Music

RECOMMENDATIONS

There are quite a few digital music services available these days

last.fm BETA

MUSIC

There are quite a few digital music services available these days

PANDORA

Spotify

Users of these services would be delighted with the huge variety to choose from

But often find it difficult to

Sift through the variety

and identify things they would like

RECOMMENDATIONS HELP USERS

NAVIGATE THE MAZE OF THE MUSIC CATALOGUES

FIND WHAT THEY ARE LOOKING FOR

FIND ARTISTS THEY MIGHT LIKE, BUT DIDN'T KNOW OF

last.fm BETA

PANDORA

MUSIC

Spotify

RECOMMENDATIONS HELP USERS

NAVIGATE THE MAZE OF ONLINE STORES

FIND WHAT THEY ARE LOOKING FOR

FIND ARTISTS THEY MIGHT LIKE, BUT DIDN'T KNOW OF

RECOMMENDATIONS HELP THESE SERVICES

SOLVE THE PROBLEM OF DISCOVERY

HOW?

last.fm BETA

PANDORA

MUSIC

Spotify

HOW?

USING DATA

WHAT USERS BOUGHT

WHAT USERS BROWSED

WHAT USERS LISTENED TO

WHAT USERS RATED

→ RECOMMENDATION ENGINE →

TOP PICKS FOR YOU!!

IF YOU LIKE THIS, YOU'LL LOVE THAT!

RECOMMENDATION ENGINE

MOST RECOMMENDATION ENGINES
USE A TECHNIQUE CALLED

COLLABORATIVE FILTERING

COLLABORATIVE FILTERING

HOW DOES THAT WORK?

HOW DO YOU NORMALLY FIND

A MOVIE TO WATCH?

A RESTAURANT TO GO TO?

AN ARTIST TO CHECK OUT?

A BOOK TO READ?

ASK A FRIEND!

SOMEONE WHO LIKES

COLLABORATIVE FILTERING

HOW DOES THAT WORK?

THE BASIC PREMISE IS THAT

IF 2 USERS HAVE THE SAME OPINION ABOUT A BUNCH OF PRODUCTS

THEY ARE LIKELY TO HAVE THE SAME OPINION ABOUT OTHER PRODUCTS TOO

COLLABORATIVE FILTERING IS
A GENERAL TERM

FOR ANY ALGORITHM THAT
RELIES ONLY ON USER BEHAVIOR
(HISTORY, RATINGS, SIMILAR
USERS ETC)

COLLABORATIVE FILTERING IS A GENERAL TERM

FOR ANY ALGORITHM THAT RELIES ONLY ON USER BEHAVIOR (HISTORY, RATINGS, SIMILAR USERS ETC)

THIS IS AS OPPOSED TO CONTENT BASED FILTERING (WHICH USES PRODUCT ATTRIBUTES LIKE GENRE, DESCRIPTION ETC)

COLLABORATIVE FILTERING IS A GENERAL TERM

FOR ANY ALGORITHM THAT RELIES ONLY ON USER BEHAVIOR (HISTORY, RATINGS, SIMILAR USERS ETC)

CF ALGORITHMS NORMALLY PREDICT USERS' RATINGS FOR PRODUCTS THEY HAVEN'T YET RATED

RATING HERE IS A GENERAL TERM

IT CAN BE A RATING THE USER HAS EXPLICITLY GIVEN

EXPLICIT RATING

NETFLIX ASKS USERS TO RATE A MOVIE ONCE THEY HAVE WATCHED IT

IT CAN BE BASED ON A PREFERENCE THE USER HAS SOMEHOW INDICATED

IMPLICIT RATING

TIMES THE USER LISTENED TO AN ARTIST

THERE ARE MANY MANY DIFFERENT ALGORITHMS TO PERFORM COLLABORATIVE FILTERING

1 POPULAR TECHNIQUE IS

LATENT FACTOR ANALYSIS

LATENT FACTOR ANALYSIS

IDENTIFY HIDDEN FACTORS THAT INFLUENCE A USER'S RATING

SOMETIMES THE FACTORS MIGHT TURN OUT TO HAVE MEANING (LIKE GENRE OR OVERALL POPULARITY)

