28/04/2025

Training Day-51

Data Transformation in Machine Learning:-

Data transformation is the most important step in a machine learning pipeline which includes modifying the raw data and converting it into a better format sothat it can be more suitable for analysis and model training purposes. In data transformation, we usually deal with issues such as noise, missing values, outliers, and non-normality.

Different Data Transformation Technique:-

Data transformation in machine learning involves a lot of techniques, let's discuss8 of the major techniques that we can apply to data to better fit our model and produce better results in the prediction process.
The choice of data transformation technique depends on the characteristics of thedata and the machine learning algorithm that we intend to use on the data. Here are the mentioned techniques discussed in details.

Handling Missing Data:-

Most of the times the data that is received from different sources miss some of the values in it, if we train our model on this data the model might behave differently or even produce error while training. Therefore, handling missing databecomes an important aspect to consider while transforming the data, there are different techniques through which we can handle the missing data which can help us improve our model performance. Let's discuss some of the techniques in details here:

- **Removing the Missing Data**: We can delete the rows or columns whichare having missing data. This is significant only when a small number
  of data is missing, if there's a large value of missing data points in our dataset we must consider some other technique otherwise deleting the rows or columns with large number of missing values will change the way our model performs since it might cause the model to train on less data. We can drop the rows which contain the missing values using the dropna method in pandas if the data we have is stored in a pandas dataframe.

- **Imputation:** In this technique we remove the missing values by filling the missing values positions with some other value, for example we canfill in the missing values with the mean of the different values from the same column which is in the same category or data type as of the

missing value. The most common types of imputation methods includemean, median, mode imputation. We can also fill in the missing valueswith a constant value that we want to be present in the data instead of the missing value. Imputation is also implanted within the sklearn library, we can impute different missing values with the help of KNNImputer(K-Nearest Neighbors) which is a part of sklearn.impute.

- **Forward Fill or Backward Fill**: Usually in time series analysis, wherethe data is produced after a constant time, if some data goes missing wecan replace the missing value with forward fill or the backward fill options, The forward fill method fills the missing value with the previous non missing value whereas backward fill method fills the missing value with the next non missing value to the missing value.

- **Interpolation**: Missing data can also be handled by interpolation technique, it involves predicting the missing values based on observed values in the dataset. There are multiple interpolation methods, the choice of the method is based on the data that we have. Most commonlyused interpolation is linear interpolation which assumes there is a linear relationship between observed values and missing data points, this
method predicts the missing value by fitting a straight line between twoadjacent non-missing points.

29/04/2025

Training Day-52

Normalization and Standardization in Machine Learning:-

Normalization and standardization are two techniques used to transform data into acommon scale. Normalization is a technique used to scale numerical data in the range of 0 to 1. This technique is useful when the distribution of the data is not known or when the data is not normally distributed. On the other hand, standardization is a technique used to transform data into a standard normal distribution. This technique is useful when the distribution of the data is known andwhen the data is normally distributed. Both techniques have different applications, and choosing the right technique based on the data and the problem you're trying tosolve is important.

What is Normalization?

Normalization in machine learning is a data preprocessing technique used to change the value of the numerical column in the dataset to a common scale withoutdistorting the differences in the range of values or losing information.

In simple terms, Normalization refers to the process of transforming features in adataset to a specific range. This range can be different depending on the chosen normalization technique.

The two most common normalization techniques are Min-Max Scaling and Z-Score Normalization, which is also called Standardization.

Now, let's discuss Min-Max Scaling.Min-

Max Scaling

This method rescales the features the features to a fixed range, usually 0 to 1. Theformula for calculating the scaled value of a feature is:

Normalized Value = Value - Min/ Max - Min

where,

Value: Original Value of the feature

Min: Minimum value of the feature across all the data points.

Max: Maximum value of the feature across all the data points.

What is Standardization?

Standardization is a data preprocessing technique used in statistics and machine learning to transform the features of your dataset so that they have a mean of 0 anda standard deviation of 1. This process involves rescaling the distribution of valuesso that the mean of observed values is aligned to 0 and the standard deviation to 1.

Standardisation aims to adjust the scale of data without distorting differences in theranges of values or losing information.

Unlike other scaling techniques, standardization maintains all original data points'information (except for cases of constant columns).

It ensures that no single feature dominates the model's output due to its scale,leading to more balanced and interpretable models.

