



POORNIMA FOUNDATION

LECTURE NOTES

Campus: PCE Course: B.Tech II in CSE Class/Section: III Yr. Section- A Date: 21/3/21
Name of Faculty: Praveen Kumar Yadav Name of Subject: Machine Learning Code: 6CS4-02
Date (Prep): 21/3/21 Date (Del): 21/3/21 Unit No: II Lect. No: 16

OBJECTIVE: To be written before taking the lecture (Pl. write in bullet points the main topics/concepts etc., which will be taught in this lecture)

Association Rule Mining.

IMPORTANT & RELEVANT QUESTIONS:

why ARM is required in ML?

FEED BACK QUESTIONS (AFTER 20 MINUTES):

what are the applⁿ of ARM concept.

OUTCOME OF THE DELIVERED LECTURE: To be written after taking the lecture (Pl. write in bullet points about students' feedback on this lecture, level of understanding of this lecture by students etc.)

Good

REFERENCES: Text/Ref. Book with Page No. and relevant Internet Websites:

scikit-learn with ML.

Association Rule Mining: associations rule mining finds interesting associations and relationship among

(ARM)

large set of data items.

is when this rule is used with sales data there it is called as Market Basket Analysis.

eg.

Tid	Items	Support	Support Count
1	Bread, Milk	[Bread]	4
2	Bread, Diaper, Beer, Eggs	[Beer]	3
3	Milk, Diaper, Beer, Cake	[Eggs]	2
4	Bread, Milk, Diaper, Beer	[Diaper]	4
5	Bread, Milk, Diaper, Cake	[Milk]	4
		[Eggs]	1

Support count (C) - Freqⁿ of occurrence of a itemset

for eg $C(\{Milk, Bread, Diaper\}) = 2$

Frequent - itemset - An itemset whose support is greater than or equal to threshold value.

Support(S) - Support of an itemset is the freq of the itemset w.r. to. number of transaction

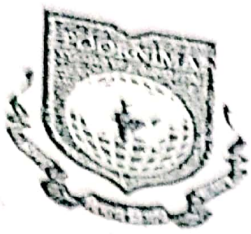
Min Support(S) - Defining support % helps us to select a threshold for freq called as - min-support minimum freq of itemset in the dataset

50% Min Support means - 50% of Total # transaction (i.e. $0.5 \times 5 = 2.5$) means for an itemset to be frequent, it should occur at least three times in 5 transaction in the dataset.

we always eliminate those itemset whose support is less than the min-support.

In ARM, the freq of pattern and association in DS are identified in the then used to predict the next relevant item in the set.

- mostly used as business decisions according to customer purchase.



POORNIMA

COLLEGE OF ENGINEERING

DETAILED LECTURE NOTES

PAGE NO. _____

Recommender system - A RS predict users interest and recommends products that the user may be interested in.
- information filtering technique that provides users with recommendation for which they might be interested in.

- Exⁿ - which websites will you find interesting?
- which degree and university are best for future
- which is the best investment for supporting the education of your children.
- which book should we buy.

How? — recommends items that are most popular among all users.
— divides all users into multiple segments based on their preference and recommends item to them based on the segment they belong to.

Purpose - Prediction Perspective - Predicts to what degree a user likes an item.

Most popular evolution scenario in research

Interaction Perspective - Given users a good feeling
- Educate users about the product domain

Retrieval Perspective - Reduce search cost
Provide correct proposals.

Recommendation Perspective - Identifying item from the long tail



POORNIMA

COLLEGE OF ENGINEERING

DETAILED LECTURE NOTES

PAGE NO

Association Rule Mining:- is a data mining technique that finds pattern in data.

→ The pattern found by Association rule mining represent relationship between items. when this is used with sales data, it is referred to as

Market Basket Analysis.

→ so association rule mining represents relationship that needs to be interpreted to be used in

strategies.

↳ helps in store-organization.

→ It helps businesses to build sales strategies - ^{attract more customers and improve their sales}

→ " " " to " Marketing strategies:-

↳ It helps shelf-life planning:- ^{ex-} olive oil with pizza dough
Knowledge of association rule enable store manager to plan their inventory as well as don't lose by overstocking.

Steps involved in Association Rule mining:-

Step 1. Find all frequent items.

An itemset is a set of items that occurs in a shopping basket

eg. [bread, butter, eggs]

A frequent itemset is one that occurs frequently in a database.

Key Term - Support Count: of an item is defined as the freq. of the item in the dataset

eg.	Tid	Items	Support Count
		[Bread]	5
	1	{bread, milk}	4
	2	{bread, Diapers, Beer, Eggs}	3
	3	{milk, Diapers, Beer, Cola}	4
	4	{Bread, milk, Diapers, Beer}	4
	5	{ " " " " " Cola }	2

Support Count doesn't take into account relative freq. i.e. the freq. w.r. to number of transactions

So Support of an itemset is the freq. of the itemset w.r. to number of transactions (measures a freq. of occurrence)

iv Percentage % of Transactions (T) that contain both 'A' and 'B'.
 $(A \Rightarrow B) = P(A \cap B)$

$$\text{support}(\text{itemset}) = \frac{\text{Freq. of itemset (Support Count)}}{\text{Total no. of transactions}}$$

itemset	Support	%
[Beer]	3/5	0.6
[Bread]	4/5	0.8
[Cola]	2/5	0.4
[Diapers]	4/5	0.8
[Milk]	4/5	0.8
[Eggs]	1/5	0.2

Defining support % helps us to setup a threshold for freq. called min-support.

60% minimum support \Rightarrow cov. of (Total A \Rightarrow 0.6 x 5 = 3 of transactions)

(For an itemset to be frequent, it should occur atleast 3 times in 5 transactions in the given dataset)

\Rightarrow we always eliminate those items whose support is less than support.

Step 2. Generate strong association rules from the frequent itemsets.
Association rules are generated by building associations from frequent itemset generated in step 1. That measure is called as confidence.

If 'A' then 'B' $\{ A \Rightarrow B \}$
 $\downarrow \quad \quad \quad \searrow$
 Antecedent Consequent

Confidence: (C) In a transaction set 'T' if 'C' is the % of times 'B' is present in all the transactions containing 'A'.

$$C = P(B/A) = \frac{P(A \cap B)}{P(A)} \quad \left\{ \begin{array}{l} \text{shows the strength} \\ \text{of Association} \end{array} \right.$$

or $= \text{support}(A \cup B) / \text{support}(A)$

Parameters - ① Finding all items that appear frequently in transactions (using min support)

② Finding strong association among freq. items & confidence.