Assignment Project Exam Help

# Recommendation Engines

https://powcoder.com

Add WeChat powcoder



HARVARD UNIVERSITY

# Agenda

| Start | End | Item |
| --- | --- | --- |
| | | Business Context |
| | | Association Rules |
| | | Code Example |
| | | Collaborative Filtering |
| | | Code Example |
| | | Personalized Reco Engine |
| | | |
| | | |
| | | |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



HARVARD UNIVERSITY

# What are Association Rules?

- Study of “what goes with what”
    - “Customers who bought X also bought Y”

Assignment Project Exam Help

- Transaction-based or event-based
    - Customer A bought peanut butter and bread.
    - When someone has body aches, and fever they also have chills

https://powcoder.com

Add WeChat powcoder

- Also called “market basket analysis” and “affinity analysis”

- Originated with study of customer transactions databases to determine associations among items purchased



HARVARD UNIVERSITY

# What are some other recommendation systems you have encountered?

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

# Many business profit by having exceptional recommendation analysis.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



HARVARD UNIVERSITY

# Generating Rules

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

# Before Data Mining

Heuristic Based - A heuristic is a mental shortcut that allows people to solve problems and make judgments quickly and efficiently.

Assignment Project Exam Help

"Hey Lumberg, you should put the salsa next to the tortilla chips in the grocery aisle."

https://powcoder.com

Add WeChat powcoder

Based on experience and/or intuition based on existing mental maps.

- Reduces Effort
- Fast and Cognitively Frugal



HARVARD UNIVERSITY

# With Data Mining - Association Rules

- Hypothesis driven
- Supported by robust transaction level data


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

# Example Association Rule

Association Rule Based – Using transactional data, identify *antecedent* & *consequent* item-sets.

Assignment Project Exam Help

If a customer buys tortilla chips then they will seek out and purchase salsa.

https://powcoder.com

Add WeChat powcoder

Enabled by computers, affinity analysis can explore relationships more complex than previously possible with heuristics alone.

- Increased number of relationships yielding additional consumer insight, and $$
- DRAWBACKS
  - Technical Acumen
  - Without shortcuts, computationally intensive



HARVARD UNIVERSITY

# Example Association Rule

If a customer buys <u>tortilla chips</u> then they will seek out and purchase <u>salsa</u>.

Assignment Project Exam Help

Association Rule

- "IF" part = **antecedent**
- "THEN" part = **consequent**

https://powcoder.com

*Association rules must be "disjoint" meaning items in the antecedent & consequent are not shared.*

Add WeChat powcoder

Affected Items

- <u>ITEMSET = {tortilla chips, salsa}</u>

Using R, we will identify many rules with antecedents and consequents.



HARVARD UNIVERSITY

# Example Association Rule

If a customer buys salsa then they will seek out and purchase tortilla chips.

Assignment Project Exam Help

Association Rule

- “IF” part = **antecedent**
- “THEN” part = **consequent**

https://powcoder.com

*Association rules must be “disjoint” meaning items in the antecedent & consequent are not shared.*

Add WeChat powcoder

Affected Items

- ITEMSET = {tortilla chips, salsa}

Because this is transaction based, there is really no 1st item to determine the antecedent/consequent order. As a result, the items are a set which can be reordered into two rules.

# What is the antecedent & consequent?

Assignment Project Exam Help

If a patient has poor circulation then check oxygen levels.

https://powcoder.com

Add WeChat powcoder



HARVARD UNIVERSITY

# What is the antecedent & consequent?

Assignment Project Exam Help

If a patient has poor circulation then check oxygen levels.

https://powcoder.com

Antecedent Consequent

Add WeChat powcoder

{ poor circulation, oxygen levels}

# What is the antecedent & consequent?

Assignment Project Exam Help

If a customer listens to Imagine Dragons then they may listen to AWOL-Nation & 21 Pilots

https://powcoder.com

Add WeChat powcoder

# What is the antecedent & consequent?

If a customer listens to Imagine Dragons then they may listen to AWOL-Nation & 21 Pilots

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Antecedent

Consequent

{ Imagine Dragons, AWOL-Nation, 21 Pilots}

# What is the antecedent & consequent?

Assignment Project Exam Help

If a customer buys bread then they will buy cheese, meat and bread.

https://powcoder.com

Add WeChat powcoder

# What is the antecedent & consequent?

Assignment Project Exam Help

If a customer buys bread then they will buy cheese, meat and bread.

