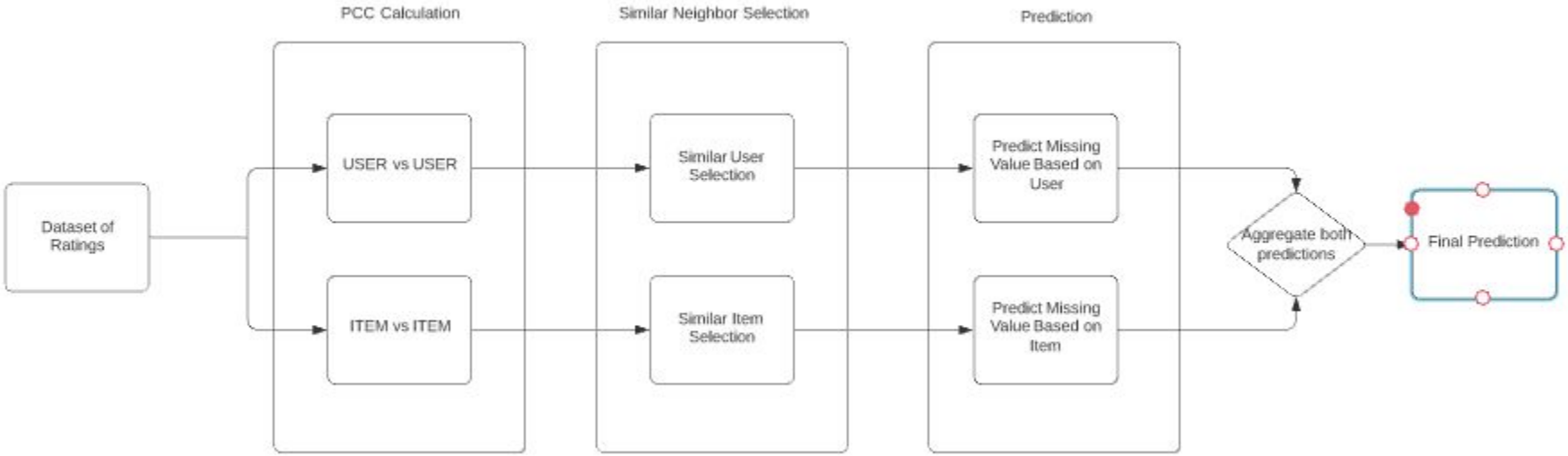
# Assignment 3
# Practical Data Science

By: Vijit Kumar (s3799493)

# Introduction

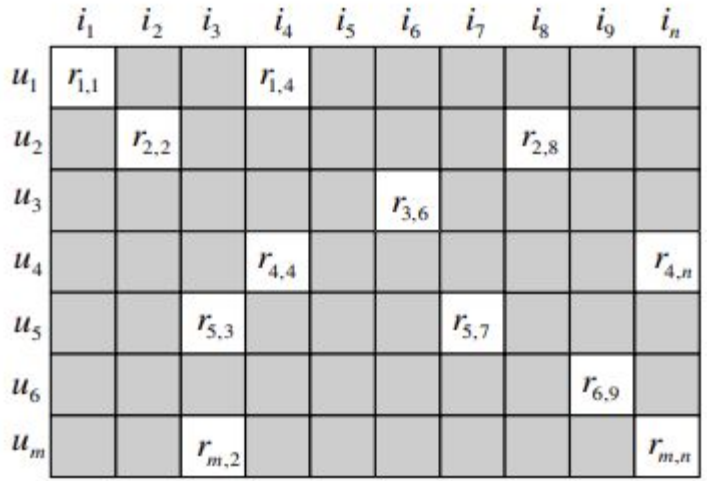
- Missing rating values
- Collaborative Filtering


# General Overview



# Dataset

- m number of users
- n number of items
- r represents the rating.



