Artificial Intelligence Assignment

SARTHAK KULSHRESTHA
Roll Number-21111051

06/August/2023

1 Q1. Explain the different types of Machine learning and also explain the five best algorithms of each type??

Machine learning can be broadly categorized into three main types:
1)Supervised Learning
2) Unsupervised Learning
3) Reinforcement Learning.

Supervised Learning: In supervised learning, the algorithm is trained on a labeled dataset, where the input data is accompanied by corresponding output labels. The goal is to learn a mapping function from input to output, so the algorithm can make predictions on new, unseen data.

Five Best Algorithms for Supervised Learning:
a. Linear Regression: A simple algorithm for regression tasks that fits a linear relationship between input features and the target variable.
b. Logistic Regression: Used for binary classification tasks, logistic regression models the probability of a binary outcome.
c. Decision Trees: Tree-like models used for both classification and regression tasks, where the data is split into branches based on feature values.
d. Random Forest: An ensemble method that combines multiple decision trees to improve accuracy and reduce overfitting.
e. Support Vector Machines (SVM): Effective for both binary and multi-class classification, SVM tries to find the best hyperplane that separates different classes.

Unsupervised Learning: In unsupervised learning, the algorithm is trained on an unlabeled dataset, and the goal is to find patterns and structures within the data without any explicit output labels.

Five Best Algorithms for Unsupervised Learning:

a. K-Means Clustering: Divides the data into K clusters based on similarity, with each cluster represented by its centroid.

b. Hierarchical Clustering: Builds a tree-like structure of clusters, allowing for hierarchical representation of the data.

c. Principal Component Analysis (PCA): Reduces the dimensionality of the data while retaining the most important features.

d. Gaussian Mixture Model (GMM): A probabilistic model that assumes data is generated from a mixture of several Gaussian distributions.

e. t-Distributed Stochastic Neighbor Embedding (t-SNE): Used for visualization, t-SNE reduces high-dimensional data to two or three dimensions, preserving local structure.

Reinforcement Learning:

In reinforcement learning, an agent learns to interact with an environment to achieve a goal. The agent receives feedback in the form of rewards or penalties based on its actions.

Five Best Algorithms for Reinforcement Learning:

a. Q-Learning: A model-free, off-policy algorithm used for discrete state and action spaces.

b. Deep Q Networks : Combines Q-learning with deep neural networks to handle high-dimensional state spaces.

c. Policy Gradient Methods: Directly optimize the policy to maximize the expected rewards.

d. Proximal Policy Optimization : A popular policy gradient algorithm that maintains a "proximal" policy update for stability.

e. Deep Deterministic Policy Gradient : Suitable for continuous action spaces, combines actor-critic methods with deep learning.

2 Explain Bagging and boosting Ensemble Learning with an example?

Bagging :

Bagging is an ensemble learning technique that aims to enhance the performance and robustness of machine learning models. It involves creating multiple subsets of the training data through random sampling with replacement. Each subset is used to train a separate base model (learner) independently. The final prediction is obtained by aggregating the predictions from all base models, typically using majority voting for classification problems or averaging for regression problems. By training models on diverse subsets of the data, Bagging reduces overfitting and improves the overall accuracy and stability of the predictions, making it particularly useful for dealing with noisy or complex datasets.

Boosting:

Boosting is another powerful ensemble learning method that sequentially combines multiple weak learners to create a strong learner. Weak learners are models that perform slightly better than random guessing. The boosting process assigns higher weights to misclassified examples, giving more attention to difficult instances during training. In each iteration, a new weak learner is trained on the updated dataset with the higher weights. The final prediction is obtained by aggregating the weighted outputs of all weak learners. Boosting is effective in handling complex decision boundaries and can significantly improve prediction accuracy by focusing on difficult examples. Unlike Bagging, Boosting adapts the model to give more importance to the most challenging instances, leading to better performance on challenging datasets.