

ADVANCED MACHINE LEARNING ALGORITHMS

Performance Evaluation

Definition

Performance evaluation is the process of assessing the effectiveness and accuracy of a machine learning model on a given dataset.

Performance Evaluation Methods

- Evaluation Metrics:

Evaluation metrics are ways to measure how well a machine learning model is performing. They help us understand if the model is making accurate predictions or if improvements are needed.

- Cross-Validation:

A technique to test how well a model will work on new data by dividing the data into several parts. The model is trained on some parts and tested on the remaining parts, and this is repeated several times.

- Hyperparameter Tuning:

Hyperparameter Tuning is the process of selecting the best set of hyperparameters for a machine learning model. They control how the model is trained and can significantly affect its performance.

- Overfitting and Underfitting:

Overfitting occurs when a model learns the training data too well, capturing noise and details that do not generalize to new data.

Underfitting occurs when a model is too simple to capture the underlying patterns in the data.

Linearity-based Models

Objective

Linearity-based models assume that the relationship between input variables (features) and the output (target) can be represented as a linear equation.

Algorithms

- Linear Regression
- Logistic Regression
- Ridge and Lasso Regression
- Support Vector Machines or SVM with Linear Kernels

Advantages

- Simple and fast to train.
- Easy to interpret.
- Works well for problems where the relationship between variables is approximately linear.

Limitations

- May not perform well on complex, non-linear problems.

Distance-based Models

Objective

These algorithms rely on measuring the "distance" between points in feature space to classify or predict new instances.

Algorithms

- K-Nearest Neighbors, KNN
- K-Means§
- DBSCAN

Advantages

- Simple and effective for smaller datasets.
- KNN can be very accurate for datasets with well-defined class boundaries.
- K-Means works well with spherical clusters.

Limitations

- KNN can be slow on large datasets due to the need to compute distances to all points.
- K-Means requires a predefined number of clusters and struggles with clusters of different shapes and sizes.
- DBSCAN may not perform well with clusters of varying densities.

Probabilistic-based Models

Objective

These models aim to predict the likelihood of different outcomes. They are especially useful when data is noisy or uncertain and can be applied to both classification and regression problems.

Algorithms

- >Naïve Bayes

- Bayesian Networks
- Hidden Markov Models, HMM

Advantages

- Simple and computationally efficient.
- Can handle uncertainty and incomplete data.
- Probabilistic interpretations make them interpretable.

Limitations

- The "naïvety" assumption (features are independent) can be a limitation if features are not independent.
- Requires more data to avoid overfitting in complex problems.

Tree-based Models

These models use a tree-like structure to represent decisions and their possible consequences. They split the data into branches based on feature values, with the goal of making more accurate predictions with each split.

Algorithms:

- Decision Tree
- Random Forest
- Gradient Boosting Machines, GBM
- AdaBoost

Advantages

- Handles both numerical and categorical data well.

- Effective for both regression and classification tasks.
- Models are interpretable, especially decision trees.

Limitations

- Prone to overfitting, especially with deep trees.
- Can be computationally expensive for large datasets.

Naïve Bayes

Definition

Naïve Bayes is a classification algorithm based on Bayes' theorem, which assumes that features are conditionally independent given the class.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Remark

Naïve Bayes assumes that the presence or absence of features is independent of each other given the class. That is why it is called Naïve.

Likelihood in Naïve Bayes

In Naïve Bayes algorithm, "likelihood" refers to the probability of observing a particular feature value given a class label. It quantifies how likely it is to see a specific feature value when we know the class of the data point.

Decision Tree

Decision Tree is a Supervised learning algorithm that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes

represent the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome.

Definition

A Decision Tree is a flowchart-like structure where each internal node represents a test on a feature, each branch represents the outcome of the test, and each leaf node represents a class label, decision taken after computing all features.

Vocabulary

- Root Node:

The initial node at the beginning of a decision tree, where the entire population or dataset starts dividing based on various features or conditions.

- Splitting:

Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.

- Decision Nodes:

Nodes resulting from the splitting of root nodes are known as decision nodes. These nodes represent intermediate decisions or conditions within the tree.

- Leaf Nodes:

Nodes where further splitting is not possible, often indicating the final classification or outcome. Leaf nodes are also referred to as terminal nodes.

- Branch or Sub-Tree:

A subsection of the entire decision tree is referred to as a branch or sub-tree. It represents a specific path of decisions and outcomes within the tree.

Decision Tree Algorithm

The algorithm to build a decision tree is as follows:

1. Start with the root node and all training data.
2. For each node, select the best feature to split on based on a criterion.
3. Split the data into subsets based on the selected feature.
4. Create child nodes for each subset and repeat the process recursively until a stopping condition is met.

Attribute Selection Measure, ASM

Definition

Attribute Selection Measure is a method used to determine the best feature to split on at each node in a decision tree. It evaluates the quality of a split based on how well it separates the data into distinct classes.

Information Gain

Definition

Information gain is a measure used to determine which feature should be used to split the data at each internal node of the decision tree. It is calculated using entropy.

Formula

$$\text{IG}(T, A) = \text{Entropy}(T) - \sum_{v \in \text{Values}(A)} \frac{|T_v|}{|T|} \cdot \text{Entropy}(T_v)$$

where T is the set of training instances, A is the attribute being evaluated, T_v is the subset of T where attribute A has value v , and $\text{Values}(A)$ is the set of all possible values for attribute A .

