INCORPORATING SIDE INFORMATION INTO PROBABILISTIC MATRIX FACTORIZATION USING GAUSSIAN PROCESSES

Ryan Prescott Adams, George E. Dahl and Iain Murray
University of Toronto and University of Edinburgh

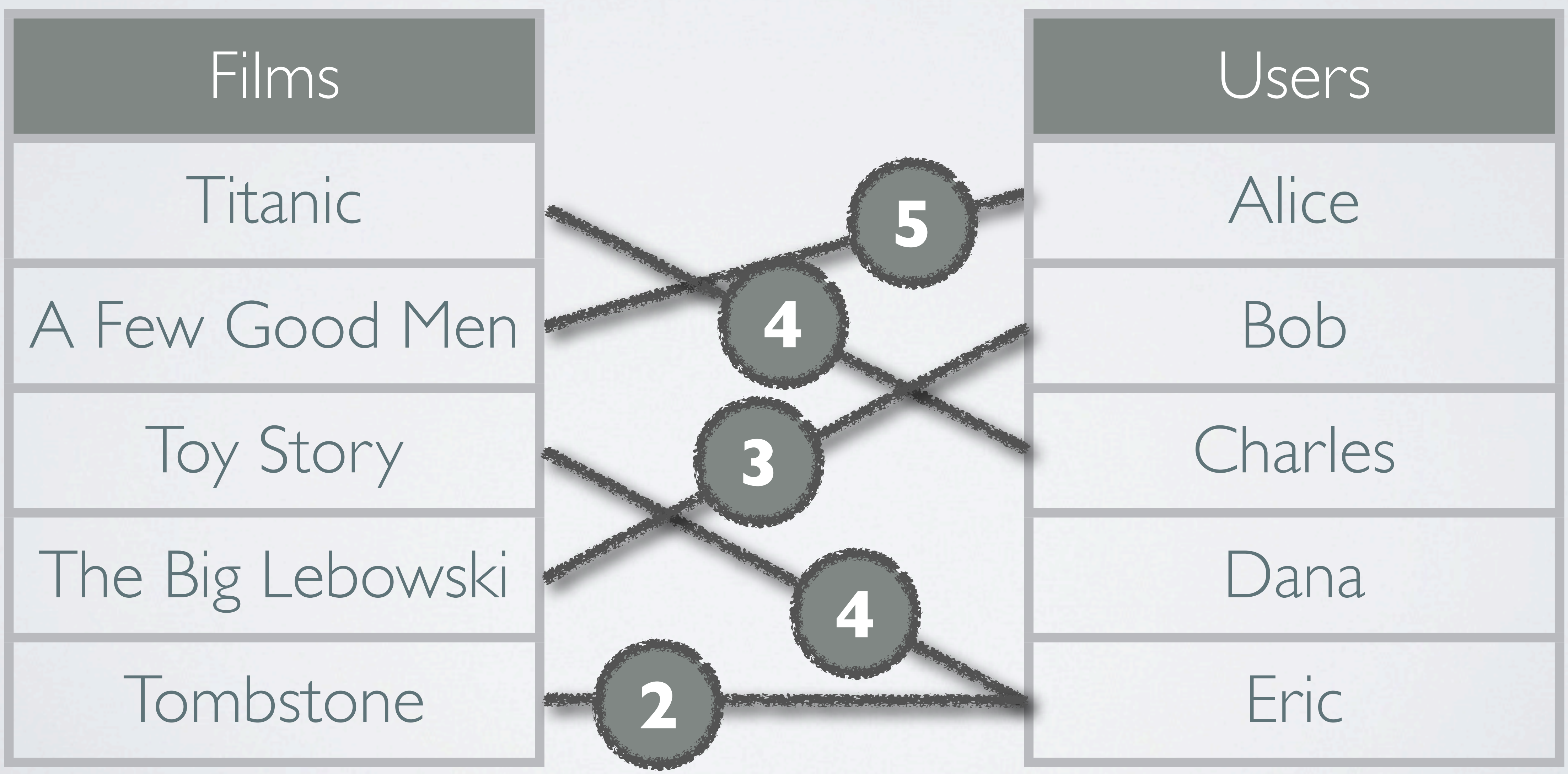
UAI: 9 July 2010

CIFAR

RELATIONAL DATA

Many data are well described as interactions between pairs of entities.

