Quiz 6 LSH & Recommendation System

1.

Increasing b

False negatives decrease

b=30, r = 5   b=20, r = 5

b=5, r = 5

b=15, r = 5

Probability

False positives increase
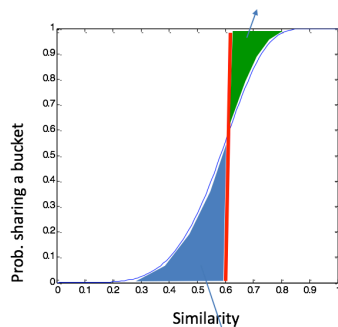
Similarity   Threshold

108
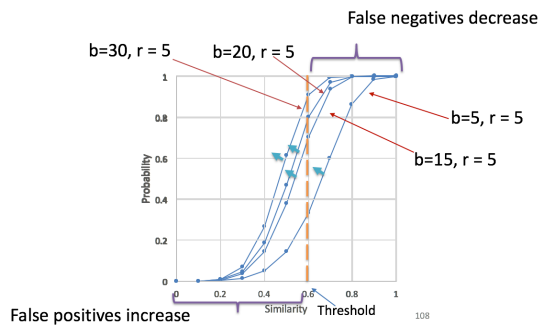
Increasing r

False negatives increase

b=20, r = 5

b=20, r = 10

b=20, r = 15

Probability

False positives decrease

Similarity   Threshold
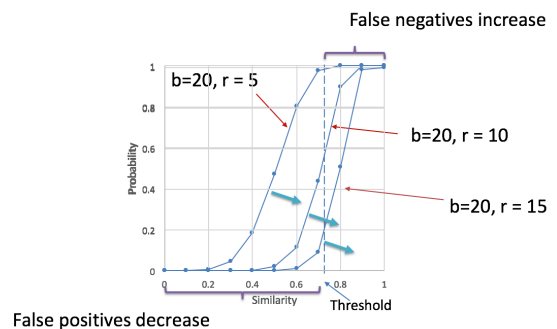
False negatives = area above the curve & to the right of line

Prob. sharing a bucket

Similarity

False positives = area below the curve & to the left of line   81

When s increases, false negative decrease and false positive increase.

2.

	Avatar	LOTR	Matrix	Pirates
Alice	1		0.2	
Bob		0.5		0.3
Carol	0.2		1	
David				0.4

For each user-item pair, value represents degree of preference of that user for that item (e.g., rating)

3. One possible example:

◆ **Example 9.2**

➢ Features of movies are set of actors and average rating

➢ Each movie has 5 actors, two in both movies

➢ Average ratings 3 and 4 (with unknown scaling factor alpha)

➢ **Must scale non-boolean components so they are not dominant or irrelevant**

$$
\begin{array}{ccccccccc}
0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 3\alpha \\
1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 4\alpha
\end{array}
$$

make sure each movie has 6 actors and 2 are in common. And average rating has a scaling factor.

4.

People usually buy things with huge popularity. But products in the long tail which have lower popularity could make more sales. They are just much less presented to customers due to the limited shelf space in traditional sales.