Chapter 2

Date 09/04/2022.

Feature :

A Feature is an attribute of a data set that is used in a Machine Learning process.

There is a view amongst certain machine Learning practitioners that only those attributes which are meaningful to a machine learning problem are to be called as features.

The features in a data set are also called as its dimensions. Hence a data set having 'n' features is called an n-dimentional data set.

Example :

In a document, the word count, character count & the number of pages could be the features. Hence, it is called as 3 dimentional data set.

Feature Engineering:

Feature Engineering is the process of selecting, transforming & creating a new features from the data set or raw data to improve the performance of machine learning models. It involves extracting meaningful information from the data & representing it in a way that is more suitable for the learning algorithm.

More, precisely, Feature Engineering refers to the process of translating data set into features such that these features are able to represent the data set more efficiently & results in a better learning performance.

It has 2 major Elements:

1. Feature Transformation
2. Feature sub-set Selection.

1 Feature Transformation:

It transforms the data either Structured or Unstructured into a new set of features which can represent the underlying problem which machine learning is trying to solve. This involves transforming the selected features to make them more suitable for the learning

2 Types:

1) Feature Construction:

This involves creating new features by combining or transforming the existing feature.

If their are 'n' features or dimentions in a data set after feature construction 'm' more featus added.

2) Feature extraction:

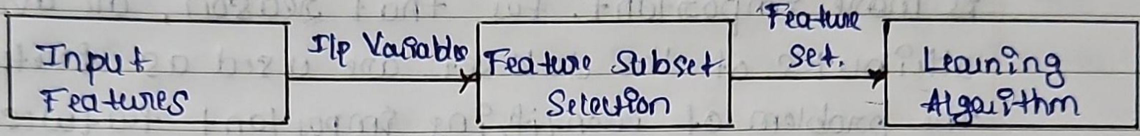
It is a process of extracting or creating new set of features from the original set of features using some functional mapping.

2. Feature Sub-Set Selection:

Also called as Feature Selection is the process of generating new Features.

The objective is to derive a subset of features from the full feature set which is most meaningful for a specific ML problem.

It involves selecting smaller subset of features that can provide similar or better performance compared to using all the available features.



classmate

Date

Page

7 Explain the Concept of Feature Transformation?

Feature Transformation:

Feature Transformation is the process of transforming the original features of a data set into a new set of features that are more suitable for ML algorithms. It involves applying mathematical or statistical operations to the original features to create new features that capture the underlying patterns or relationships in the data.

However, often it is not clear which feature is more important. For that reason, all available attributes of the data set are used as features & the problem of identifying important features is left to the learning model. This is definitely not a feasible approach.

For example:

Consider a data set that contains info. about the height & weight of individuals. The ML algorithm used to predict a person's BMI may perform better if the features are transformed.

To deal with these problems Feature Transformation come into play & It is used as an effective tool for dimensionality reduction & hence for boosting the learning model performance.

2 Goals of Feature Transformation:

1. Achieving best reconstruction of the original features in the data set.
2. Achieving highest efficiency in Learning task.

One common Feature Transformation technique is Normalization. And also Log Transformation, Principal Component Analysis (PCA) & so on

2. Disscuss the Concept of text Specific Feature Construction with Example

~~Text Specific~~ Feature Construction :

Feature Construction is the process of creating new features from the original features in a data set, which is more powerfull in order improve the performance of a ML algorithm.

Let's Consider a example of a real estate data set having details of all apartments sold in a specific region. the data set has 3 Features - apartment length, apartment bredth & price. If it is used as an input to a regression problem, such data can be the training data for the regression model. So given the training data the model should be able to predict the price of an apartment whose price is not known. However instead of using length & bredth of the apartment as a predicator, it is much convenient & makes more sense to use the area of the apartment, which is not an existing feature of the data set. Hence such a feature, namely apartment area, can be added to the data set. In other words we transform the 3-Dimentional data set to 4-Dimentional data set.

Text-Specific Feature Construction:

In the current world, text is arguably the most predominant medium of communication. Whether we think about Social Networks like, Facebook, Twitter or email or whatsApp, text plays a major role in the flow of information.