Formula of Standardization

Z = (x-mean)/standard deviation

29/04/2025

Training Day-53

Applications of Machine Learning:-

Machine learning is one of the most exciting technologies that one would have ever come across. As is evident from the name, it gives the computer that which makes it more similar to humans: The ability to learn. Machine learning is activelybeing used today, perhaps in many more places than one would expect. Some of the most common applications are:

Image Recognition

Image Recognition is one of the reasons behind the boom one could haveexperienced in the field of Deep Learning. The task which started from
classification between cats and dog images has now evolved up to the level of FaceRecognition and real-world use cases based on that like employee attendance tracking.

Also, image recognition has helped revolutionized the healthcare industry by employing smart systems in disease recognition and diagnosis methodologies.

Speech Recognition

Speech Recognition based smart systems like Alexa and Siri have certainly comeacross and used to communicate with them. In the backend, these systems are
based basically on Speech Recognition systems. These systems are designed suchthat they can convert voice instructions into text.

One more application of the Speech recognition that we can encounter in our day-to-day life is that of performing Google searches just by speaking to it.

Recommender Systems

As our world has digitalized more and more approximately every tech giants try toprovide customized services to its users. This application is possible just because of the recommender systems which can analyze a user's preferences and search history and based on that they can recommend content or services to them.

An example of these services is very common for example youtube. It recommendsnew videos and content based on the user's past search patterns. Netflix recommends movies and series based on the interest provided by users whensomeone creates an account for the very first time.

Fraud Detection

In today's world, most things have been digitalized varying from buying toothbrushes or making transactions of millions of dollars everything is accessibleand easy to use. But with this process of digitization cases of fraudulent
transactions and fraudulent activities have increased. Identifying them is not thateasy but machine learning systems are very efficient in these tasks.

Due to these applications only whenever the system detects red flags in a user's activity than a suitable notification be provided to the administrator so, that thesecases can be monitored properly for any spam or fraud activities.

Self Driving Cars

It would have been assumed that there is certainly some ghost who is driving a carif we ever saw a car being driven without a driver but all thanks to machine
learning and deep learning that in today's world, this is possible and not a story from some fictional book. Even though the algorithms and tech stack behind thesetechnologies are highly advanced but at the core it is machine learning which has made these applications possible.

The most common example of this use case is that of the Tesla cars which are well-tested and proven for autonomous driving.

Medical Diagnosis

If you are a machine learning practitioner or even if you are a student then youmust have heard about projects like breast cancer Classification, Parkinson's
Disease Classification, Pneumonia detection, and many more health-related tasks which are performed by machine learning models with more than 90% of accuracy.

Not even in the field of disease diagnosis in human beings but they work perfectly fine for plant disease-related tasks whether it is to predict the type of disease it is orto detect whether some disease is going to occur in the future.

30/04/2025

Training Day-54

Neural Networks: A Comprehensive Guide

Introduction

Neural networks are computing systems inspired by biological neural networks in human brains. They form the foundation of deep learning, a subset of machine learning that excels at pattern recognition and problem-solving.

Basic Components

1. Neurons (Nodes)
   - Basic processing units
   - Receive input, process it, and generate output
   - Each neuron has a weight and bias

2. Layers
   - Input Layer: Receives initial data
   - Hidden Layers: Process information
   - Output Layer: Produces final results

How Neural Networks Work

1. Input Processing

- Data enters through input layer
- Each input is multiplied by associated weights
- Bias is added to weighted sum

2. Activation Functions

Common types:

- ReLU (Rectified Linear Unit)
- Sigmoid

- Tanh
- Softmax (for classification)

3. Training Process
- Forward Propagation
- Backward Propagation
- Weight Adjustment
- Error Minimization

Applications

1. Computer Vision
- Image Recognition
- Object Detection
- Facial Recognition
- Medical Image Analysis

2. Natural Language Processing
- Language Translation
- Text Generation
- Sentiment Analysis
- Speech Recognition

3. Business Applications
- Customer Behavior Prediction
- Risk Assessment
- Market Analysis
- Fraud Detection

Types of Neural Networks

1. Feedforward Neural Networks
   - Simplest form
   - Information flows in one direction
   - Used for pattern recognition

2. Convolutional Neural Networks (CNN)
   - Specialized for image processing
   - Uses convolution operations
   - Excellent at feature detection

3. Recurrent Neural Networks (RNN)
   - Processes sequential data
   - Has memory capabilities
   - Used for time series analysis

Limitations and Challenges

- Requires large amounts of data
- Computationally intensive
- Black box nature
- Potential for bias

Future Prospects

- Improved efficiency
- Better interpretability
- Enhanced automation
- Broader application