Pearson Correlation Coefficient: Issues

- Underlying assumption is that users dislike what they rated below average
- This is not true in practice (we rate only what we liked or highly disliked)
- The correlation *flattens* in case of uniformly distributed ratings

Deviation from average rating on shared items

$$sim(a,b) = \frac{\sum_{p \in P}(r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P}(r_{a,p} - \bar{r}_a)^2}\sqrt{\sum_{p \in P}(r_{b,p} - \bar{r}_b)^2} + \varepsilon}$$

!!! Will be zero in case of uniform rating !!!

User-based Collaborative Filtering: Prediction

- Let's use a different prediction function that is **mean-centered** in order to remove bias:

$$R_U = \overline{r_a} + \frac{\sum_{b \in N} sim(a,b) * (r_{b,p} - \overline{r_b})}{\sum_{b \in N} sim(a,b)}$$

Item-based Collaborative Filtering: Example

- For **User 3**, we need to predict ratings for **Item 1** and **Item 6**

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Mean
User 1	7	6	7	4	5	4	5.5
User 2	6	7	?	4	3	4	4.8
User 3	?	3	3	1	1	?	2
User 4	1	2	2	3	3	4	2.5
User 5	1	?	1	2	3	3	2

Item-based Collaborative Filtering: Example

- Although we can use any similarity measures discussed previously but we are going to use Adjusted Cosine similarity for this example
  - *It is Cosine similarity that is mean-adjusted*

$$sim(\vec{a}, \vec{b}) = \frac{\sum_{u \in U}(r_{u,a} - \overline{r_u})(r_{u,b} - \overline{r_u})}{\sqrt{\sum_{u \in U}(r_{u,a} - \overline{r_u})^2}\sqrt{\sum_{u \in U}(r_{u,b} - \overline{r_u})^2}}$$

Item-based Collaborative Filtering: Example

- For **User 3**, we need to predict ratings for **Item 1** and **Item 6**:

$$\text{Adj Cosine}(I1, I3) = \frac{(1.5*1.5)+(-1.5*-0.5)+(-1*-1)}{\sqrt{1.5^2+(-1.5)^2+(-1)^2}\,.\,\sqrt{1.5^2+(-0.5)^2+(-1)^2}} = 0.912$$

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	1.5	0.5	1.5	-1.5	0.5	-1.5
User 2	1.2	2.2	?	-0.8	-1.8	-0.8
User 3	?	1	1	-1	-1	?
User 4	-1.5	-0.5	-0.5	0.5	0.5	1.5
User 5	-1	?	-1	0	1	1

Item-based Collaborative Filtering: Example

- For **User 3**, we need to predict ratings for **Item 1** and **Item 6**:

$$\text{Adj Cosine}(I1, I3) = \frac{(1.5*1.5)+(-1.5*-0.5)+(-1*-1)}{\sqrt{1.5^2+(-1.5)^2+(-1)^2}\,.\,\sqrt{1.5^2+(-0.5)^2+(-1)^2}} = 0.912$$

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	1.5	0.5	1.5	-1.5	0.5	-1.5
User 2	1.2	2.2	?	-0.8	-1.8	-0.8
User 3	?	1	1	-1	-1	?
User 4	-1.5	-0.5	-0.5	0.5	0.5	1.5
User 5	-1	?	-1	0	1	1

- In the same manner, calculate similarity of I_1 with **all other items**

Item-based Collaborative Filtering: Prediction

- The final predicted ratings of **Item 1** for **User 3** is given by:

$$R_{31} = \frac{(3*0.735)+(3*0.912)}{|0.735|+|0.912|} = 3$$

Note: We assume 0.735 to be the similarity score between Item 1 and the second most similar item (Item 2). You can calculate it yourself and verify

Item-based Collaborative Filtering: Exercise

- **Task 1:** What will be the predicted rating for **Item 6** of **User 3**?

- **Task 2:** Predict all the missing ratings and find the top (unseen) item that can be recommended to each user