**Application of Machine learning**

* Image recognition
* Speech recognition
* Traffic prediction
* Self- driving cars
* Email Spam and Malware Filtering
* Virtual Personal Assistant
* Online Fraud Detection

**Association Rules mining and “Apriori Principle”**

* Association Rule Mining, as the name suggest, association rules are simple if/Then statements that help discover relationship between seemingly independent relational database or other data repositories.
* Most machine learning algorithms works with numeric datasets and hence tend to be mathematical. However, association rule mining is suitable for non-numeric categorical data and requires just a little bit more than simple counting.
* Association Rule Mining in is an Unsupervised Non-linear algorithm to uncover how the items are associated with each other. In it, frequent Mining shows which items appear together in a transaction or relation. It is majorly used by retailers, grocery stores, an online marketplace that has a large transactional database. The same way when any online social media, marketplace, and e-commerce websites know what you buy next using recommendations engines. The recommendations you get on item or variable, while you check out the order is because of Association rule mining boarded on past customer data.
* There are three common ways to measure association:

1. Support
2. Confidence
3. Lift

**Support**

* it says how popular an item is, as measured in the proportion of transactions in which an item set appears **Support (A => B) = P (A U B)**

**Confidence**

* it says how likely item Y is.

**Application of Association Rule mining**

* Market basket Analysis
* Medical Diagnosis
* Census Data
* Protein Sequence

**Apriori Principle**

* Read each item in the transaction.
* Calculate the support of every item.
* If support is less then minimum support, discard the item. Else, insert into frequent itemset.
* Calculate confidence for each non-empty subset.
* If confidence is less than minimum confidence, discard the subset. Else, it into strong rules

**Strengths and the weaknesses of K-means clustering.**

**Strength**

* Simple – easy to understand and to implement.
* Efficient – Time complexity
* K -means is the most popular clustering algorithm.

**Weakness**

* The algorithm is only applicable if the mean is defined.
* For categorical data, k-mode – the centroid is represented by most frequent values.
* The user needs to specify K.
* The algorithm is sensitive to **outliers.**
* Outliers are data points that are very far away from other data points.

**Advantages and Disadvantages of Principal Component Analysis (PCA)**

**Advantages**

* Removes Correlated Features
  + In a real-world scenario, this is very common that you get thousands of features in your dataset. You cannot run your algorithm on all the features as it will be reducing the performance of your algorithm and it will not be easy to visualize that many features in any kind of graph. So, you must reduce the number of features in your dataset.
  + You need to find out the correlation among the features (Correlated variables). Finding correlation manually in thousands of features is nearly impossible, frustrating, and time-consuming. PCA does this for you efficiently.
  + After implementing the PCA on your dataset, all the principal components are independent of one another. There is no correlation among them.
* Improves Algorithm Performance
  + With so many features, the performance of your algorithm will drastically degrade. PCA is a very common way to speed up your Machine Learning algorithm by getting rid of correlated variables which do not contribute to any decision making. The training time of the algorithms reduces significantly with a smaller number of features.
  + So, if the input dimensions are too high, then using PCA to speed up the algorithm is a reasonable choice.
* Reduces Overfitting
  + Overfitting mainly occurs when there are too many variables in the dataset. So, PCA helps in overcoming the overfitting issue by reducing the number of features.
* Improves Visualization
  + It is very hard to visualize and understand the data in high dimensions. PCA transform a high dimensional data (2 dimension) so that it can be visualized easily.

**Disadvantages**

* Independent variables become less interpretable:
  + After implementing PCA on the dataset, your original features will turn into Principal Components. Principal Components are the linear combination of your original features. Principal Components are not as readable and interpretable as original features.
* Data standardization is must before PCA:
  + You must standardize your data before implementing PCA, otherwise PCA will not be able to find the optimal Principal Components.   
      
    For instance, if a feature set has data expressed in units of Kilograms, Light years, or Millions, the variance scale is huge in the training set. If PCA is applied on such a feature set, the resultant loadings for features with high variance will also be large. Hence, principal components will be biased towards features with high variance, leading to false results.  
      
    Also, for standardization, all the categorical features are required to be converted into numerical features before PCA can be applied.  
      
    PCA is affected by scale, so you need to scale the features in your data before applying PCA. Use **StandardScaler**from **Scikit Learn** to standardize the dataset features onto unit scale (mean = 0 and standard deviation = 1) which is a requirement for the optimal performance of many Machine Learning algorithms.
* Information Loss:
  + Although Principal Components try to cover maximum variance among the features in a dataset, if we do not select the number of Principal Components with care, it may miss some information as compared to the original list of features.

**Market Basket Analysis**

* Market Basket Analysis is one of the key techniques used by large retailers to uncover associations between items. It works by looking for combinations of items that occur together frequently in transactions. To put it another way, it allows retailers to identify relationships between the items that people buy.
* Association Rules are widely used to analyze retail basket or transaction data and are intended to identify strong rules discovered in transaction data using measures of interestingness, based on the concept of strong rules.