OTHER TIMES, THEY MIGHT BE ABSTRACT FACTORS WITH NO REAL LIFE MEANING

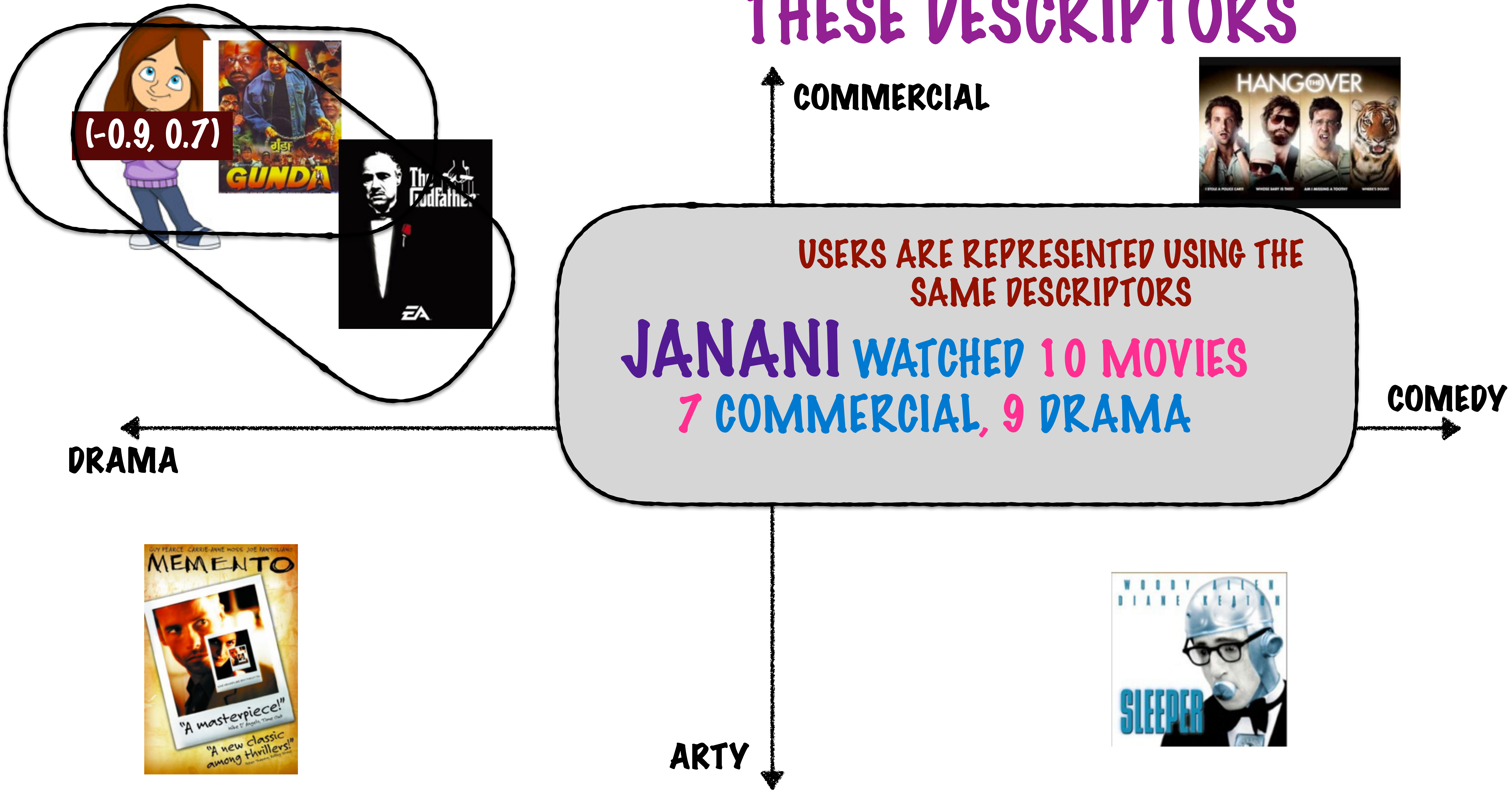
LATENT FACTOR ANALYSIS

LET'S SAY YOU WERE DOING THIS WITH MOVIES

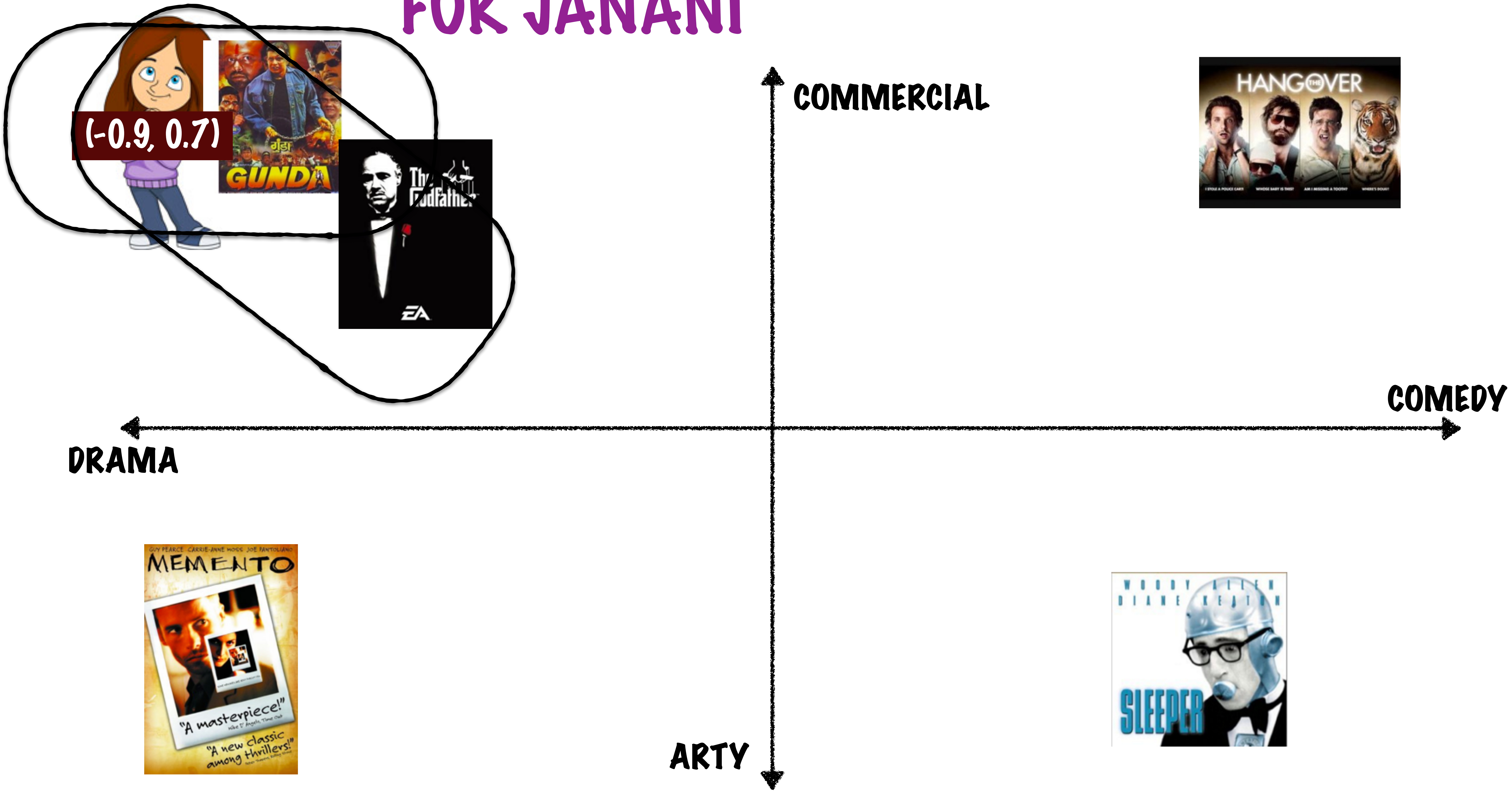
AND THE HIDDEN FACTORS ARE

1. COMMERCIAL APPEAL
2. DRAMATIC VS COMEDIC NATURE

MOVIES ARE REPRESENTED USING THESE DESCRIPTORS



RECOMMENDATIONS FOR JANANI



LATENT FACTOR COLLABORATIVE FILTERING

TO IDENTIFY HIDDEN FACTORS

YOU NEED A USER-PRODUCT-RATING MATRIX

	PROD 1	PROD 2	PROD 3	PROD 4	.. .. ..	PROD D
USER 1	4	-	4	-	-	-
USER 2	-	3	4	-	-	-
USER 3	5	3	2	-	-	5
USER 4	2	-	2	-	-	4
..	-	-	-	4	-	-
..	-	1	-	-	-	-
USER N	4	3	4	-	-	5

IT REPRESENTS USERS BY THEIR RATINGS FOR DIFFERENT PRODUCTS

LATENT FACTOR COLLABORATIVE FILTERING

IT REPRESENTS USERS BY THEIR RATINGS FOR DIFFERENT PRODUCTS

	PROD 1	PROD 2	PROD 3	PROD 4	.. .. ..	PROD D
USER 1	4	-	4	-	-	-
USER 2	-	3	4	-	-	-
USER 3	5	3	2	-	-	5
USER 4	2	-	2	-	-	4
:	-	-	-	4	-	-
:	-	1	-	-	-	-
USER N	4	3	4	-	-	5