Text data, due to the inherent unstructured nature of the data it is so straight-forward that it do not have readily available features like Structured data sets on which the ML tasks can be executed.

All Machine Learning Models need numerical data as input. So the text data in the data sets need to be transformed into numerical features.

Text Data or Corpus which is the more popular keyword, is converted to numerical representation following a process known as Vectorization.

In this process, word occurences in all documents belonging to the corpus are consolidated in the form of bag-of-words

There are 3 major steps that are followed:

1. Tokenize
2. Count
3. Normalize.

In order to Tokenize a corpus, the Blankspace & punctuations are used as delimeters to separate the words, or tokens.

Then the number of occurances of each token is Counted, for each document.

Lastly tokens are weighted with reducing importance when they occur in the majority of the documents. A matrix is then formed with each token corresponding to Column & a document of corpus representing as Row. Each cell contains the count of tokens. This mat is called Document-term matrix.

Ex: Consider the 3 sample Doc

Doc 1 : Intelligent app Creates Intelligent process

Doc 2 : Bots are intelligent app

Doc 3 : Bussiness Inteligente.

	Intelliget	app	Create	process	Bots	...	Business
Doc1	2	1	1	1	0		0
Doc2	1	1	0	0	1		0
Doc3	1	0	0	0	0		1

classmate

Date
Page

6 Explain the Key Drivers in Feature Selection?

Key Drivers in Feature Selection:

Feature Selection is arguably the most critical pre-processing activity in any ML project. It intends to select a sub-set of system attributes or features which makes a most meaningful contribution in ML activity. It is also called as Feature Sub-Set Selection.

Key Drivers:

1 Feature Relevance

2 Feature Redundancy.

1 Feature Relevance:

It refers to which an input feature contributes to the output of a machine learning model. In other words, it is a measure of how important a feature is in determining the final outcome of the model.

It is typically measured using statistical or ML technologies. These techniques will help to identify the most important features & remove irrelevant or redundent features.

In supervised learning in case a variable is not contributing any information then it is said to be irrelevant. In case th info contribution for prediction is very little, then the variable is said to be weakly relevant. Remaining variables which makes significant contribution said to be strongly relevant.

Ex:

Roll number of a student does not contribute any significant info in predicting the weight ~~of~~ the student.

2 Feature Redundancy:

A Feature may contribute information which is similar to the info contributed by one or more other features.

So it refers to the situation where 2 or more input features in a ML model contains similar or redundent information.

For example:

In the weight prediction problem, both the features Age & Height contribute similar information. This is because with an increase in Age, & Height weight is expected to increase.

So in context to this problem age & height contribute similar information.

But, Redundant Features can negatively impact the performance of the model by introducing noise, computing complexity & reducing the explainability of the model.

Theory or Theorem.

Bayes Theorem :

Bayes Theorm is a statistical formula that desribes the relationship b/w conditional probablities. It is named after Thomas Bayes.

In ML, Bayes Theorem is used to calculate the probablity of a Hypothesis or an event based on prior knowledge or evidence.

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

It states that the probablity of an event A given that event B has already occured is equal to the probablity of an event B given that event A has occured, multiplied by the probablity of event A, divided by probablity of event B.

3. Explain the Various terms associated with Bayes Theorm:

Terms associated with Bayes Theorm:

1 Prior
2 Posterior
3 Likelihood

1. Prior:

Also called as Prior Probablity is the Prior knowledge or belief about the probablities of various Hypothesis H is called Prior.

It is based on our prior knowledge, assumptions, or beliefs about the probablity of the event.

For example:

If we have to determine whether a perticular type of tumour is maligont for a patient,

Let's say we wan't to predict whether a customer will buy a product or not based on their age, gender & income, before we predict we should have some prior knowledge.

2 Posterior:

The probablity that a perticular Hypothesis holds for a data set based on the Prior is called the Posterior or Posterior Probablity.

It is based on the Prior Probablity of the Hypothesis or model.