WRONG!

https://powcoder.com

Add WeChat powcoder

{ BREAD, meat, cheese, BREAD}

Not Disjoint

# Tiny Example: Phone Faceplates

| Transaction | Faceplate Colors Purchased | | | |
|---|---|---|---|---|
| 1 | red | white | green | |
| 2 | white | orange | | |
| 3 | white | blue | | |
| 4 | red | white | orange | |
| 5 | red | blue | | |
| 6 | white | blue | | |
| 7 | white | orange | | |
| 8 | red | white | blue | green |
| 9 | red | white | blue | |
| 10 | yellow | | | |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



HARVARD UNIVERSITY

# Many Rules are Possible

Transaction 1 supports several rules, such as

- "If red, then white" ("If a red faceplate is purchased, then so is a white one")
- "If white, then red"
- "If red and white, then green"
- + several more

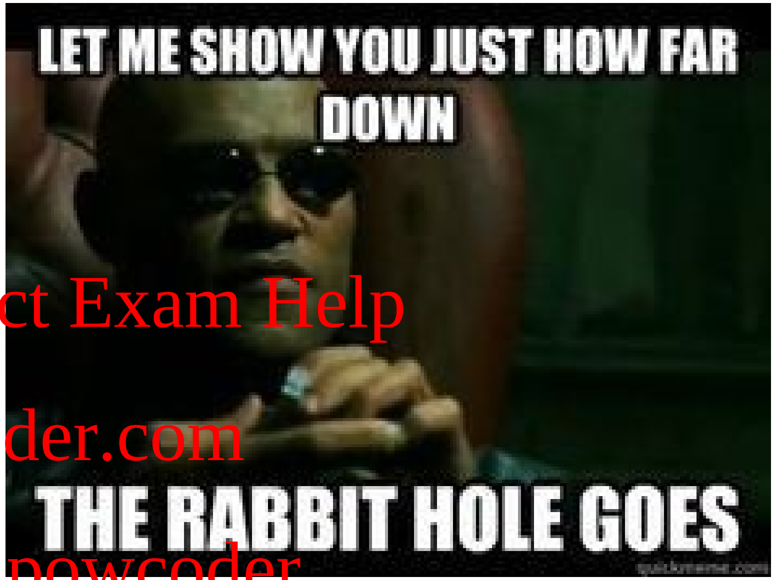
| Transaction | Faceplate Colors Purchased | | |
|---|---|---|---|
| 1 | red | white | green |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

HARVARD UNIVERSITY

# Rules on Rules on Rules…10 transactions yet many possibilities

| Transaction | Faceplate Colors Purchased | | | |
|---|---|---|---|---|
| 1 | red | white | green | |
| 2 | white | orange | | |
| 3 | white | blue | | |
| 4 | red | white | orange | |
| 5 | red | blue | | |
| 6 | white | blue | | |
| 7 | white | orange | | |
| 8 | red | white | blue | green |
| 9 | red | white | blue | |
| 10 | yellow | | | |




Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Single Antecedent

{red, white}
{red, white, green}
{white, red}
{white, green}
{red, green}
{green, …}

Double Antecedent

{red, white, green}
{white, green, red}
{red, green, white}

Triple Antecedent

{red, white, blue, green}

*Haven't even gotten to orange, blue, and yellow.*

# Frequent Item Sets

- Ideally, we want to create all possible combinations of items
- **Problem:** computation time grows exponentially as # items increases
- **Solution:** consider only "frequent item sets"
- Criterion for frequent: *support*

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Focusing on frequent item sets keeps your rules from finding pockets of associations that have little evidence or business value.

E.g. If a person buys bread at Wal-Mart, then they will also buy a bike lock…sure that happens but likely not as often as other consequent items.

