Advance Machine Learning

01CO1301
4 Credits

Marwadi University
Department of
Computer Engineering

Ravikumar R Natarajan

Course Outcomes

- At the end of the course, students will be able to:
- To understand key concepts, tools and approaches for pattern recognition on complex data sets.
- To learn Kernel methods for handling high dimensional and non-linear patterns.
- To implement state-of-the-art algorithms such as Support Vector Machines and Bayesian networks.
- To Solve real-world machine learning tasks: from data to inference.
- To apply theoretical concepts and the motivations behind different learning frameworks.

Marwadi University

Department of Computer Engineering

Filter Methods

Unit #5

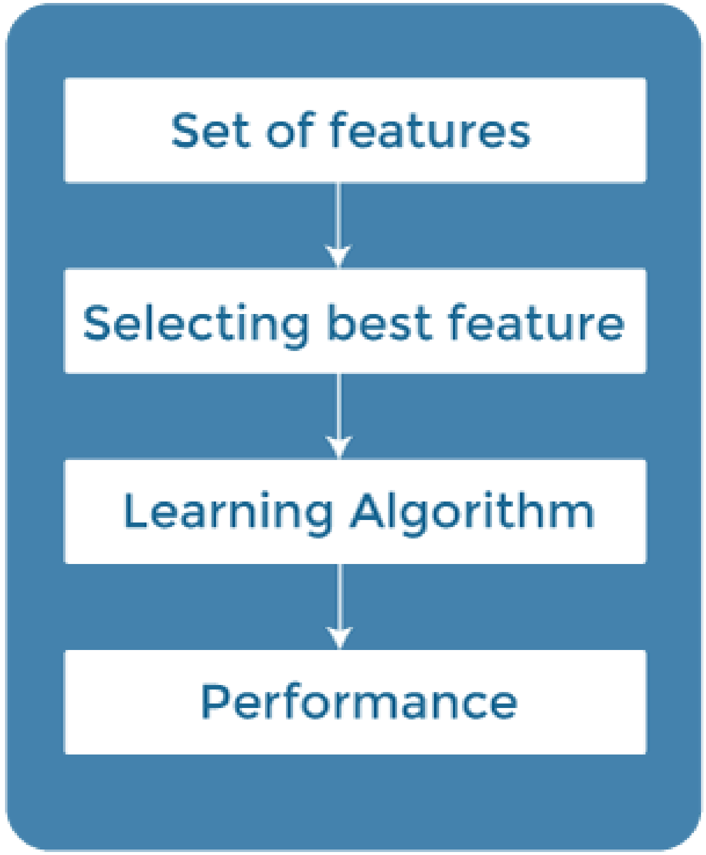
Content

- Filter Methods
  - Sub-space approaches
  - Embedded methods
  - Low-Rank approaches
- Recommender Systems
- Application areas –
  - Security
  - Business
  - Scientific

Filter Methods

- Filter methods are generally used as a preprocessing step. The selection of features is independent of any machine learning algorithms.
- Instead, features are selected on the basis of their scores in various statistical tests for their correlation with the outcome variable.
- In Filter Method, features are selected on the basis of statistics measures. This method does not depend on the learning algorithm and chooses the features as a pre-processing step.
- The filter method filters out the irrelevant feature and redundant columns from the model by using different metrics through ranking.
- The advantage of using filter methods is that it needs low computational time and does not overfit the data

Filter Methods



Filter Methods

- **Filter methods:**
  - information gain
  - chi-square test
  - fisher score
  - correlation coefficient
  - variance threshold
- **Wrapper methods:**
  - recursive feature elimination
  - sequential feature selection algorithms
  - genetic algorithms
- **Embedded methods:**
  - L1 (LASSO) regularization
  - decision tree

Sub-space Approaches

In machine learning the random subspace method, also called attribute bagging or feature bagging.

