Chapter 2, Data Mining Techniques used in Recommender Systems

How to solve a data analysis problem

- Take a problem
- Identify data sources & Acquire data
- Data Cleansing
- Visualize results & Do predictors
- Model Validation
- Build Statistical Model
- Exploratory Analysis

Euclidean Distance \( (x, y) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} \)

similarity = \( \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} \)

\( \rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \)
CLUSPLOT( iris )

Component 1
These two components explain 95.02% of the point variability.

Reduction in cost for values of 'k'

Within-cluster sum of squares

Clusters
Petal.Length < 2.45

setosa

Petal.Length < 4.85

Sepal.Length < 5.15

versicolor

Petal.Width < 1.75

versicolor

versicolor

virginica

> summary(model)

    var     rel.inf
Petal.Length  Petal.Length  67.440852
Petal.Width   Petal.Width   24.942084
Sepal.Width   Sepal.Width   7.017065
Sepal.Length  Sepal.Length  0.000000
sensitivity or true positive rate (TPR)

\[ TPR = \frac{TP}{P} = \frac{TP}{(TP + FN)} \]

specificity (SPC) or true negative rate (TNR)

\[ SPC = \frac{TN}{N} = \frac{TN}{(TN + FP)} \]

precision or positive predictive value (PPV)

\[ PPV = \frac{TP}{(TP + FP)} \]

Chapter 3, Recommender Systems
Heatmap of the first rows and columns

Dimensions: 10 x 15

Heatmap of the top users and movies

Dimensions: 10 x 17
Heatmap of the top users and movies

Dimensions: 12 x 7

Heatmap of the top users and movies

Dimensions: 27 x 17
Heatmap of the top users and movies

Dimensions: 27 x 17

Heatmap of the first rows and columns

Dimensions: 20 x 20
Distribution of the number of items for UBCF

Distribution of movies by user

\[ \text{distance}(item_1, item_2) = \frac{|item_1 \cap item_2|}{|item_1 \cup item_2|} \]
Chapter 4, Evaluating the Recommender Systems

unknown items by the users

rowCounts(getData(eval_sets, "unknown"))

Number of repetitions in the training set

\[ \text{distance}(user_1, user_2) = \frac{\text{user}_1 \cap \text{user}_2}{\text{user}_1 \cup \text{user}_2} \]
Distribution of movies per user

Distribution of the RMSE by user
Chapter 5, Case Study - Building Your Own Recommendation Engine

Binary rating matrix

Dimensions: 50 x 50

Distribution of the number of users

n_users

count

0 500 1000 1500