

Feature extraction in Data Mining

Data mining refers to extracting or mining knowledge from large amounts of data. In other words, Data mining is the science, art, and technology of discovering large and complex bodies of data in order to discover useful patterns. Theoreticians and practitioners are continually seeking improved [techniques](#) to make the process more efficient, cost-effective, and accurate.

"Data is the new oil for IT industry." – by Clive Humby

"The world is one big data problem." – by Andrew McAfee, co-director of the MIT Initiative

"Data is the new science. Big Data holds the answers." – By Pat Gelsinger

This is indeed true because there is an ample amount of data available to us that we can definitely make use of in one or the other fields once it has been processed. However, industries require only a subset of data from the whole lot. So in reality, we need some mechanisms to get access to the part of data that we actually need. However, this job can not be done manually as it would be time-consuming and at times would become out of scope for us.

Suppose if we want to build a machine learning project for some firm or for our own project requirements where we required images to make a project on object detection. Making such kinds of projects require an image dataset which might contain numerous attributes. So in order to work with them, we must first extract the features we needed. So in this case feature extraction plays a major role to make our life easy.

Feature Extraction

Feature Extraction is basically a process of dimensionality reduction where the raw data obtained is separated into related manageable groups. A distinctive feature of these large datasets is that they contain a large number of variables and additionally these variables require a lot of computing resources in order to process them. Hence Feature Extraction can be useful in this case in selecting particular variables and also combining some of the related variables which in a way would reduce the amount of data. The results obtained would be evaluated with the help of precision and recall measures. PCA is one of the most used linear dimensionality reduction techniques. It is an unsupervised learning algorithm.

Feature Generation

Feature Generation is the process of inventing new features from the already existed features. As the sizes of the datasets vary a lot, it becomes impossible to manage the larger ones. Thus this process of feature generation can play a vital role in order to ease the task. To avoid generating meaningless

features, we make use of some mathematical formulae and statistical models to enhance clarity and accuracy. This process usually adds more information to the model to make it more accurate. So enhancing model accuracy is something that can be achieved through this process. This process in a way ignores the meaningless interaction by detecting meaningful interactions.

Feature Evaluation

It is of utmost importance to initially prioritize the features to get our work done in a well-organized manner and thus feature evaluation can be a tool for this. Here each and every feature is being evaluated in order to score them objectively and henceforth utilize them based on the current needs. The unimportant ones can be ignored. So feature evaluation is an important task to perform in order to get a proper final output of the model by reducing the bias ness and inconsistency in the data.

Linear and Non-Linear Feature Extraction

Feature Extraction can be divided into two broad categories i.e. **linear** and **non-linear**. One of the examples of linear feature extraction is PCA (Principal Component Analysis). A principal component is a normalized linear combination of the original features in a dataset. PCA is basically a method to obtain required variables (important ones) from a large set of variables available in a data set. PCA tends to use orthogonal transformation to transform data into a lower-dimensional space which in turn maximizes the variance of the data.

PCA can be used for anomaly and outlier detection as these are considered as noise or irrelevant data in the entire dataset.

Steps followed in building PCA from scratch are:

- Firstly, standardize the data
- Thereafter, calculate the Covariance-matrix
- Then, calculate the Eigenvector & Eigenvalues for the Covariance-matrix.
- Arrange all Eigenvalues in decreasing order.
- Normalize the sorted Eigenvalues.
- Horizontally stack the Normalized_Eigenvalues

PCA fails when the data is non-linear which can be considered as one of the biggest disadvantages of PCA. This is where Kernel-PCA plays its role. Kernel-PCA is similar to SVM because both of them implements Kernel-Trick to convert the non-linear data to higher dimensional data up to the point when the data is separable. Non-Linear approaches could be used in the case of face recognition to extract features over large datasets.

Applications of Feature Extraction

- **Bag of Words:** This is the most widely used technique in the field of Natural Language Processing. Here, firstly sentences are tokenized, lemmatized, and stop words are removed. After that, the words are individually classified into the frequency of use. Since the features are usually extracted from a sentence present in a document or website, feature extraction plays a vital role in this case.
- **Image Processing:** Image processing is one of the most explorative domains where feature extraction is widely used. Since images represent different features or attributes such as shapes, hues, motion in the case of digital images thus processing them is of utmost importance so that only specified features are extracted. The image processing also makes use of many algorithms in addition to feature extraction.
- **Auto-encoders:** This is mainly used when we want to learn a compressed representation of raw data. The procedure carried out is basically unsupervised in nature. Thus, Feature Extraction plays a major role in identifying the key features from the data which will help us to code by learning from the coding of the original data set in order to derive new ones.
- Effective Feature Extraction also plays a major role in solving under-fitting and overfitting related problems in Machine Learning related projects.
- Feature Extraction also gives us a clear and improvised visualization of the data present in the dataset as only the important and required data has been extracted.
- Feature Extraction helps in training the model in a more efficient manner which in turn basically speeds up the whole process.

How is it Different from Feature Selection?

Feature Selection aims to rank the importance of the features previously existing in the dataset and in turn remove the less important features. However, Feature Extraction is concerned with reducing the dimensions of the dataset to make the dataset more crisp and clear.

PCA fails when the data is non-linear which can be considered as one of the biggest disadvantages of PCA. This is where Kernel-PCA plays its role. Kernel-PCA is similar to SVM because both of them implements Kernel-Trick to convert the non-linear data to a higher dimensional data upto the point when the data is separable. Non-Linear approaches could be used in case of face recognition to extract features over large datasets.

So Feature Extraction has its diverse usage in most of the domains. This appears at the beginning stage of any project which makes use of a large dataset. So the whole procedure of feature extraction must be executed and evaluated carefully to get an optimized result with greater accuracy which in turn will help us to get a better insight of the relationship between the variables present in the dataset and henceforth plan for the next stage of the execution.