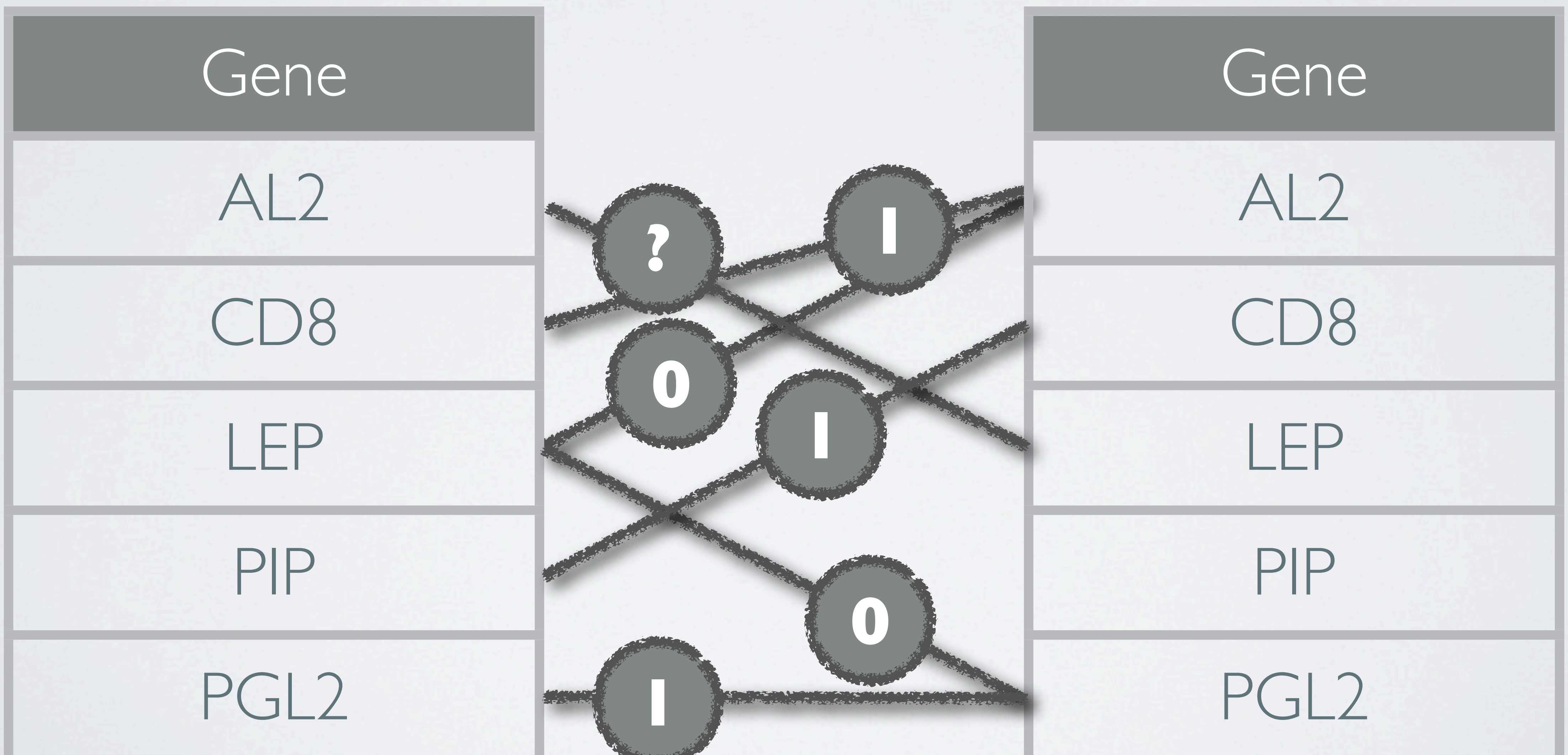
Example: Collaborative Filtering



RELATIONAL DATA

Many data are well described as interactions between pairs of entities.