**An example of Association rules**

* Assume there are 100 customers.
* 10 of them bought milk, 8 bought butter and 6 bought both.
* Bought milk => bought butter
* Support = P (Milk & Butter) = 6/100 = 0.06
* Confidence = support/P(Butter) = 0.06/0.08 = 0.75
* Lift = confidence/P(Milk) = 0.75/0.10 = 7.5

**Support Vector Machine – Classification (SVM)**

* A Support vector machine (SVM) performs classifications by finding the hyperplane that maximizes the margin between the two classes. The vectors (cases) that define the hyperplane are the support vectors.

**Algorithm**

* Define an optimal hyperplane: maximize margin.
* Extend the above definition for non-linearly separable problems: have a penalty term for misclassifications.
* Map data to high dimensional space where it is easier to classify with linear decision surfaces: reformulate problem so that data is mapped implicitly to this space.

Applications of ML

Explain Association Rules mining. What is the “Apriori Principle”?

strength(s) and the weakness(es) of K-means clustering.

difference between classification and clustering?

different types of Learning/Training models in Machine Learning (ML)

advantages and disadvantages of Principal Component

Analysis (PCA),

Market Basket Analysis

SVM hyperplane.

You are training a Multilayer Perceptron (MLP) neural network for a particular

classification task. After, some investigation, your neural network is constructed with 5 input.

variables, one hidden layer with 12 nodes and one output layer with 3 nodes (the classes). How?

many network parameters are required to be tuned/trained? Show your detailed calculations.

Indicative answer: (5+1) x 12 + (12+1) x 3 = 111. We must also include the bias.

“weights” which is a compulsory component in MLP structure.

Market Basket Analysis is one of the key techniques used by large retailers to uncover associations between items. It works by looking for combinations of items that occur together frequently in transactions. Association Rules are widely used to analyze retail basket or transaction data. You have been given the following transaction database that consists of items (a, b, c, d &e) bought in a store by customers.

Find all the closed frequent item sets which are not maximal, along with their support, for a minsupp threshold of 0.3. Procedure: Define first, all the frequent itemsets (10 marks), then all the closed frequent itemsets (10 marks) and finally all the closed frequent itemsets which are not maximal (10 marks). Show all steps/results of your work and justify any decision you have taken in your analysis.

Indicative answer:

We check first the 1-size item sets.

a 5/10 = 0.5

b 0.7

c 0.5

d 0.9

e 0.6

Everything is above min support, so they are candidates for the creation of 2-size item sets.

ab 0.3

ac 0.2 reject, due to less 0.3

ad 0.4

ae 0.4

bc 0.3

bd 0.6

be 0.4

cd 0.4

ce 0.2 reject, due to less 0.3

de 0.6

8 2-itemsets are suitable; they will create the 3-size item sets in the next stage.

Abc reject as it contains ac.

Abd

Abe

Acd reject due to ac.

Ace reject due to ac.

Ade

Bcd

Bce reject due to ce.

Bde

Cde reject due to ce.

Thus, only 5 3-itemsets are suitable candidates. We need to calculate their support.

Abd 0.2 reject, due to less 0.3

Abe 0.2 reject, due to less 0.3

Ade 0.4

Bcd 0.2 reject, due to less 0.3

Bde 0.4

These 2 3-itemsets are ok for the creation of 4-itemsets.

Abcd reject as it includes ac.

Abce reject as it includes bc.

Abde reject as it includes abd.

Acde 0.1 reject, less 0.3

Acde 0.1 reject, less 0.3

So, finally we have 5 1-itemsets, 8 2-itemsets and 2 3-itemsets (all frequent)

In summary, these are the following frequent item sets.

items support

[1] {c} 0.5

[2] {a} 0.5

[3] {e} 0.6

[4] {b} 0.7

[5] {d} 0.9

[6] {b, c} 0.3

[7] {c, d} 0.4

[8] {a, e} 0.4

[9] {a, b} 0.3

[10] {a, d} 0.4

[11] {b, e} 0.4

[12] {d, e} 0.6

[13] {b, d} 0.6

[14] {a, d, e} 0.4

[15] {b, d, e} 0.4

Based on definitions:

Closed Frequent Itemset: An itemset is closed if none of its immediate supersets has the same support as that.

of the itemset.

Maximal frequent itemset: The definition says that an itemset is maximal frequent if none of its immediate

supersets are frequent.

Let us try to find the closed frequent item sets.

items support

[1] {c} 0.5

[2] {a} 0.5

[3] {b} 0.7

[4] {d} 0.9

[5] {b, c} 0.3

[6] {c, d} 0.4

[7] {a, b} 0.3

[8] {d, e} 0.6

[9] {b, d} 0.6

[10] {a, d, e} 0.4

[11] {b, d, e} 0.4

There are 4 item sets from the frequent list (e,

SVM

SVM or Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyperplane which separates the data into classes.

At first approximation what SVMs do is to find a separating line(or hyperplane) between data of two classes. SVM is an algorithm that takes the data as an input and outputs a line that separates those classes if possible.

Lets begin with a problem. Suppose you have a dataset as shown below and you need to classify the red rectangles from the blue ellipses(let’s say positives from the negatives). So your task is to find an ideal line that separates this dataset in two classes (say red and blue).

According to the SVM algorithm we find the points closest to the line from both the classes.These points are called support vectors. Now, we compute the distance between the line and the support vectors. This distance is called the margin. Our goal is to maximize the margin. The hyperplane for which the margin is maximum is the optimal hyperplane.