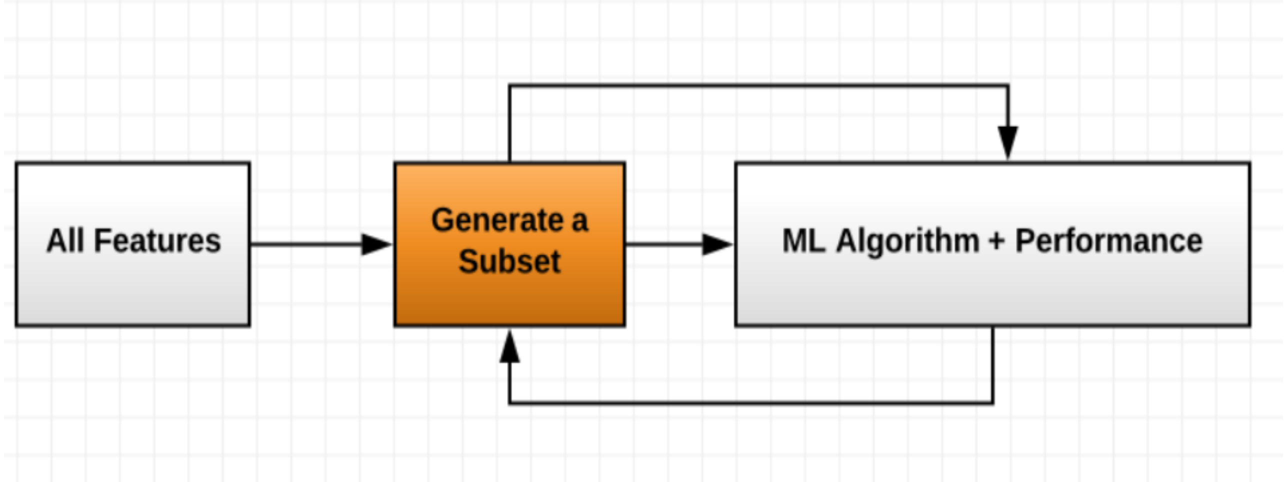
It is an ensemble learning method that attempts to reduce the correlation between estimators in an ensemble by training them on random samples of features instead of the entire feature set.

https://www.sciencedirect.com/topics/computer-science/subspace-method

Embedded Methods

The main goal of feature selection's embedded method is learning which features are the best in contributing to the accuracy of the machine learning model.

They have built-in penalization functions to reduce overfitting:



These encompass the benefits of both the wrapper and filter methods, by evaluating interactions of features but also maintaining reasonable computational cost.

Embedded Methods

The typical steps for embedded methods involve training a machine learning algorithm using all the features, then deriving the importance of those features according to the algorithm used. Afterward, it can remove unimportant features based on some criteria specific to the algorithm.

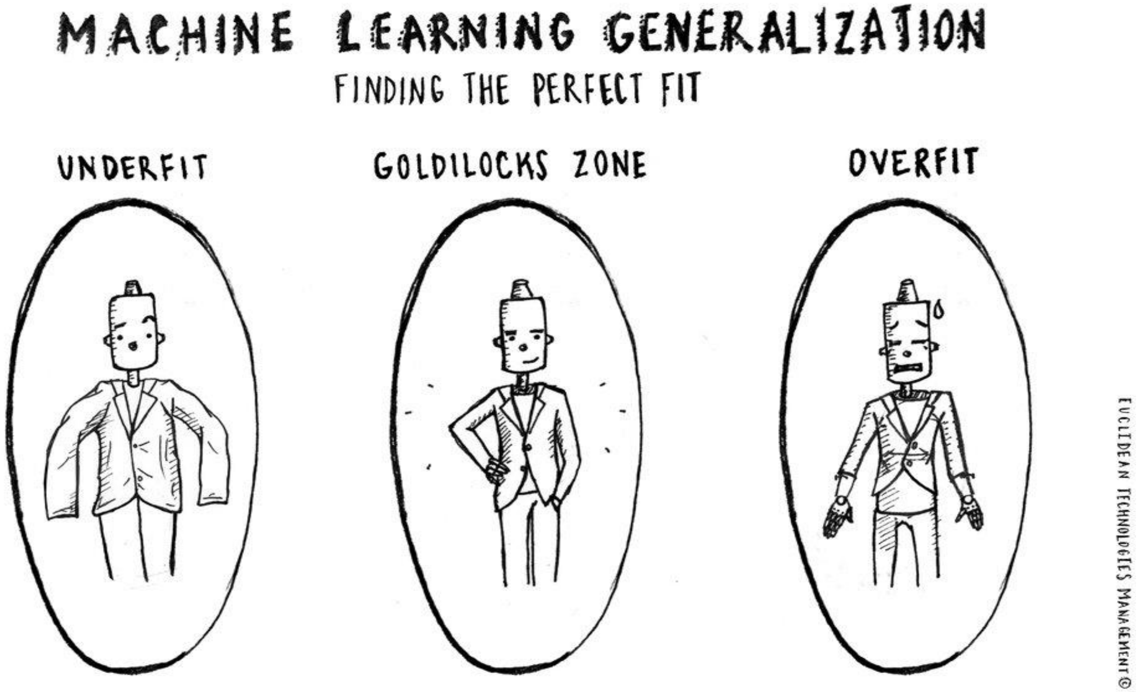
It’s implemented by algorithms that have built-in feature selection methods.