# Pearson Correlation Coefficient (PCC)

- User-User correlation matrix
- Item-Item correlation matrix

# PCC Similarity Equation

User-User Based

$$Sim(a,u) = \frac{\sum_{i \in I(a) \cap I(u)} (r_{a,i} - \overline{r}_a) \cdot (r_{u,i} - \overline{r}_u)}{\sqrt{\sum_{i \in I(a) \cap I(u)} (r_{a,i} - \overline{r}_a)^2} \cdot \sqrt{\sum_{i \in I(a) \cap I(u)} (r_{u,i} - \overline{r}_u)^2}},$$

Item-Item Based

$$Sim(i,j) = \frac{\sum_{u \in U(i) \cap U(j)} (r_{u,i} - \overline{r}_i) \cdot (r_{u,j} - \overline{r}_j)}{\sqrt{\sum_{u \in U(i) \cap U(j)} (r_{u,i} - \overline{r}_i)^2} \cdot \sqrt{\sum_{u \in U(i) \cap U(j)} (r_{u,j} - \overline{r}_j)^2}},$$

# PCC Similarity Equation with significance weighting

$$Sim'(a,u) = \frac{Min(|I_a \cap I_u|, \gamma)}{\gamma} \cdot Sim(a,u),$$

User-User based

$$Sim'(i,j) = \frac{Min(|U_i \cap U_j|, \delta)}{\delta} \cdot Sim(i,j),$$

Item-Item based

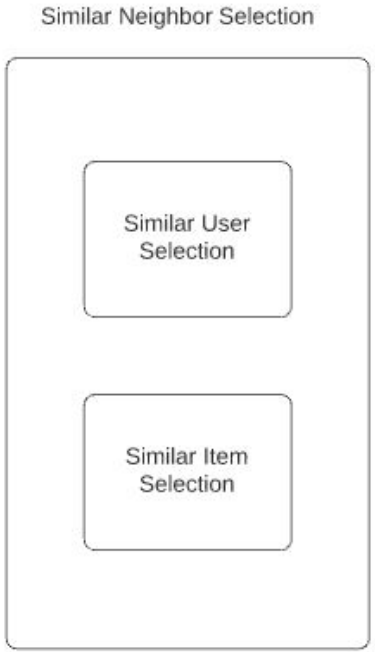
# Similar Neighbor Selection

Similar users selection with ITA parameter

$$S(u) = \{u_a | Sim'(u_a, u) > \eta, u_a \neq u\},$$

Similar item selection with THETA parameter

$$S(i) = \{i_k | Sim'(i_k, i) > \theta, i_k \neq i\},$$



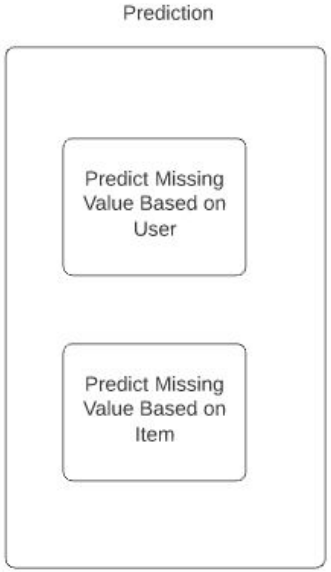
# Missing Data Prediction

User based prediction component

$$P(r_{u,i}) = \overline{u} + \frac{\sum_{u_a \in S(u)} Sim'(u_a, u) \cdot (r_{u_a,i} - \overline{u}_a)}{\sum_{u_a \in S(u)} Sim'(u_a, u)}.$$

Item based prediction component

$$P(r_{u,i}) = \overline{i} + \frac{\sum_{i_k \in S(i)} Sim'(i_k, i) \cdot (r_{u,i_k} - \overline{i}_k)}{\sum_{i_k \in S(i)} Sim'(i_k, i)}.$$



# Final Prediction Case 1

- LAMBDA parameter to aggregate the final prediction.
- Similar users are present during neighbor selection
- Similar items are present during neighbor selection

$$P(r_{u,i}) = \lambda \times (\overline{u} + \frac{\sum_{u_a \in S(u)} Sim'(u_a, u) \cdot (r_{u_a,i} - \overline{u}_a)}{\sum_{u_a \in S(u)} Sim'(u_a, u)}) +$$

$$(1 - \lambda) \times (\overline{i} + \frac{\sum_{i_k \in S(i)} Sim'(i_k, i) \cdot (r_{u,i_k} - \overline{i}_k)}{\sum_{i_k \in S(i)} Sim'(i_k, i)}),$$