# Support

| Transaction | Faceplate Colors Purchased | | |
|---|---|---|---|
| 1 | red | white | green |
| 2 | white | orange | |
| 3 | white | blue | |
| 4 | red | white | orange |
| 5 | red | blue | |
| 6 | white | blue | |
| 7 | white | orange | |
| 8 | red | white | blue | green |
| 9 | red | white | blue |
| 10 | yellow | | |

*Support for an itemset* = # of transactions that include an itemset

- Example: support for the item set {red, white} is 4 out of 10 transactions, or 40%

*Support for a rule* = # of transactions that include both the antecedent *and* the consequent

$$\text{SUPPORT} = \frac{\text{Number of transactions with Item Set}}{\text{Total Number of Transactions}}$$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Assignment Project Exam Help

# Apriori Algorithm

https://powcoder.com

Add WeChat powcoder

HARVARD
UNIVERSITY

# Generating Frequent Item Sets

For *k* products…

1. User sets a minimum support criterion
2. Next, generate list of one-item sets
3. Reduce the set of 1 items to only those meeting the support criterion
4. Use the reduced list of one-item sets to generate list of two-item sets, omitting any items that were previously removed.
5. Reduce the set of 2 items to only those meeting the support criterion
6. Use the reduced list of two-item sets to generate list of three-item sets, omitting any items that were previously removed.
7. Reduce the set of 3 items to only those meeting the support criterion
8. Continue up through *k*-item sets

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

By calculating the first run of 1 item sets (if white, then red) you learn about frequencies as you find more complex item sets. If item sets don't have support in the first run, they won't have support in later runs.

# Apriori Algo

| Transaction | Faceplate Colors Purchased | | |
|---|---|---|---|
| 1 | red | white | green |
| 2 | white | orange | |
| 3 | white | blue | |
| 4 | red | white | orange |
| 5 | red | blue | |
| 6 | white | blue | |
| 7 | white | orange | |
| 8 | red | white | blue | green |
| 9 | red | white | blue |
| 10 | yellow | | |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

If yellow then white. → 0% support

Since {yellow, and any other color} has no support, there is no need to check it for subsequent item sets such as {yellow, white, blue} or {yellow, white, blue, green}

# Apriori Algo

| Transaction | Faceplate Colors Purchased | | |
|---|---|---|---|
| 1 | red | white | green |
| 2 | white | orange | |
| 3 | white | blue | |
| 4 | red | white | orange |
| 5 | red | blue | |
| 6 | white | blue | |
| 7 | white | orange | |
| 8 | red | white | blue | green |
| 9 | red | white | blue |
| 10 | yellow | | |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

If yellow then white. → 0% support

Support Criteria = 40%

- {Red} 5/10
- {White} 8/10
- ~~{Green} 2/10~~
- {Orange} 3/10
- {Blue} 4/10
- ~~{Yellow} 1/10~~

Since {yellow, and any other color} has no support, there is no need to check it for subsequent item sets such as {yellow, white, blue} or {yellow, white, blue, green}



HARVARD UNIVERSITY

# Apriori Algo

| Transaction | Faceplate Colors Purchased | | | |
|---|---|---|---|---|
| 1 | red | white | green | |
| 2 | white | orange | | |
| 3 | white | blue | | |
| 4 | red | white | orange | |
| 5 | red | blue | | |
| 6 | white | blue | | |
| 7 | white | orange | | |
| 8 | red | white | blue | green |
| 9 | red | white | blue | |
| 10 | yellow | | | |

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

If red then white. → 40% support

If white then orange. → 30% support

Since {red, white} and {white, orange} are above our 30% criterion, we should check for three item sets. However if our support criterion is 40% we wouldn't explore {white, orange …} because its support is 30%

# Apriori Algo

| Transaction | Faceplate Colors Purchased | | | |
|---|---|---|---|---|
| 1 | red | white | green | |
| 2 | white | orange | | |
| 3 | white | blue | | |
| 4 | red | white | orange | |
| 5 | red | blue | | |
| 6 | white | blue | | |
| 7 | white | orange | | |
| 8 | red | white | blue | green |
| 9 | red | white | blue | |
| 10 | yellow | | | |

If red then white. → 40% support

If white then orange. → 30% support

If white then blue. → 40% support

Support Criteria = 40%

- {Red/White} 4/10
- ~~{White/Orange} 3/10~~
- {Blue/White} 4/10
- ~~{Red/Blue} 3/10~~
- …

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Since {red, white} and {white, orange} are above our fake 30% criterion, we should check for three item sets.

