BAZE UNIVERSITY ABUJA IN PARTNERSHIP WITH DOMINEUM

IDEAS Emerging Technology Skills Scholarship Program

INTRODUCTION TO MACHINE LEARNING (ML)

Presented by: Khadijah Saad Mohammed

SUPPORTED BY THE WORLD BANK, NITDA, Innovation Development and Effectiveness in the Acquition of Skills

Content

- Introduction
- Types of Machine Learning
- Supervised Learning
- Common Algorithms
- Evaluating Model Performance
- Preparing Data and Engineering Features

What is Machine Learning

Machine learning is a field of study that enables computers to learn from data and make predictions or decisions without being explicitly programmed.

- Machine learning is a subset of artificial intelligence.
- ML generates rules for us
- Before (classical programming) we give data and rules to get answers
- Now (Machine learning programming) We give it data and answers and it gives us rules.

TYPES OF ML

Supervised Learning

In supervised learning, the algorithm learns from labeled data, making predictions or decisions based on that data.

Unsupervised Learning

Unsupervised learning deals with unlabeled data, where the algorithm tries to find patterns or structure without specific guidance.

Reinforcement Learning

Reinforcement learning involves training an algorithm to make sequential decisions by rewarding or penalizing its actions.

Getting Started with Supervised Learning

Supervised learning involves training a model on a dataset with labeled examples, where each example consists of input features and a corresponding output.

Examples

- Imagine you're trying to predict whether a student will pass or fail an exam based on their study hours and previous exam scores. In this scenario, the input features are study hours and previous exam scores, and the output is whether the student passed or failed.
- Predicting the price of a house

COMMON ALGORITHMS

Linear Regression

- Linear regression is a straightforward algorithm used for predicting continuous values. It works by fitting a straight line to the data, aiming to minimize the difference between the observed and predicted values using techniques like least squares.

Logistic Regression

- Logistic regression is a widely used algorithm for binary classification tasks. It estimates the probability that a given input belongs to a particular class using a logistic function. It's particularly useful when the relationship between the features and the target variable is linear.

Decision Trees

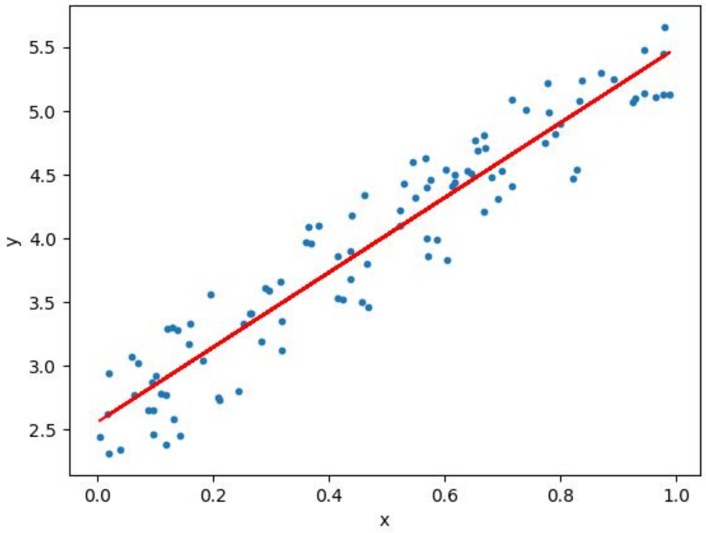
- Decision trees are versatile models capable of handling both classification and regression tasks. They work by recursively splitting the data into subsets based on the features' values, aiming to maximize the purity or homogeneity of each subset. Decision trees are easy to interpret and can capture complex relationships in the data.

K-Nearest Neighbors (KNN)

- KNN is a non-parametric algorithm used for both classification and regression tasks. It predicts the target value of a new instance by averaging or voting among the k nearest neighbors in the feature space. KNN is simple to implement and can capture complex decision boundaries, making it suitable for various applications. However, it can be computationally expensive, especially with large datasets.

LINEAR REGRESSION

The goal of linear regression is to fit a line to a set of points.



Importance of Model Evaluation

Evaluating models is crucial to understand how well they perform on unseen data.

Key Metrics

Accuracy: Measures the proportion of correctly classified instances.

Precision: Measures the proportion of true positives among all positive predictions.

Recall: Measures the proportion of true positives that were correctly identified.

F1-Score: Harmonic mean of precision and recall.

ROC Curve and AUC: Measures the tradeoff between true positive rate and false positive rate.

Machine Learning - Data-Driven

1. Collect data of images and their labels/classes
2. Use machine learning to train this to get a classifier
3. Test this classifier on a different set of data (Excluding the ones used for training)

Preparing Data and Engineering Features

Data Pre-processing

- Cleaning: Removing errors and inconsistencies from the data.
- Transformation: Scaling or normalizing features to ensure uniformity.
- Encoding: Converting categorical variables into numerical representations.
- Feature Engineering: Creating new features or transforming existing ones to improve model performance.

Avoiding Common Problems

Bias-Variance Tradeoff:

- Balancing bias and variance to achieve optimal model performance.

Overfitting/Underfitting:

- Overfitting occurs when the model learns the training data too well but fails to generalize to new data.
- Underfitting happens when the model is too simple to capture the underlying patterns in the data.

Strategies to Address Overfitting/Underfitting:

- Cross-validation, regularization, and using more training data.

CONCLUSION

- Recap of Week 1:
  - Covered the basics of machine learning, including types, algorithms, evaluation metrics, data pre-processing, and common pitfalls.
  - Understanding these fundamentals is crucial for building a strong foundation in machine learning.
- Preview of Coming Weeks:
  - Stay tuned for more advanced topics and practical applications in subsequent sessions.

1.

BAZE UNIVERSITY ABUJA
IN PARTNERSHIP WITH
DOMINEUM

THANK YOU

Q&A