Some of the most popular examples of these methods are LASSO and RIDGE regression which have inbuilt penalization functions to reduce overfitting.

LASSO Regression

- LASSO stands for Least Absolute Shrinkage and Selection Operator.
- **WHY Lasso?**
- When we have less or insufficient data, the model suffers from underfitting. Underfitting reduces the accuracy of our machine learning model. Its occurrence simply means that our model does not fit the data well enough.
- Did you ever try to fit in oversized clothes? A normal Person trying to fit in an extra-large dress refers to the underfitting problem. The same problem occurs in the dataset if you increase the number of features to decrease the cost function.

LASSO Regression



Underfit happens in linear models when dealing with fewer data. If we cannot get rid of this problem, it affects the model performance. Here, Lasso regression comes into the picture. It reduces the underfitting problem in a dataset by using some metrics.

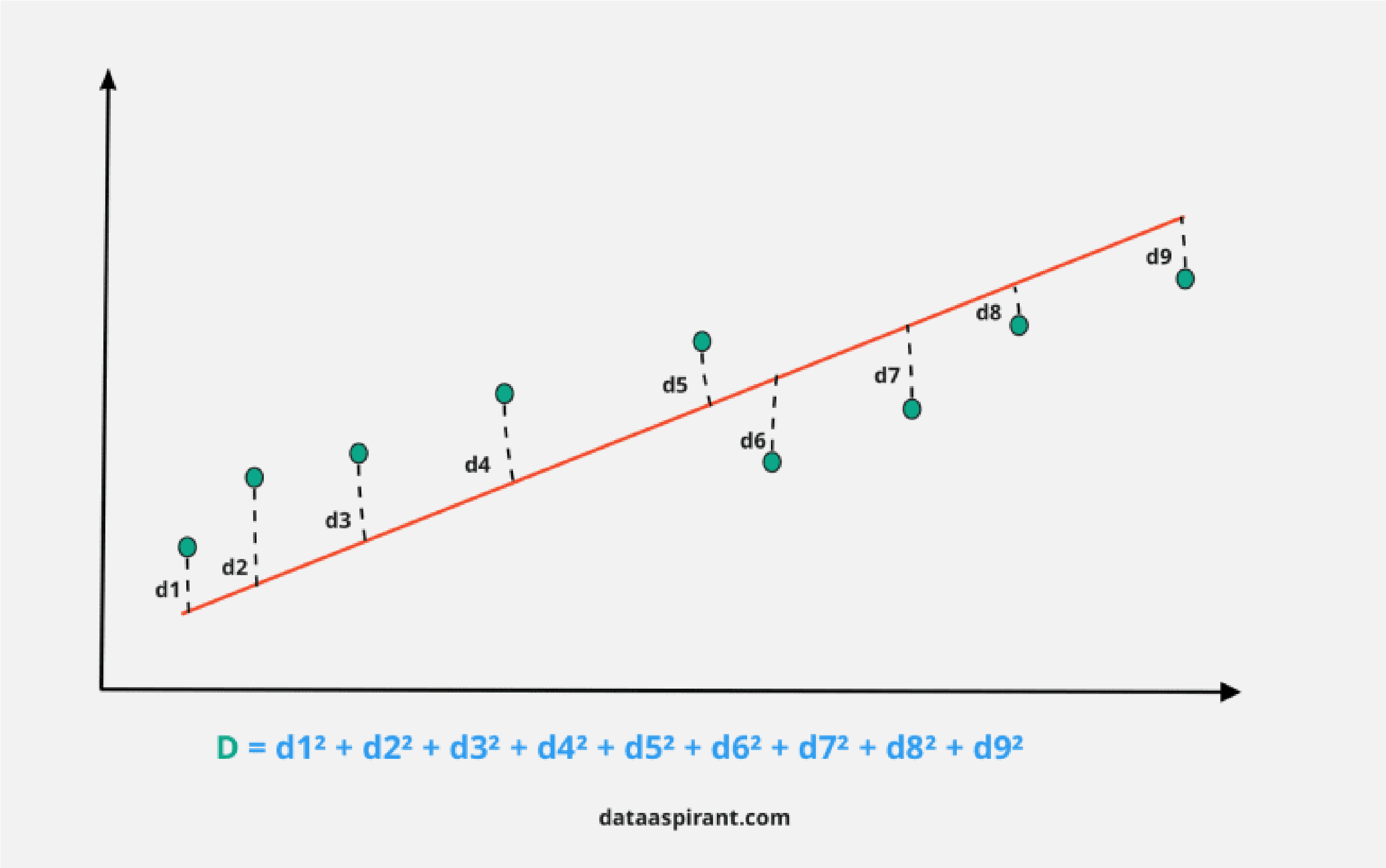
L1 regularization adds penalty equivalent to the absolute value of the magnitude of coefficients.