Entropy

Definition

Entropy is a metric to measure the impurity in a given attribute. It specifies randomness in data. In a decision tree, the goal is to decrease the entropy of the dataset by creating more pure subsets of data. Since entropy is a measure of impurity, by decreasing the entropy, we are increasing the purity of the data.

$$\text{Entropy}(S) = - \sum_{i=1}^c p_i \cdot \log_2(p_i)$$

where p_i is the proportion of instances in class i and c is the total number of classes.

Gini Impurity

Definition

Gini impurity is a measure of how often a randomly chosen element from the dataset would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the dataset.

$$\text{Gini}(S) = 1 - \sum_{i=1}^c p_i^2$$

where p_i is the proportion of instances in class i and c is the total number of classes.

Random Forest

Definition

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes, for classification, or mean prediction, for regression, of the individual trees.

Random Forest Algorithm

The algorithm to build a random forest is as follows:

1. From the training dataset, create multiple bootstrap samples.
2. For each bootstrap sample, train a decision tree using a random subset of features at each split.
3. Repeat the process to create a large number of trees (the forest).
4. For classification, aggregate the predictions of all trees by majority vote; for regression, average the predictions.

Neural Networks

Definition

Neural networks are a class of machine learning models inspired by the structure and function of the human brain. They consist of interconnected nodes, neurons, organized in layers, where each connection has a weight that is adjusted during training to minimize prediction errors.

Neural Networks Techniques

- Feedforward Neural Networks
- Convolutional Neural Networks, CNN
- Recurrent Neural Networks, RNN
- Long Short-Term Memory Networks, LSTM

Artificial Neural Networks, ANN

Artificial Neural Networks are computational models inspired by the human brain's neural networks. They consist of layers of nodes, called neurons, which are interconnected and work together to process information and make predictions or decisions based on input data.

Artificial Neural Networks Architecture

The architecture of an Artificial Neural Network typically consists of three main types of layers: input layer, hidden layers, and output layer. Each layer is made up of neurons that are connected to neurons in the adjacent layers through weighted connections.

Vocabulary

- Weights:

Each connection between neurons has a weight, indicating how strong the connection is. If the weight is close to zero, changing the input won't affect the output much. Negative weights mean that increasing the input will decrease the output.

- Activation Function:

Each neuron in the hidden and output layers applies a function to the sum of its inputs multiplied by their weights. This function is called an activation function. Choosing the right activation function is important and affects how the network learns.

- Deep Network:

A neural network with two or more hidden layers is called a deep neural network.

Deep Learning

Definition

Deep Learning is a subset of machine learning that focuses on using neural networks with many layers, deep neural networks, to model complex patterns in data. It has been particularly successful in areas such as image and speech recognition, natural language processing, and game playing.

Feature Engineering

Definition

Feature engineering is the process of using domain knowledge to extract or create features from raw data that can improve the performance of machine learning models.

Feature Engineering Techniques

- Handling Missing Data

the values or data that is not stored or not present, for some variables in the given dataset

- Outlier Detection

Outliers are data points that differ significantly from other observations. They can occur due to variability in the data or measurement errors.

- Scaling and Normalization

Scaling and normalization are techniques used to adjust the range of numerical features in a dataset. This ensures that all features contribute equally to the model's learning process.

- Encoding Categorical Data

Encoding categorical data involves converting categorical variables into numerical representations that can be used by machine learning algorithms.

Scaling

Definition

Scaling is the process of transforming features to a specific range, often $[0, 1]$ or $[-1, 1]$. This is important because many machine learning algorithms are sensitive to the scale of the input data.

Normalisation

Definition

Normalisation is the process of transforming features to a specific range, often $[0, 1]$ or $[-1, 1]$. This is important because many machine learning algorithms are sensitive to the scale of the input data.

Z-score Normalization, Standardization

Definition

Z-score normalization transforms features to have a mean of 0 and a standard deviation of 1. This is achieved by subtracting the mean and dividing by the standard deviation for each feature.

Formula

$$z = \frac{(x - \mu)}{\sigma}$$

Min-Max scaling

Definition

Min-Max scaling transforms features to a fixed range, typically $[0, 1]$. It is calculated by subtracting the minimum value and dividing by the range for each feature.

Formula

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Hyperparameters

Definition

Hyperparameters are parameters that are set before training a machine learning model. They control the behavior of the learning algorithm and can significantly impact the model's performance.

Learning Rate

Definition

The learning rate is a hyperparameter that controls how much to change the model in response to the estimated error each time the model weights are

updated.

Regularisation

Definition

Regularisation is a technique used to prevent overfitting in machine learning models. It adds a penalty term to the loss function, which discourages the model from learning overly complex patterns.

Hyperparameter Tuning Techniques

- **Manual Search:**

Selecting hyperparameters based on intuition or experience, typically used for simple models or when hyperparameter space is small.

- **Grid Search:**

Exhaustively tests all possible combinations of hyperparameters within specified ranges.

- **Random Search:**

Randomly selects combinations of hyperparameters and tests them.

- **Cross-Validation:**

A technique where the dataset is split into multiple parts or folds. The model is trained on some parts and tested on the remaining part, and this is repeated multiple times.

- **K-Fold Cross-Validation:**

The dataset is divided into K subsets or folds. The model is trained on K-1 folds and tested on the remaining fold. This process is repeated K times,

each time using a different fold as the test set.

Definition

Hyperparameter tuning techniques are methods used to find the optimal values for hyperparameters in machine learning models. These techniques help improve model performance by systematically adjusting the model's parameters.