Collaborative Filtering

- Use preferences of a group of people to make recommendations to other people.

- Build system for finding people who share tastes and for making automatic recommendations based on things that other people like.

  - E.g. Amazon

- Low tech way to get recommendations - ask your friends!
- But some friends have better taste than others i.e., their taste is more "similar" to yours

Collaborative Filtering Algorithms

- Search a large group of people & find a smaller set with tastes similar to yours.
- Looks at other thing they like & makes recommendations.

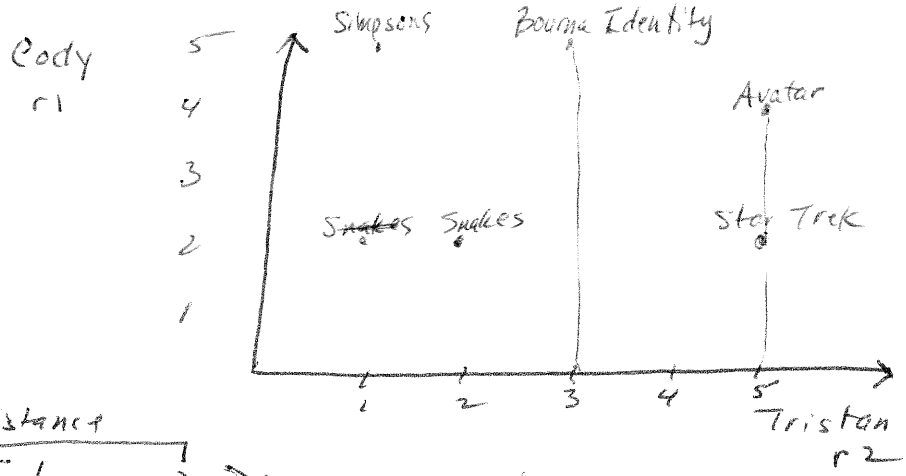
Note: "Using Collaborative filtering to weave an information tapestry," David Goldberg 1992, Xerox PARC ← coined term

Step 1 - Collect Preferences,
1a) Normalize ratings

Step 2 - Find Similar Users

2a) Euclidean distance score.

- Take items that two users have ranked in common and uses them as an axis in a chart (called a "preference" space).

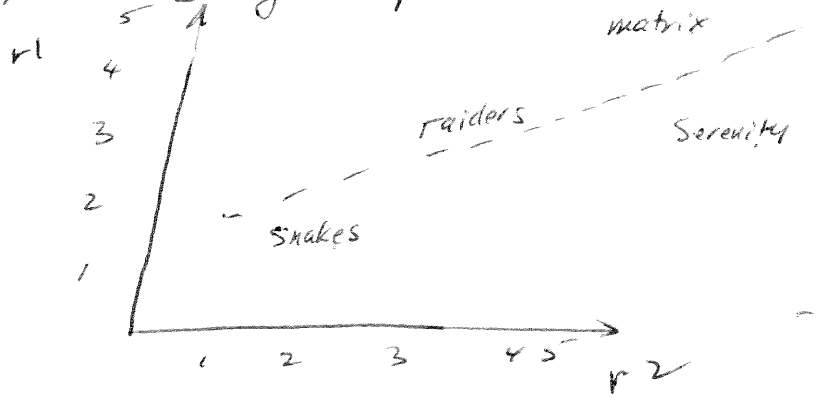


Euclidean distance

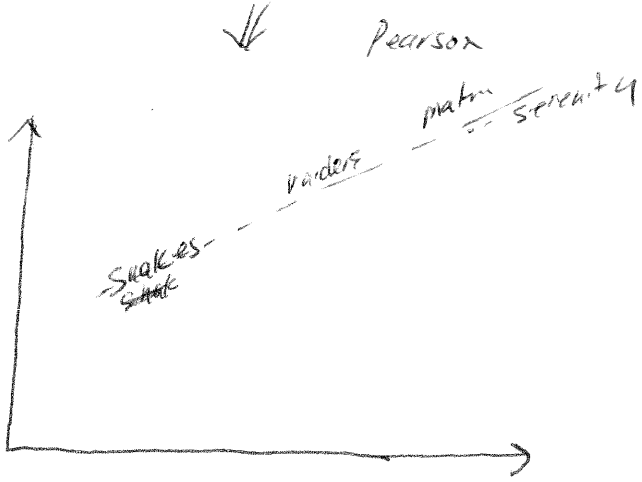
$$d = \sqrt{\sum_i (r1_i - r2_i)} \rightarrow sc = \frac{1}{(1 + \underbrace{sqrt(\underbrace{sum}_{\text{sum of squares}}(pow(r1 - r2)))}_{\text{sum of squares}})}$$

Pearson Correlation Score

- Measures how well two sets of data fit on a straight line.
- Normalizes grading curve



- Say r2 consistently rates movies lower



Sum all preferences

$$Sum1 = \sum_i (x_{1i})$$

$$Sum_2 = \sum_i x_{2i}$$

Sum of squares

$$sum1sq = pow\left(\sum_i x_{1i}, 2\right)$$

$$sum2sq = pow\left(\sum_i x_{2i}, 2, 2\right)$$

Sum of products

$$pSum = \sum x_{i1} * x_{i2}$$

$$= \frac{\sum_i (x1_i * x2_i) - \left(\sum_i x1_i * \sum_i x2_i\right)/n}{\sqrt{\left[\left(\sum_i x1_i^2\right) - \frac{\left(\sum_i x1_i\right)^2}{n}\right] * \left[\sum_i x2_i^2 - \frac{\left(\sum x2_i\right)^2}{n}\right]}}$$

$$\left[sum1sq - \frac{sum1^2}{n}\right] * \left[sum2sq - \frac{sum2^2}{n}\right]$$

Ranking other users (wrt yourself)

- Use similarity coefficient and rank order other users.

Recommending Items

1.) numerator
   - sum each potential recommendation rating from a particular user * that user's similarity to you

2) denominator
   - normalize by sum of all similarities.

$$\frac{(\text{sum}\ (r.\text{rating\_norm} * s.sc))}{\text{sum}\ (s.sc)}$$