LASSO Regression

- **What is LASSO?**
- Lasso regression performs **L1** regularization.
- Lasso Regression is almost identical to Ridge Regression, the only difference is the absolute value as opposed to the squaring the weights when computing the ridge regression penalty.
- **Lasso regression is like linear regression, but it uses a technique “shrinkage” where the coefficients of determination are shrunk towards zero to avoid overfitting and make them work better on different datasets.**
- This type of regression is used when the dataset shows high multicollinearity or when you want to automate variable elimination and feature selection.

LASSO Regression

The Statistics of Lasso Regression?



d1, d2, d3, etc., represents the distance between the actual data points and the model line in the above graph.

LASSO Regression

The Statistics of Lasso Regression?

Least-squares is the sum of squares of the distance between the points from the plotted curve.

In linear regression, the best model is chosen in a way to minimize the least-squares.

While performing lasso regression, we add a penalizing factor to the least-squares. That is, the model is chosen in a way to reduce the below loss function to a minimal value. During the Lasso fitting algorithm, the model tries to minimize the difference between the predicted and estimated value of the observation with the penalty.

D = least-squares + lambda * summation (absolute values of the magnitude of the coefficients)

Lasso regression penalty consists of all the estimated parameters. Lambda can be any value between zero to infinity. This value decides how aggressive regularization is performed. It is usually chosen using cross-validation. Lasso penalizes the sum of absolute values of coefficients. As the lambda value increases, coefficients decrease and eventually become zero.

This way, lasso regression eliminates insignificant variables from our model.

Low Rank Approach

- In mathematics, low-rank approximation is a minimization problem, in which the cost function measures the fit between a given matrix (the data) and an approximating matrix (the optimization variable), subject to a constraint that the approximating matrix has reduced rank.
- The problem is used for mathematical modeling and data compression. The rank constraint is related to a constraint on the complexity of a model that fits the data.
- In applications, often there are other constraints on the approximating matrix apart from the rank constraint, e.g., non-negativity and Hankel structure.

- Low-rank approximation is closely related to:
- principal component analysis,
- factor analysis,
- total least squares,
- latent semantic analysis
- orthogonal regression, and
- dynamic mode decomposition.

