Nov 12

Collaborative filtering (Netflix)

Each user = a ranking on movies/shows on Netflix

Hypothesis: user 1 is "close" to user 2 if user 1's ranking is "close" to user 2's ranking.

Assumption: Each user ranks ALL movies/shows (unrealistic)

Input: A ranking $a_1, \dots, a_n$ (permutation on $\{1, \dots, n\}$)

Output: Number of inversions

$\underbrace{1, \dots, n}_{\text{"true" ranking}}$

Recall: (i,j) is an inversions

(1) $i < j$ AND (2) $a_i > a_j$

Ex1: $a = (1, \dots, n)$; # inversions $= 0$ ← lowest possible value

all pairs are NOT inversions

(if $i < j \Rightarrow a_i = i < j = a_j$)

$a_1, \dots, a_n$ being sorted $\equiv$ # inversions $= 0$

Ex2: $a = (n, \dots, 1)$; # inversions $= \binom{n}{2} = \frac{n(n-1)}{2}$

i j

$a_i > a_j$

all pairs ARE inversions

General "bad case"

Goal: Count # pairs (i,j) $1 \le i \le n'$ $1 \le j \le m$ s.t. $l_i > r_j$

Merge Count (a_L, a_R)

0. $c = 0$
1. $i, j = 1$
2. While $i \le n'$ and $j \le m$
   - If $l_i < r_j$
     - i++
   - else
     - c += $(n' - i + 1)$
     - j++
3. Return c