For example:

Suppose we have a bag containing 5 Red & 3 Blue Balls. We randomly pick one ball without looking, & we wan to know the probablity that the ball is Red

We can use Baye's Theorem to calculate the posterior probablity of the ball being red, given that we know the prior probablity of the ball being red & the likelyhood of obseving a red ball.

Lets Assume, Prior probablity of picking red ball is,
$P(Red) = 5/8 = 0.625$ & the Blue Ball is,
$P(Blue) = 3/8 = 0.375$.

Now let's say we observe that the ball we picked is not blue. So now we can find the Posterior probablity of the ball being red given the observed data:

$$P(Red \mid Not\ Blue) = \frac{P(Not\ Blue \mid Red) * P(Red)}{P(Not\ Blue)}$$

$$= \frac{0.625}{1 - 0.375} = 1.$$

So the posterior probablity of Ball being red is 1 or 100% given that we observed a non-blue ball.

3. Likelihood:

A likelihood refers to the probablity of observing the data or evidence given a certain Hypothesis. It measures how well the Hypothesis explains the observed data.

In mathematical notation, it is denoted as P(Data | Hypothesis) where data represents the observed data & Hypothesis represents the Hypothesis or model being tested.

likelihood

Prior Probablity

Posterior Probablity

$$P(hypothesis \mid data) = \frac{P(data \mid hypothesis) * P(hypothesis)}{P(data)}$$

Marginal Likelihood.

4. Explain the Concept of Naives Bayes Classifier?

Naives Bayes Classifier:

The Naive Bayes Classifier is a probablistic algorithm used for classification tasks in ML.

It calculates the probablity of a data point belonging to a certain class based on it's features using Bayes theorem.

The term "naive" indicates that the features are conditionally independent of each other in a given class.

It is not a single algorithm but a family of algorithms where all of them share a common principle i.e every pair of features being classified is independent of each other.

Feature Matrix:

Example	Sky	AirTemp	Wind	Water	EnjoySport
1	Sunny	Warm	Strong	Warm	YES
2.	Rainy	Cold	Strong	Warm	NO

The dataset is divided into 2 parts namely:

1) Feature Matrix:

Contains all the rows or vector of data set in which each vector consist of the value of dependent features. In the above data set features are, 'Sky', 'AirTemp', 'Wind', 'Water'.

2) Response Vector:

Contains the value of class variable i.e output or prediction for each row of feature matrix. In the above data set the class variable name is 'EnjoySport'.

Steps in Naive Bayes Classifier:

It calculates the probability of an event using in the following steps.

Step 1: Calculate the Prior Probablity for a given class labels.

Step 2: Find the Likelihood probablity with each attribute for each class.

Step 3: Put these values in Bayes Formula & Calculate Posterior Probablity.

Step 4: See which class has a higher probablity, given the input belongs to the higher probablity class.

Application of Naive Bayes Classifier:

Naive Bayes Classifier is a widely used algorithm in ML due to its simplicity, speed & accuracy in many applications.

1
2
3
4
5

1 Text Classification:

Naive Bayes classifier is widely used for text classification such as spam filtering, document classification & etc. It can quickly & accurately classify large amounts of text data, making it an ideal choice for such tasks.

2 Recommendation Systems:

It is used in recommendation systems to recommend products, services or content to users based on their past behaviours or preferences.

3 Fraud Detection:

It is used in Fraud Detection systems to identify suspicious transactions or activities. It can quickly & accurately identify patterns in large data sets & detect fraudulent behaviour.

4 Image Classification:

It is also used in image classification tasks such as facial recognition, object detection & image segmentation. It can classify image based on their features & accurately identify objects or people in the image.

5. Medical Diagnosis:

It is used to classify patients based on their symptoms & medical history. It can help to identify the likelihood of a patient having a certain diseases or condition which can help in diagnosis & treatment.

6 Spam Filtering:

Spam Filtering is the best known use of Naive Bayesian classification.

Presently, almost all the email providers have this as a built-in functionality, which makes use of a Naive Bayes classifier to identify spam email on the basis of certain conditions.