Reference Paper
https://arxiv.org/pdf/1808.04521.pdf

Low Rank Approach

- **Applications**
- Linear system identification, in which case the approximating matrix is Hankel structured.
- Machine learning, in which case the approximating matrix is nonlinearly structured.
- Recommender systems, in which cases the data matrix has missing values and the approximation is categorical.
- Distance matrix completion, in which case there is a positive definiteness constraint.
- Natural language processing, in which case the approximation is nonnegative.
- Computer algebra, in which case the approximation is Sylvester structured.

Recommendation System

- Recommender systems are the systems that are designed to recommend things to the user based on many different factors. These systems predict the most likely product that the users are most likely to purchase and are of interest to. Companies like Netflix, Amazon, etc. use recommender systems to help their users to identify the correct product or movies for them.

- The recommender system deals with a large volume of information present by filtering the most important information based on the data provided by a user and other factors that take care of the user's preference and interest. It finds out the match between user and item and imputes the similarities between users and items for recommendation.

- Both the users and the services provided have benefited from these kinds of systems. The quality and decision-making process has also improved through these kinds of systems.

Recommendation System

- **Types of Recommendation System**

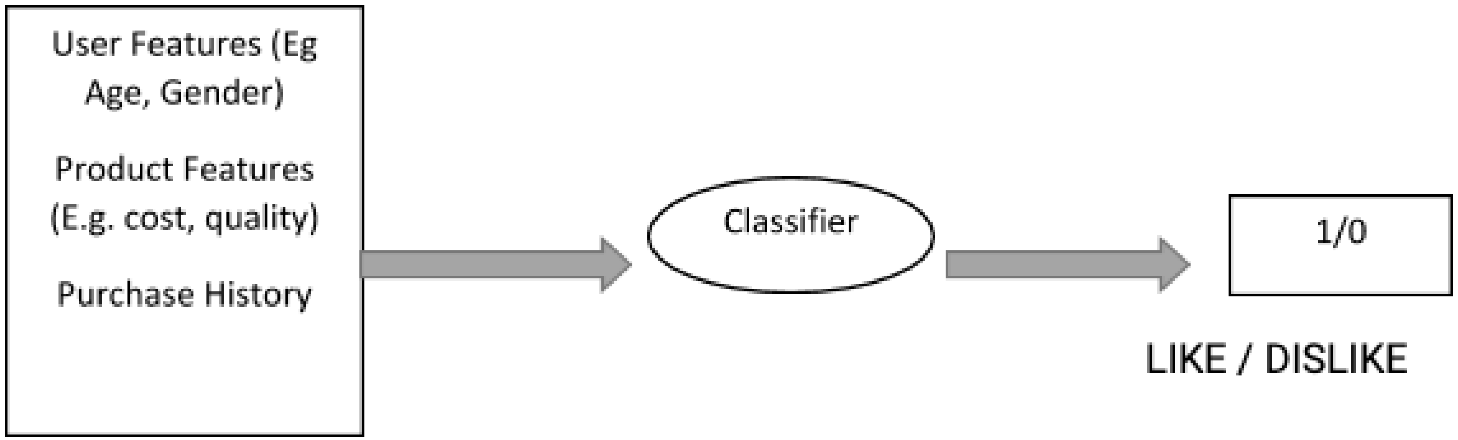
1. Popularity-Based Recommendation System
2. Classification Model
3. Content-Based Recommendation System
    1. Euclidean Distance
    2. Cosine Similarity
    3. Jaccard Similarity
4. Collaborative Filtering
    1. User-based nearest-neighbor collaborative filtering
    2. Item-based nearest-neighbor collaborative filtering
    3. Singular value decomposition and matrix-factorization
        1. Regularization
        2. Bias terms
        3. Minimizing with Stochastic Gradient Descent (SGD)
5. Hybrid recommendation system

Recommendation System

- **Popularity-Based Recommendation System**
- It is a type of recommendation system which works on the principle of popularity and or anything which is in trend. These systems check about the product or movie which are in trend or are most popular among the users and directly recommend those.
- For example, if a product is often purchased by most people then the system will get to know that that product is most popular so for every new user who just signed it, the system will recommend that product to that user also and chances becomes high that the new user will also purchase that.
- **Merits of popularity based recommendation system**
- It does not suffer from cold start problems which means on day 1 of the business also it can recommend products on various different filters.
- There is no need for the user's historical data.
- **Demerits of popularity based recommendation system**
- Not personalized
- The system would recommend the same sort of products/movies which are solely based upon popularity to every other user.
- **Example**
- Google News: News filtered by trending and most popular news.
- YouTube: Trending videos.

