

COS 435: Collaborative filtering equations given in class 4/13/06

$r(u,i)$ = rating of i^{th} item by user u

I_u = set of items rated by user u

$I_{u,v}$ = set of items rated by both users u and v

$$r_u^{\text{avg}} = (1/|I_u|) * \sum_{i \text{ in } I_u} r(u,i) \quad \text{average rating by user } u$$

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

similarity between users u and v
(Pearson correlation coefficient)

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

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 .