However if our support criterion is 40% we wouldn't explore {white, orange …} because its support is 30%

# Apriori Algo

| Transaction | Faceplate Colors Purchased | | |
|---|---|---|---|
| 1 | red | white | green |
| 2 | white | orange | |
| 3 | white | blue | |
| 4 | red | white | orange |
| 5 | red | blue | |
| 6 | white | blue | |
| 7 | white | orange | |
| 8 | red | white | blue | green |
| 9 | red | white | blue |
| 10 | yellow | | |

Support Criteria = 40%

- ~~{Red/White/Green} 2/10~~
- ~~{Red/White/Orange} 3/10~~
- ~~{Red/White/Blue} 2/10~~

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

If red , white then blue. → 20% support

If red , white then orange. → 30% support

If red , white then green. → 20% support

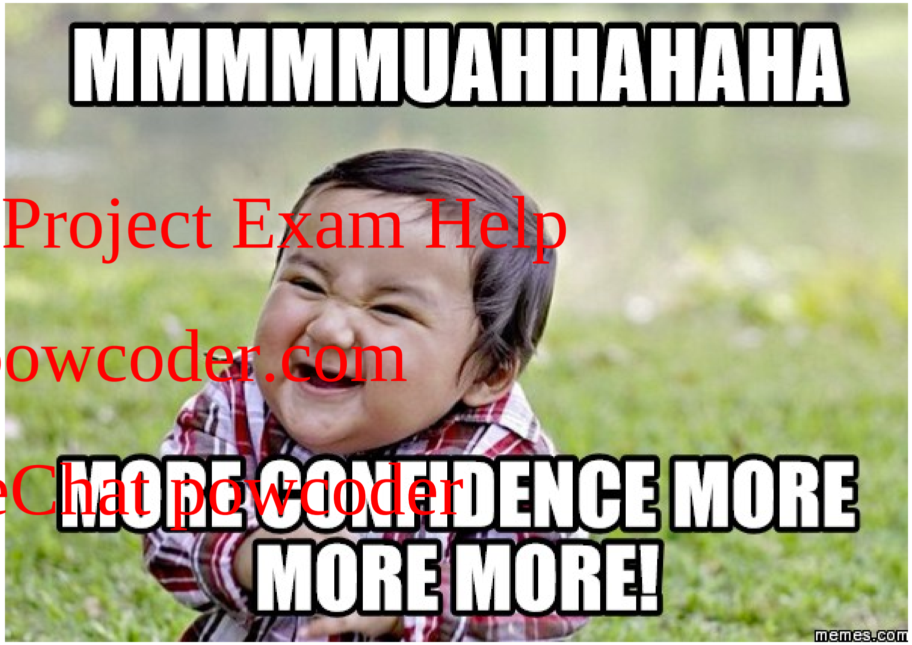
You only check rules that are still in the pool, so no sets with {yellow} or with {white, orange} . *We show {white, red, orange} but its not really in the pool any longer.*



HARVARD UNIVERSITY

# Measures of Rule Performance

***Confidence:*** the % of antecedent transactions that also have the consequent item set




Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

How confident are you that the ***antecedent*** isn't just naturally occurring.

# Confidence

Suppose you have 100,000 transactions.

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

# KPI Example – Baseline Transactions

Suppose you have 100,000 transactions.