Recommendation System

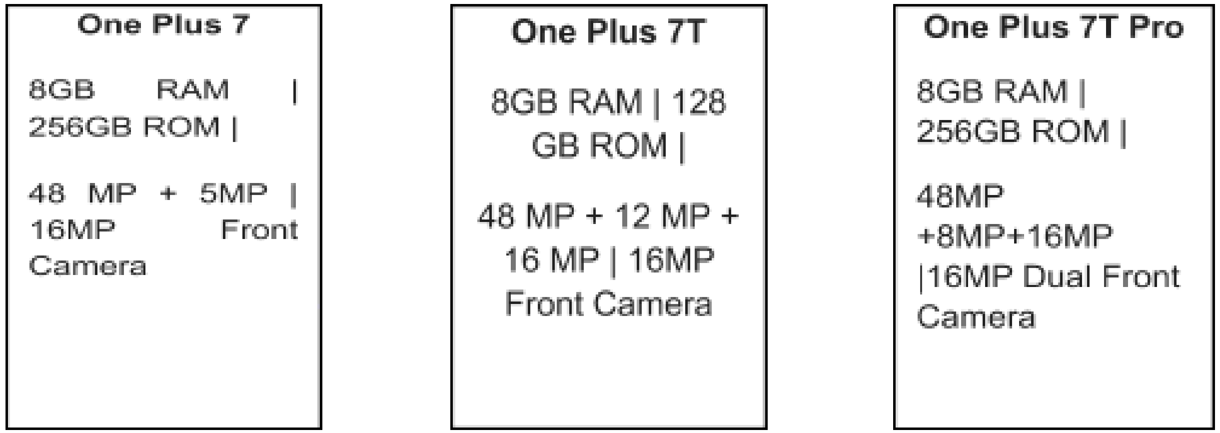
- **Classification Model**
- The model that uses features of both products as well as users to predict whether a user will like a product or not.



- **Limitations of Classification Model**
- It is a rigorous task to collect a high volume of information about different users and also products.
- Also, if the collection is done then also it can be difficult to classify.
- Flexibility issue.

Recommendation System

- **Content-Based Recommendation System**
- It is another type of recommendation system which works on the principle of similar content. If a user is watching a movie, then the system will check about other movies of similar content or the same genre of the movie the user is watching. There are various fundamentals attributes that are used to compute the similarity while checking about similar content.
- To explain more about how exactly the system works, an example is stated below



Different models of one plus.

Recommendation System

- **Content-Based Recommendation System**
- **Euclidean Distance:** Distance between two points can be calculated by the equation;

$$\text{Inner}(x, y) = \sum_i x_i y_i = (x, y)$$

Cosine Similarity: Cosine of the angle between the two vectors of the item, vectors of A and B is calculated for imputing similarity. If the vectors are closer, then small will be the angle and large will be the cosine.

$$\text{Similarity}(X,Y) = \frac{X.Y}{|X| \times |Y|}$$

Jaccard Similarity: Users who have rated item A and B divided by the total number of users who have rated either A or B gives us the similarity. It is used for comparing the similarity.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Recommendation System

- **Content-Based Recommendation System**

Merits

- There is no requirement for much of the user's data.
- We just need item data that enable us to start giving recommendations to users.
- A content-based recommender engine does not depend on the user's data, so even if a new user comes in, we can recommend the user as long as we have the user data to build his profile.
- It does not suffer from a cold start.