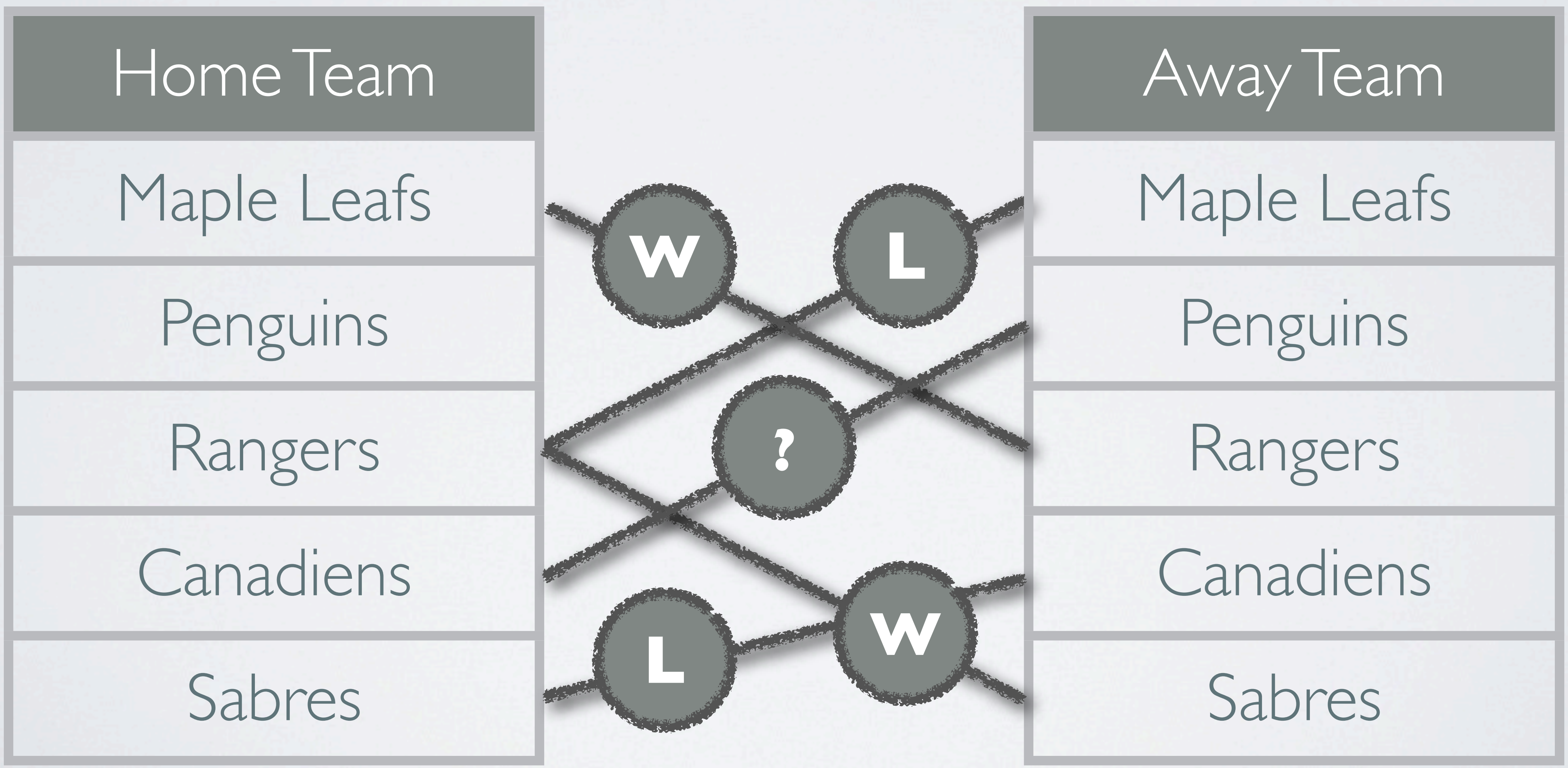
Example: Biological Pathway Analysis



RELATIONAL DATA

Many data are well described as interactions between pairs of entities.