Number of Antecedent Rules

If candles, cake mix, then any other item = 5000 Transactions

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Bike Lock Rule

If candles, cake mix, then bike lock = 100 Transactions

Balloon Rule

If candles, cake mix, then balloon = 500 Transactions



HARVARD UNIVERSITY

# SUPPORT – *frequency among all transactions*

## Suppose you have 100,000 transactions.

| | | | |
|---|---|---|---|
| Number of Antecedent Rules | If candles, cake mix, then any other item | = | 5000 Transactions |
| Bike Lock Rule | If candles, cake mix, then bike lock | = | 100 Transactions |
| Balloon Rule | If candles, cake mix, then balloon | = | 500 Transactions |
| Bike Lock Support | 100 Transactions | ÷ | 100 / 100,000 or .001 |
| Balloon Support | 500 Transactions | ÷ | 500 / 100,000 or .005 |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



HARVARD UNIVERSITY

# CONFIDENCE – *better than natural occurrence*

| | | | |
|---|---|---|---|
| Number of Antecedent Rules | If candles, cake mix, then any other item | = | 5000 Transactions |
| Bike Lock Rule | If candles, cake mix, then bike lock | = | 100 Transactions |
| Balloon Rule | If candles, cake mix, then balloon | = | 500 Transactions |
| Bike Lock Confidence | 100 Transactions | ÷ | 100 / 5,000 or .02 |
| Balloon Confidence | 500 Transactions | ÷ | 500 / 5,000 or .10 |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



HARVARD UNIVERSITY

# Confidence

| Item Set | Support | Confidence |
|---|---|---|
| {candles, cake mix, bike lock} | .001 | .02 |
| {candles, cake mix, balloon} | .005 | .10 |

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Both have low support, meaning its an infrequent purchase. BUT when it does happen the confidence is higher that balloons will be purchased.



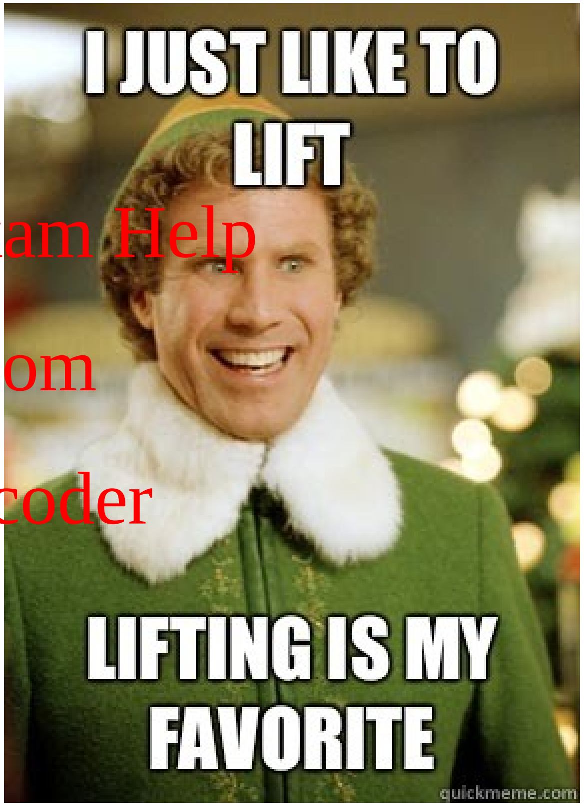
HARVARD UNIVERSITY

# Measures of Rule Performance

**Lift** = *confidence/(benchmark confidence)*

***Benchmark confidence*** = transactions with consequent as % of all transactions

Lift > 1 indicates a rule that is useful in finding consequent items sets (i.e., more useful than just selecting transactions randomly)




Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

People naturally buy or select the ***<u>consequent</u>*** at some rate, how much better is this specific antecedent item(s)?

# Lift

## <u>Ballooon confidence is 0.1</u>

<u>Benchmark</u>

<u>CONFIDENCE</u>

"then balloon" occurs 600 times out of 100k transactions ÷ 600 / 100K or .006

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

# Lift

## Now you know confidence is .1.

Benchmark CONFIDENCE

"then balloon" occurs 600 times out of 100k transactions ÷ 600 / 100K or .006

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Lift

For balloon purchases, how much better than the average occurrence is the rule? ÷ .1 / .006 or 16.6

This rule is providing lift over the natural propensity to purchase balloons.



