\[ r(u,i) = \text{rating of } i^{th} \text{ item by user } u \]

\[ I_u = \text{set of items rated by user } u \]

\[ I_{u,v} = \text{set of items rated by both users } u \text{ and } v \]

\[ r_u^{avg} = \left( \frac{1}{|I_u|} \right) \sum_{i \in I_u} r(u,i) \]

average rating by user \( u \)

\[ \sum_{i \in I_{u,v}} (r(u,i) - r_u^{avg}) (r(v,i) - r_v^{avg}) \]

\[ \text{sim}(u,v) = \frac{\sum_{i \in I_{u,v}} (r(u,i) - r_u^{avg}) (r(v,i) - r_v^{avg})}{\left( \sum_{i \in I_{u,v}} (r(u,i) - r_u^{avg})^2 \sum_{i \in I_{u,v}} (r(v,i) - r_v^{avg})^2 \right)^{\frac{1}{2}}} \]

similarity between users \( u \) and \( v \)

(Pearson correlation coefficient)

\[ \sum_{v \in S} \text{sim}(u,v) * (r(v,i) - r_v^{avg}) \]

\[ r_{pred}(u,i) = r_u^{avg} + \frac{\sum_{v \in S} \text{sim}(u,v) * (r(v,i) - r_v^{avg})}{\sum_{v \in S} | \text{sim}(u,v) |} \]

predicted rating of \( i^{th} \) item by user \( u \)

where \( S \) is either the set of all users other than \( u \) or a set of “most similar users” to \( u \).

For Problem Set 4, take \( S \) to be all users other than \( u \).