BREAK THIS DOWN TO IDENTIFY THE HIDDEN FACTORS

	F1	F2	F3
USER 1			
USER 2			
USER 3			
USER 4			
:			
:			
USER N			

*

	PROD 1	PROD 2	PROD 3	PROD 4	.. .. ..	PROD D
F1						
F2						
F3						

LATENT FACTOR COLLABORATIVE FILTERING

R

USER-PROD RATING MATRIX

	PROD 1	PROD 2	PROD 3	PROD 4	..	.. .. PROD D
USER 1	4	-	4	-	-	-
USER 2	-	3	4	-	-	-
USER 3	5	3	2	-	-	5
USER 4	2	-	2	-	-	4
:	-	-	-	4	-	-
:	-	7	-	-	-	-
USER N	4	3	4	-	-	5

DECOMPOSED INTO

P

	F1	F2	F3
USER 1			
USER 2			
USER 3			
USER 4			
:			
:			
USER N			

USER-FACTOR MATRIX

EACH ROW IS A USER DESCRIBED BY THEIR INTEREST IN THE HIDDEN FACTORS

*

Q

	PROD 1	PROD 2	PROD 3	PROD 4	..	.. .. PROD D
F1						
F2						
F3						

PRODUCT-FACTOR MATRIX

EACH COLUMN IS A PRODUCT DESCRIBED BY ITS RELEVANCE TO THE HIDDEN FACTORS

LATENT FACTOR COLLABORATIVE FILTERING

R

	PROD 1	PROD 2	PROD 3	PROD 4	.. …	PROD D
USER 1	4	-	4	-	-	-
USER 2	-	3	4	-	-	-
USER 3	5	3	2	-	-	5
USER 4	2	-	**2**	-	-	4
..	-	-	-	4	-	-
..	-	7	-	-	-	-
USER N	4	3	4	-	-	5