HARVARD UNIVERSITY

# Alternate Data Format: Binary Matrix

| Transaction | Red | White | Blue | Orange | Green | Yellow |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| 2 | 0 | 1 | 0 | 1 | 0 | 0 |
| 3 | 0 | 1 | 1 | 0 | 0 | 0 |
| 4 | 1 | 1 | 0 | 1 | 0 | 0 |
| 5 | 1 | 0 | 1 | 0 | 0 | 0 |
| 6 | 0 | 1 | 1 | 0 | 0 | 0 |
| 7 | 1 | 0 | 1 | 0 | 0 | 0 |
| 8 | 1 | 1 | 1 | 0 | 1 | 0 |
| 9 | 1 | 1 | 1 | 0 | 0 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 1 |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

# Process of Rule Selection

Generate all rules that meet specified support & confidence

1. Find frequent item sets (those with sufficient support – see previous)
   - Avoids investigating possible all rules (a priori algorithm)
2. From the frequent item sets, generate rules with sufficient confidence & lift
   - Only use rules that you are confident have antecedents that occur more than natural (confidence)
   - Only use rules where the antecedent / consequent relationship is stronger than (lift) how often the consequent occurs

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

HARVARD UNIVERSITY

# Generating Rules in R

P(green) if you use the rule

How much better your chances of getting a green are if you use the rule than if you select randomly

Assignment Project Exam Help

```
     lhs                 rhs support confidence lift
15 {Red,White}    => {Green} 0.2     0.5      2.500000
5 {Green}         => {Red}   0.2     1.0      1.666667
14 {White,Green} => {Red}   0.2     1.0      1.666667
4 {Orange}        => {White} 0.2     1.0      1.428571
6 {Green}         => {White} 0.2     1.0      1.428571
13 {Red,Green}    => {White} 0.2     1.0      1.428571
```

https://powcoder.com
Add WeChat powcoder

- **Support**: red, white, green occurs 2 out of 10 times (0.20)
- **Confidence**: 50% of the time, red/white will result with green.
- **Lift**: Red/white will result in green 2.5 times more often than green usually.

HARVARD UNIVERSITY

# Interpretation

- *Support* measures overall occurrence
- *Confidence* shows the rate at which consequents will be found given an antecedent (useful in learning costs of promotion)
- *Lift ratio* shows how effective the rule is in finding consequents above or below the natural selection of the consequence

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

HARVARD UNIVERSITY

# Example: Charles Book Club

**TABLE 14.7** SUBSET OF BOOK PURCHASE TRANSACTIONS IN BINARY MATRIX FORMAT

| ChildBks | YouthBks | CookBks | DoItYBks | cefBks | ArtBks | GeogBks | ItalCook | ItalAtlas | ItalArt | Florence |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder


Row 1, e.g., is a transaction in which books were bought in the following categories: Youth, Do it Yourself, Geography

# Let’s Practice! Open A_AssociationRules.R

| VARIABLE NAMES | DESCRIPTIONS |
|---|---|
| Seq# | Sequence number in the partition |
| ID# | Identification number in the full (unpartitioned) market test data set |
| Gender | O=Male, 1=Female |
| M | Monetary- Total money spent on books |
| R | Recency- Months since last purchase |
| F | Frequency- Total number of purchases |
| FirstPurch | Months since first purchase |
| ChildBks | Number of purchases from the category: Child books |
| YouthBks | Number of purchases from the category: Youth books |
| CookBks | Number of purchases from the category: Cookbooks |
| DoItYBks | Number of purchases from the category Do It Yourself books |
| RefBks | Number of purchases from the category: Reference books (Atlases, Encyclopedias, Dictionaries) |
| ArtBks | Number of purchases from the category: Art books |
| GeoBks | Number of purchases from the category: Geography books |
| ItalCook | Number of purchases of book title: "Secrets of Italian Cooking." |
| ItalAtlas | Number of purchases of book title: "Historical Atlas of Italy." |
| ItalArt | Number of purchases of book title: "Italian Art." |
| Florence | =1 "The Art History of Florence." was bought,<br>=0 if not |
| Related purchase | Number of related books purchased |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