Example: Modeling Sports



RELATIONAL DATA AS MATRICES

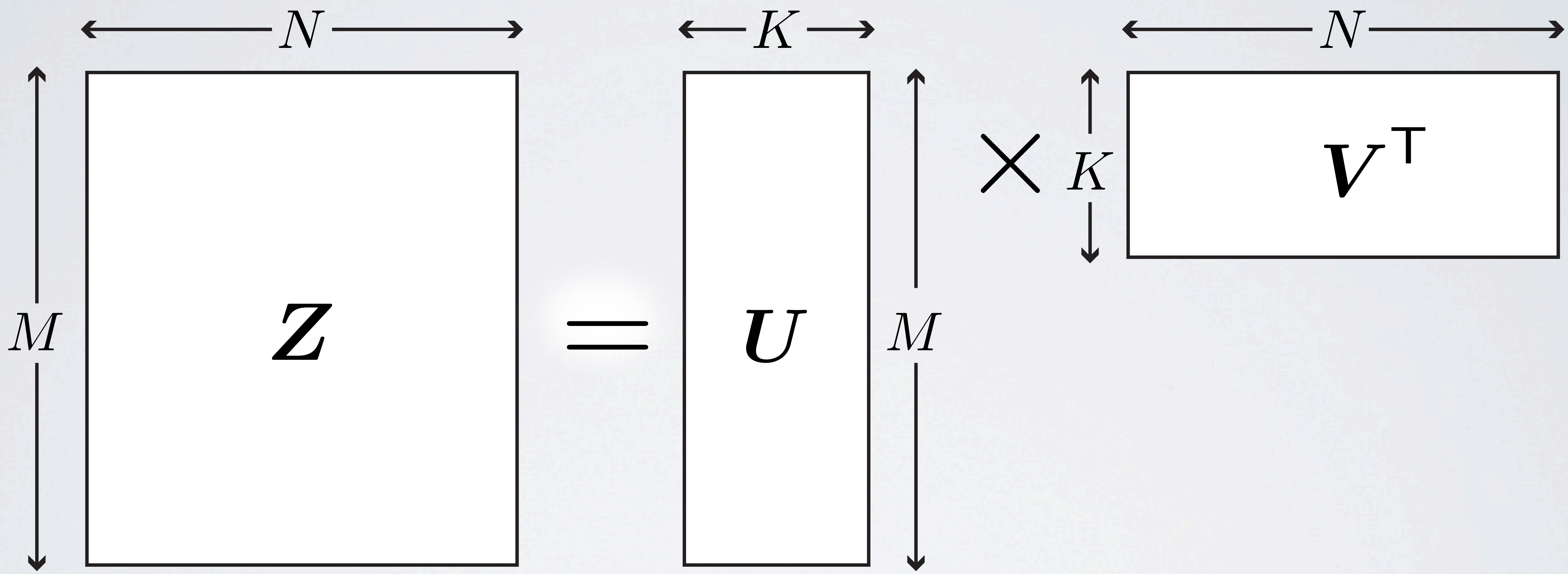
A useful view on relational data is to place in a matrix where the rows and columns correspond to the entities.

	Alice	Bob	Charles	Dana	Eric
Titanic			4		2
A Few Good Men	1		?	5	
Toy Story		3			4
The Big Lebowski	5	5		?	
Tombstone			4		2

INTRODUCING STRUCTURE

How to build dependencies into a relational matrix?

Consider the matrix to have low rank.



