize the performance metric. Optimization can involve different methods or algorithms, such as gradient descent, stochastic gradient descent, or genetic algorithms.
* Gradient descent: An optimization algorithm that iteratively updates the model parameters by moving in the opposite direction of the gradient (or slope) of the cost function with respect to the parameters. Gradient descent can help to find the local minimum or global minimum of the cost function, depending on the initial values or settings of the parameters.
* Stochastic gradient descent: A variant of gradient descent that updates the model parameters by using only a single or a small batch of randomly selected examples at each iteration. Stochastic gradient descent can help to speed up the convergence and avoid local minima by introducing randomness and noise in the gradient calculation.

If you want to learn more about these terminologies, you can check out these web search results: [Machine Learning Glossary | Google for Developers](https://developers.google.com/machine-learning/glossary/), [Machine Learning Terminology - W3Schools](https://www.w3schools.com/ai/ai_ml_terminology.asp), [Machine Learning Glossary (A to Z Terms) in ML - EnjoyAlgorithms](https://www.enjoyalgorithms.com/blogs/machine-learning-glossary/).