HARVARD UNIVERSITY

# Summary – Association Rules

- Association rules (or *affinity analysis*, or *market basket analysis*) produce rules on associations between items from a database of transactions
- Widely used in **recommender systems**
- Most popular method is **Apriori algorithm**

Assignment Project Exam Help

- To reduce computation, we consider only "frequent" item sets (=support)

https://powcoder.com

- Performance of rules is measured by *confidence* and *lift*

Add WeChat powcoder

- Can produce a profusion of rules; review is required to identify useful rules and to reduce redundancy

HARVARD UNIVERSITY

# Agenda

| Start | End | Item |
|---|---|---|
| | | Business Context |
| | | Association Rules |
| | | Code Example |
| | | Collaborative Filtering |
| | | Code Example |
| | | Personalized Reco Engine |
| | | |
| | | |
| | | |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



HARVARD UNIVERSITY

# Collaborative Filtering

- User based methods
- Item based methods

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder


HARVARD UNIVERSITY

# Item-user matrix

- Cells are user preferences, $r_{ij}$, for items
- Preferences can be ratings, or binary (buy, click, like)

Assignment Project Exam Help

https://powcoder.com

| | Item ID | | | |
|---|---|---|---|---|
| User ID | $I_1$ | $I_2$ | $\cdots$ | $I_p$ |
| $U_1$ | $r_{1,1}$ | $r_{1,2}$ | $\cdots$ | $r_{1,p}$ |
| $U_2$ | $r_{2,1}$ | $r_{2,2}$ | $\cdots$ | $r_{2,p}$ |
| $\vdots$ | | | | |
| $U_n$ | $r_{n,1}$ | $r_{n,2}$ | $\cdots$ | $r_{n,p}$ |

Add WeChat powcoder

# More efficient to store as rows of triplets

Each row has the user ID, the item ID, and the user's rating of that item

$(U_u, I_i, r_{ui})$

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

```
> head(songListens)
                                    userID            song_id listens
1 b80344d063b5ccb3212f76538f3d9e43d87dca9e SOAKIMP12A8C130995       1
2 b80344d063b5ccb3212f76538f3d9e43d87dca9e SOBBMDR12A8C13253B       2
3 b80344d063b5ccb3212f76538f3d9e43d87dca9e SOBXHDL12A81C204C0       1
4 b80344d063b5ccb3212f76538f3d9e43d87dca9e SOBYHAJ12A6701BF1D       1
5 b80344d063b5ccb3212f76538f3d9e43d87dca9e SODACBL12A8C13C273       1
6 b80344d063b5ccb3212f76538f3d9e43d87dca9e SODDNQT12A6D4F5F7E       5
```

Triplets restructure as a longer format but are more memory efficient.

HARVARD UNIVERSITY

# User-based Collaborative Filtering

- Start with a single user who will be the target of the recommendations
- Find other users who are most similar, based on comparing preference vectors

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Association rules is a frequentist look at items or transactions. Collaborative filtering is seeking to identify similarities by user preference or item features.

