Last class (Market Basket Analysis)

1) Problem statement
2) Data preprocessing
3) Apriori Algorithm
4) Association Rule
5) Various metrics (Few Remaining)
6) ~~Conclusion~~

Today's class

1) Leverage, Conviction
2) Summary of market basket analysis
3) Intro to Reco Sys
4) Formulating Problem for Reco Sys
5) Collaborative Filtering
6) Item-item and User-user similarity
7) Cold-start problem
8) Content-based Reco Sys

9) Reco Sys as a Classification / Regression Problem

10) Summary

11) Matrix Factorization

Hybrid Model

features

User i → (U_i)

age	gender	location	device used
30	M	Delhi	Android

item j → (I_j)

genre	length of movie	language
action	150 min	hindi + tamil

U_i I_j

$$\begin{bmatrix} 30 & m & D & A \end{bmatrix} \cdot \begin{bmatrix} A & 150 & H+T \end{bmatrix}$$

← features →

X

Output / Target (Y)

R_{ij} → 0 to 10

(7.5)

Regression

$W x + b = y$

→ Linear Regression model

$W x_p + b = y_p$

DT, RF, Boosting, NN, Polynomial, KNN,] Any regression technique

Confidence: $C(x \to y) = P(y/x)$

$= \frac{P(x \cap y)}{P(x)}$

Lift: $L(x \to y) = \frac{P(x \cap y)}{P(x) \cdot P(y)}$

$= \frac{P(y/x)}{P(y)}$ ←

$$P(Br) = \frac{1.25}{100} \qquad N = 600,000$$

$$P(Milk) = \frac{10}{100}$$

$$P(Br \cap M) = \frac{1}{100}$$

$$C(Br \rightarrow M) = \frac{P(Br \cap M)}{P(Br)}$$

$$= \frac{\frac{1}{100}}{\frac{1.25}{100}} = \frac{1}{1.25} = 0.8$$

$$L(Br \rightarrow M) = \frac{\frac{1}{100} \times \frac{4}{5}}{\frac{1.25}{100} \times \frac{10}{100}} = \frac{400}{50}$$

$$= 8$$

How to decide which is antecedent and which is consequent?

Our dataset has more support for antecendent green but, with lower confidence.

d) Between high support, low confidence

&

low support, high confidence

which one should be chosen when?