# Final Prediction Case 2

- No LAMBDA parameter
- Only Similar users are present in neighbor selection

Only User based component will be used

$$P(r_{u,i}) = \overline{u} + \frac{\sum_{u_a \in S(u)} Sim'(u_a, u) \cdot (r_{u_a,i} - \overline{u}_a)}{\sum_{u_a \in S(u)} Sim'(u_a, u)}.$$

# Final Prediction Case 3

Only item based component will be used

- No LAMBDA parameter
- Only Similar items are present in neighbor selection

$$P(r_{u,i}) = \bar{i} + \frac{\sum_{i_k \in S(i)} Sim'(i_k, i) \cdot (r_{u,i_k} - \bar{i}_k)}{\sum_{i_k \in S(i)} Sim'(i_k, i)}.$$

# Final Prediction Case 4

- No Similar users are selected
- No similar items are selected
- There is no prediction the rating remains 0.

# Final output

| | $i_1$ | $i_2$ | $i_3$ | $i_4$ | $i_5$ | $i_6$ | $i_7$ | $i_8$ | $i_9$ | $i_n$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $u_1$ | $r_{1,1}$ | | | $r_{1,4}$ | | | | | | |
| $u_2$ | | $r_{2,2}$ | | | | | | $r_{2,8}$ | | |
| $u_3$ | | | | | | $r_{3,6}$ | | | | |
| $u_4$ | | | | $r_{4,4}$ | | | | | | $r_{4,n}$ |
| $u_5$ | | | $r_{5,3}$ | | | | $r_{5,7}$ | | | |
| $u_6$ | | | | | | | | | $r_{6,9}$ | |
| $u_m$ | | | $r_{m,2}$ | | | | | | | $r_{m,n}$ |

Initial Matrix

| | $i_1$ | $i_2$ | $i_3$ | $i_4$ | $i_5$ | $i_6$ | $i_7$ | $i_8$ | $i_9$ | $i_n$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $u_1$ | $r_{1,1}$ | 0 | $\hat{r}_{1,3}$ | $r_{1,4}$ | 0 | $\hat{r}_{1,6}$ | 0 | $\hat{r}_{1,8}$ | $\hat{r}_{1,9}$ | 0 |
| $u_2$ | 0 | $r_{2,2}$ | 0 | $\hat{r}_{2,4}$ | $\hat{r}_{2,5}$ | 0 | $\hat{r}_{2,7}$ | $r_{2,8}$ | 0 | $\hat{r}_{2,n}$ |
| $u_3$ | $\hat{r}_{3,1}$ | 0 | $\hat{r}_{3,3}$ | $\hat{r}_{3,4}$ | $\hat{r}_{3,5}$ | $r_{3,6}$ | 0 | $\hat{r}_{3,8}$ | $\hat{r}_{3,9}$ | 0 |
| $u_4$ | $\hat{r}_{4,1}$ | $\hat{r}_{4,2}$ | 0 | $r_{4,4}$ | $\hat{r}_{4,5}$ | $\hat{r}_{4,6}$ | $\hat{r}_{4,7}$ | 0 | $\hat{r}_{4,9}$ | $r_{4,n}$ |
| $u_5$ | $\hat{r}_{5,1}$ | $\hat{r}_{5,2}$ | $r_{5,3}$ | 0 | $\hat{r}_{5,5}$ | 0 | $r_{5,7}$ | $\hat{r}_{5,8}$ | $\hat{r}_{5,9}$ | $\hat{r}_{5,n}$ |
| $u_6$ | $\hat{r}_{6,1}$ | $\hat{r}_{6,2}$ | 0 | $\hat{r}_{6,4}$ | $\hat{r}_{6,5}$ | $\hat{r}_{6,6}$ | $\hat{r}_{6,7}$ | 0 | $r_{6,9}$ | $\hat{r}_{6,n}$ |
| $u_m$ | $\hat{r}_{m,1}$ | 0 | $r_{m,2}$ | $\hat{r}_{m,4}$ | 0 | $\hat{r}_{m,6}$ | 0 | $\hat{r}_{m,8}$ | $\hat{r}_{m,9}$ | $r_{m,n}$ |

Final Matrix