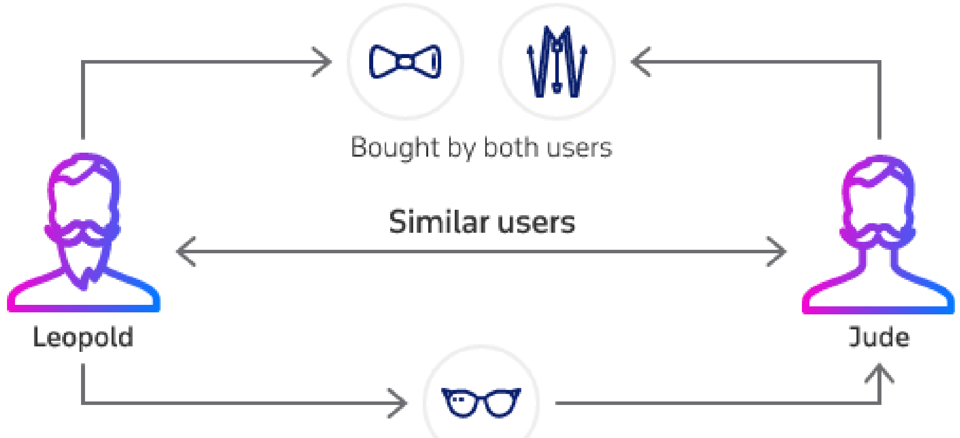
Demerits

- Items data should be in good volume.
- Features should be available to compute the similarity.

Recommendation System

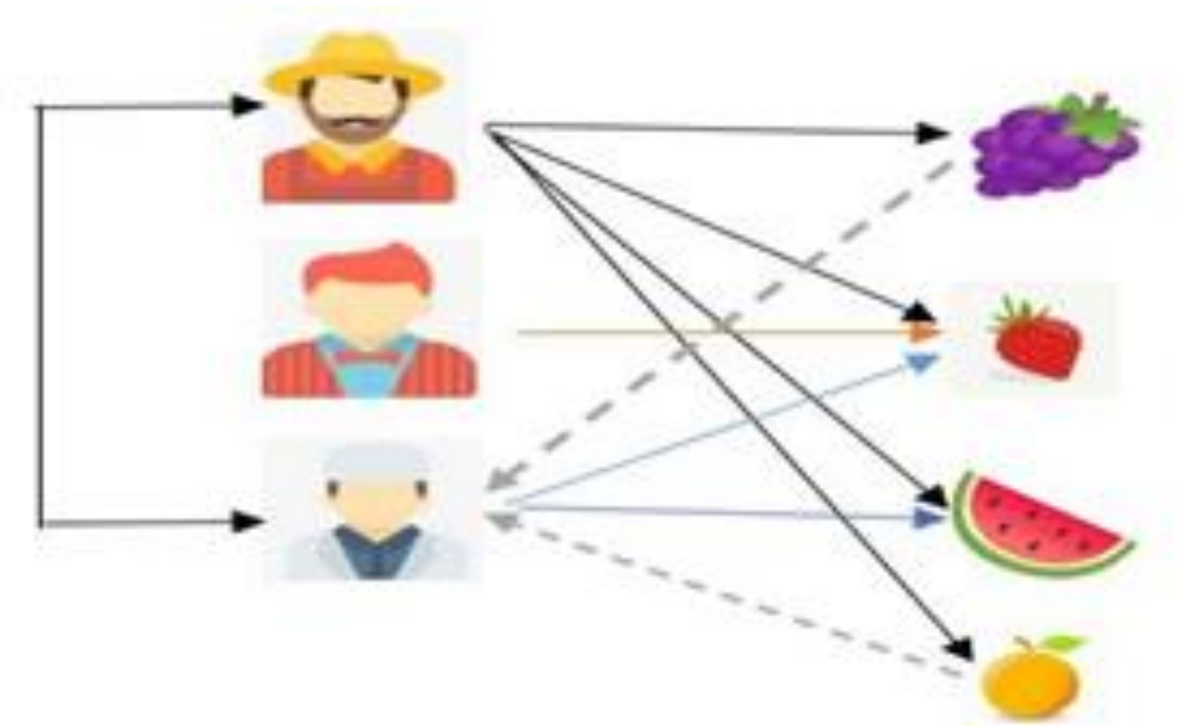
- **Collaborative Filtering**
  - It is considered to be one of the very smart recommender systems that work on the similarity between different users and also items that are widely used as an e-commerce website and also online movie websites. It checks about the taste of similar users and does recommendations.
  - The similarity is not restricted to the taste of the user moreover there can be consideration of similarity between different items also. The system will give more efficient recommendations if we have a large volume of information about users and items.

