CSE-454: Data Mining

Hands-on Data Mining

Lec Md. Jakaria
Department of CSE
Military Institute of Science & Technology



Affinity Analysis

Affinity analysis is a type of data mining that gives similarity between samples (objects). This could be the similarity between the following:

- 1. Users on a website, in order to provide varied services or targeted advertising
- 2. Items to sell to those users, in order to provide recommended movies or products
- 3. Human genes, in order to find people that share the same ancestors



Product Recommendations

We are going to focus on a basic product recommendation service which is based on the following idea:

- ▶ When two items are historically purchased together, they are more likely to be purchased together in the future.
- A very simple algorithm for this type of product recommendation algorithm is to simply find any historical case where a user has brought an item and to recommend other items that the historical user brought.

3/6



Product Recommendations

To simplify the coding, we will consider only two items at a time. As an example, people may buy bread and milk at the same time at the supermarket.

If a person buys product X, then they are likely to purchase product Y

= 4/6 = 2000



Loading the dataset with NumPy

The dataset we are going to use for this example is a NumPy two-dimensional array

5/6



Loading the dataset with NumPy

The dataset we are going to use for this example is a NumPy two-dimensional array

The array looks like a table, with rows representing different samples and columns representing different features.

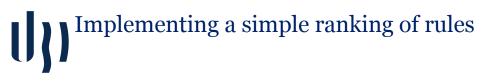


Loading the dataset with NumPy

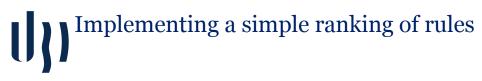
The dataset we are going to use for this example is a NumPy two-dimensional array

The array looks like a table, with rows representing different samples and columns representing different features.

The row represents items purchased in the a transaction. Each column (vertical row) represents each of the items. They are bread, milk, cheese, apples, and bananas, respectively.



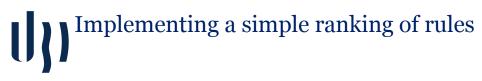
We wish to find rules of the type *If a person buys product X, then they are likely to purchase product Y.*



We wish to find rules of the type If a person buys product X, then they are likely to purchase product Y.

Rules of this type can be measured in many ways, of which we will focus on two: **support** and **confidence.**

<u>6/6</u> ■ ♥990



We wish to find rules of the type *If a person buys product X, then they are likely to purchase product Y.*

Rules of this type can be measured in many ways, of which we will focus on two: **support** and **confidence.**

1. Support is the number of times that a rule occurs in a dataset

Implementing a simple ranking of rules

We wish to find rules of the type *If a person buys product X, then they are likely to purchase product Y.*

Rules of this type can be measured in many ways, of which we will focus on two: **support** and **confidence.**

- 1. Support is the number of times that a rule occurs in a dataset
- 2. Confidence measures how accurate they are when they can be used.