USER-PROD RATING MATRIX

DECOMPOSED INTO

P

	F1	F2	F3
USER 1			
USER 2			
USER 3			
USER 4		p_4	
..			
USER N			

USER-FACTOR MATRIX

*

Q

	PROD 1	PROD 2	PROD 3	PROD 4	.. …	PROD D
F1						
F2			q_3			
F3						

PRODUCT-FACTOR MATRIX

EACH RATING HAS TO BE DECOMPOSED INTO 2 VECTORS

$$r_{4B} = p_4 \, . \, q_B$$

YOU CAN WRITE SUCH AN EQUATION FOR EACH RATING OF AN PROD i BY USER u

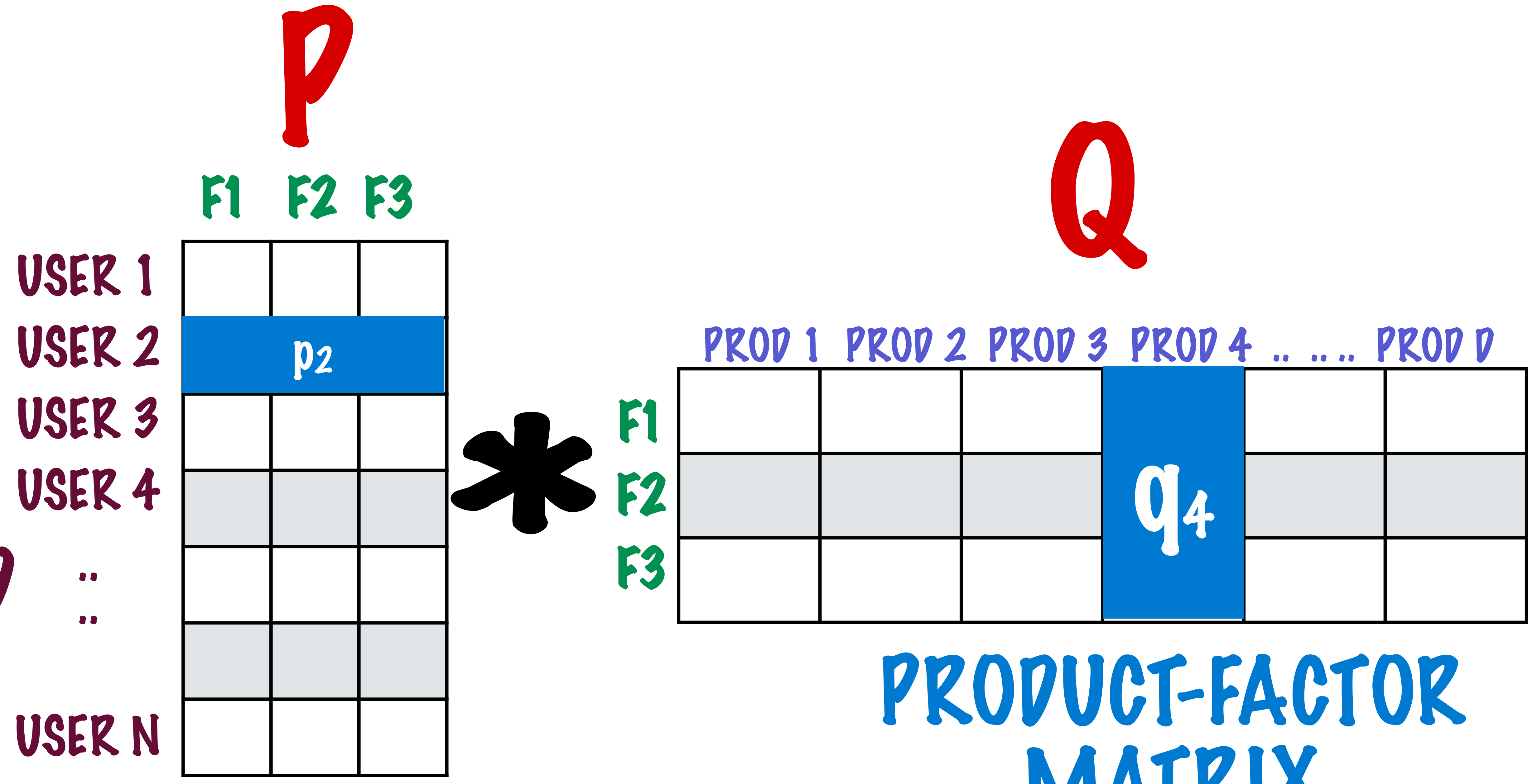
LATENT FACTOR COLLABORATIVE FILTERING

R

	PROD 1	PROD 2	PROD 3	PROD 4	.. .. ..	PROD D
USER 1	4	-	4	-	-	-
USER 2	-	3	4	?	-	-
USER 3	5	3	2	-	-	5
USER 4	2	-	2	-	-	4
::	-	-	-	4	-	-
	-	1	-	-	-	-
USER N	4	3	4	-	-	5

USER-PROD RATING MATRIX

DECOMPOSED INTO



USER-FACTOR MATRIX

PRODUCT-FACTOR MATRIX

SOLVE THIS SET OF EQUATIONS FOR THE SET OF RATINGS WHICH EXIST

$$r_{ui} = p_u . q_i$$

USE THE RESULTING P'S AND Q'S TO FIND THE RATING OF ANY USER FOR ANY PRODUCT

LATENT FACTOR COLLABORATIVE FILTERING

SOLVE THIS SET OF EQUATIONS FOR THE SET OF RATINGS WHICH EXIST (TRAINING SET)

$$r_{ui} = p_u . q_i$$

ONCE WE'VE SOLVED THIS SET OF EQUATIONS WE CAN PREDICT WHAT RATING THE USER WOULD GIVE TO ANY PRODUCT

LATENT FACTOR COLLABORATIVE FILTERING

SOLVE THIS SET OF EQUATIONS FOR THE SET OF RATINGS WHICH EXIST (TRAINING SET)

$$r_{ui} = p_u . q_i$$

SORT THE PREDICTED RATINGS IN DESCENDING ORDER TO FIND THE TOP RECOMMENDATIONS FOR A USER

LATENT FACTOR COLLABORATIVE FILTERING

SOLVE THIS SET OF EQUATIONS
FOR THE SET OF RATINGS
WHICH EXIST (TRAINING SET)

$$r_{ui} = p_u . q_i$$

ALTERNATING LEAST
SQUARES

IS A TECHNIQUE TO SOLVE
THIS SET OF EQUATIONS

ALTERNATING LEAST SQUARES

ALS IS A STANDARD OPTIMIZATION TECHNIQUE THAT CAN BE APPLIED TO MANY PROBLEMS

SPARK'S MLLIB HAS A BUILT-IN CLASS FOR APPLYING ALS ON ANY USER-PRODUCT-RATING MATRIX

MLLIB COMPLETELY ABSTRACTS AWAY THE TECHNICAL IMPLEMENTATION DETAILS OF ALS