# Measuring Proximity Pearson Correlation

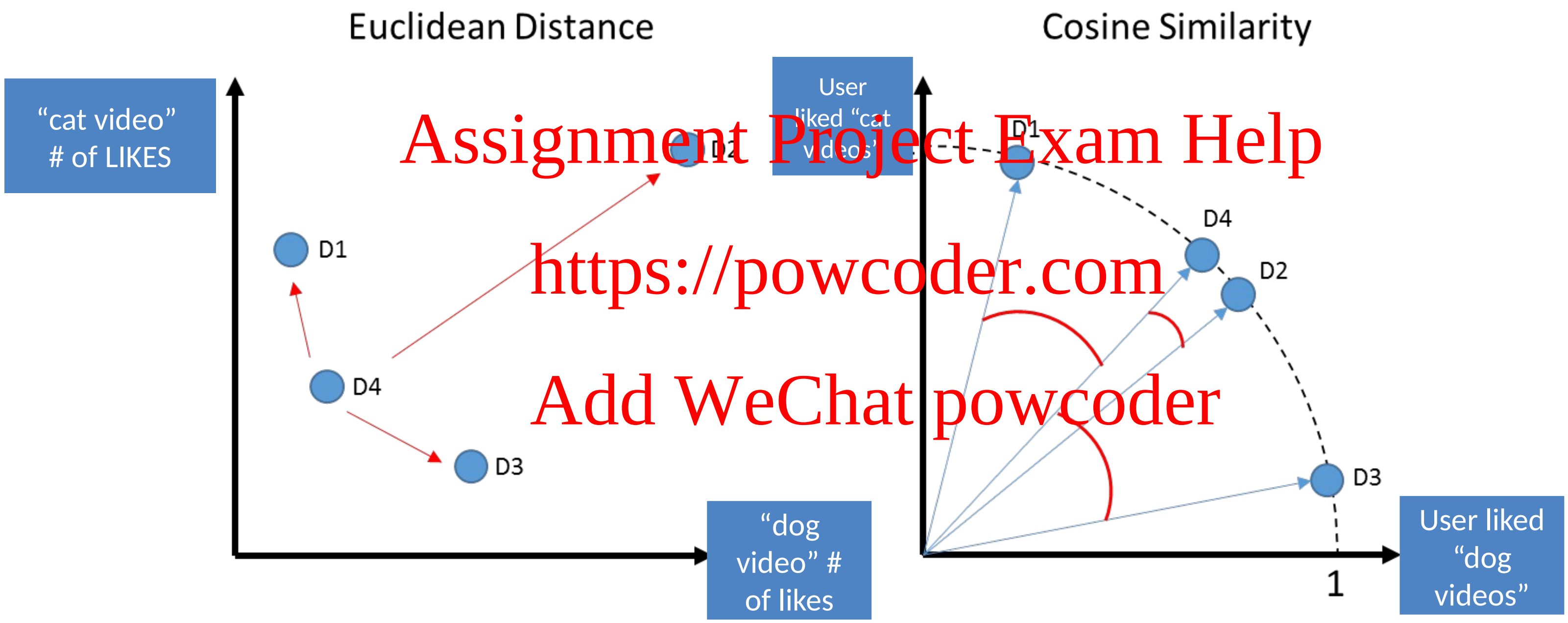
- Like nearest-neighbor algorithm
- But Euclidean distance does not do well
- Correlation proximity does better (Pearson)
- For each user pair, find the co-rated items, calculate correlation between the vectors of their ratings for those items

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

$$\mathrm{Corr}(U_1, U_2) = \frac{\sum (r_{1,i} - \bar{r}_1)(r_{2,i} - \bar{r}_2)}{\sqrt{\sum (r_{1,i} - \bar{r}_1)^2}\sqrt{\sum (r_{2,i} - \bar{r}_2)^2}}$$

MEH?!

# Proximity Measure - Cosine Similarity





Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder


*K-Means Clustering*

Outliers affect algo.

*Spherical K-Means*

Outliers have no affect.

HARVARD UNIVERSITY

# Major Challenge to Collaborative Filtering

COLD START

For users with just one item, or items with just one neighbor, neither cosine similarity nor correlation produces useful metric.  What do you do with new users?

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

# Open B_CollaborativeFiltering.R


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

# Summary – Collaborative Filtering

- User-based – for a new user, find other users who share his/her preferences, recommend the highest-rated item that new user does not have.

- Item-based – for a new user considering an item, find another item that is most similar in terms of user preferences represented as previous transactions.
  - Ability to calculate item-item correlations in advance greatly speeds up the algorithm

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

# Association Rules vs. Collaborative Filtering

- AR: focus entirely on frequent (popular) item combinations. Data rows are single transactions. Ignores user dimension. Often used in displays (what goes with what).
- CF: focus is on user preferences. Data rows are user purchases or ratings over time for a user or a particular item. Can capture "long tail" of user preferences – useful for recommendations involving unusual or a large number of items

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder



HARVARD UNIVERSITY