Collaborative filtering



Recommendation System

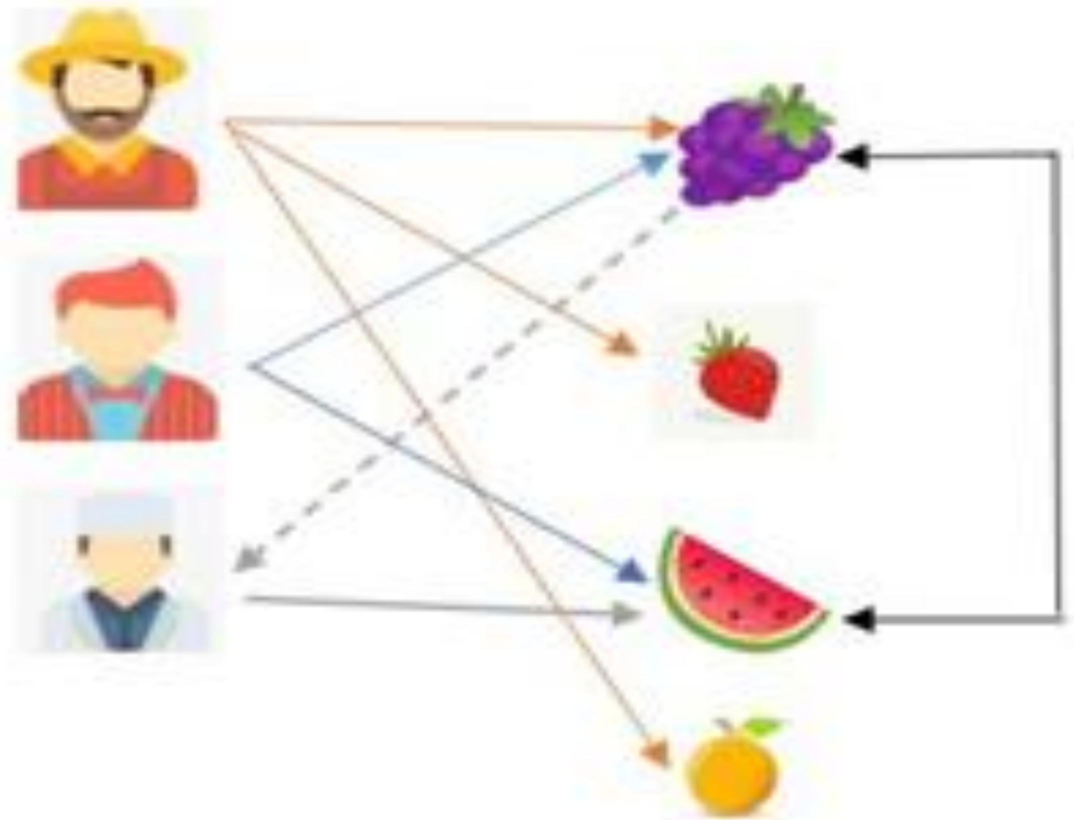
- **User-based nearest-neighbor collaborative filtering**



user-user collaborative filtering where there are three users A, B and C respectively and their interest in fruit. The system finds out the users who have the same sort of taste of purchasing products and similarity between users is computed based upon the purchase behavior. User A and User C are similar because they have purchased similar products.

Recommendation System

- **Item-based nearest-neighbor collaborative filtering**



The system checks the items that are similar to the items the user bought. The similarity between different items is computed based on the items and not the users for the prediction. Users X and Y both purchased items A and B so they are found to have similar tastes.

Recommendation System

- **Collaborative filtering**

Limitations

- Enough users required to find a match. To overcome such cold start problems, often hybrid approaches are made use of between CF and Content-based matching.
- Even if there are many users and many items that are to be recommended often, problems can arise of user and rating matrix to be sparse and will become challenging to find out about the users who have rated the same item.
- The problem in recommending items to the user due to sparsity problems.

Recommendation System

- **SVD – already discussed**

End of Unit 5

Thank You.