

Nov 12

# Collaborative filtering (Netflix)

Each user  $\equiv$  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  
Recall:  $(i, j)$  is an inversion if  $i < j$  AND  $a_i > a_j$

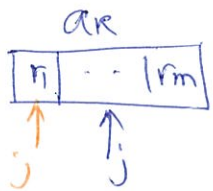
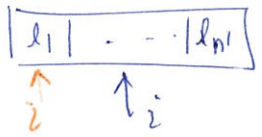
$1, \dots, n$   
"true" ranking

Ex 1:  $a = (1, \dots, n)$ ; #inversions = 0  $\leftarrow$  lowest possible value  
all pairs are NOT inversions

$a_1, \dots, a_n$  being sorted  $\equiv$  #inversions = 0  
(if  $i < j \Rightarrow a_i = i < a_j = j$ )

Ex 2:  $a = (n, \dots, 1)$ ; #inversions =  $\binom{n}{2} = \frac{n(n-1)}{2}$   
all pairs ARE inversions  
 $i < j$   
 $a_i > a_j$

General "bad case"



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

Merge Count  $(a_L, a_R)$

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