M: "Number of Movies" Z: Relational Data Matrix

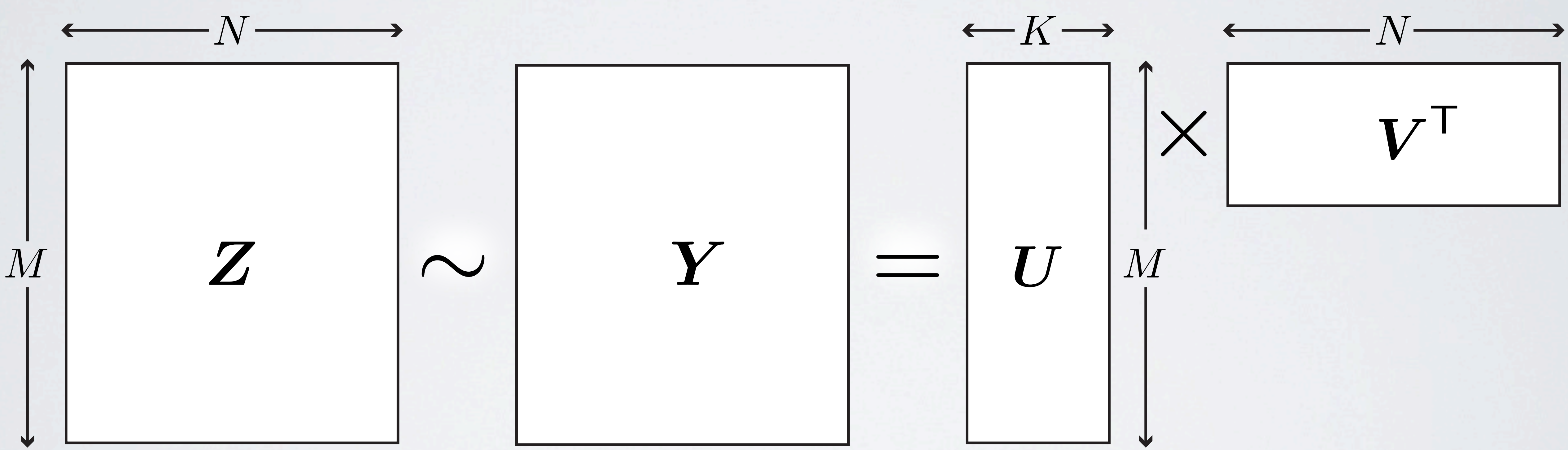
N: "Number of Users" U: Latent "Movie Features"

K: Something You Choose V: Latent "User Features"

PROBABILISTIC MATRIX FACTORIZATION

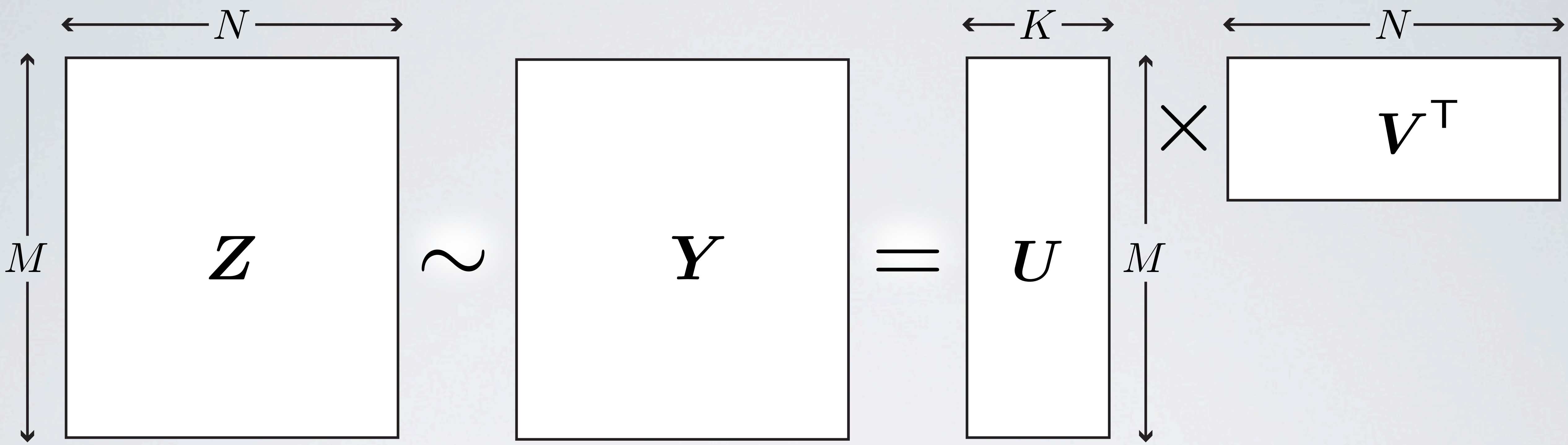
The observed matrix is some kind of "noisy" version of the low-rank matrix.

What "noisy" means will depend on your data.



The matrix Y is now unobserved.

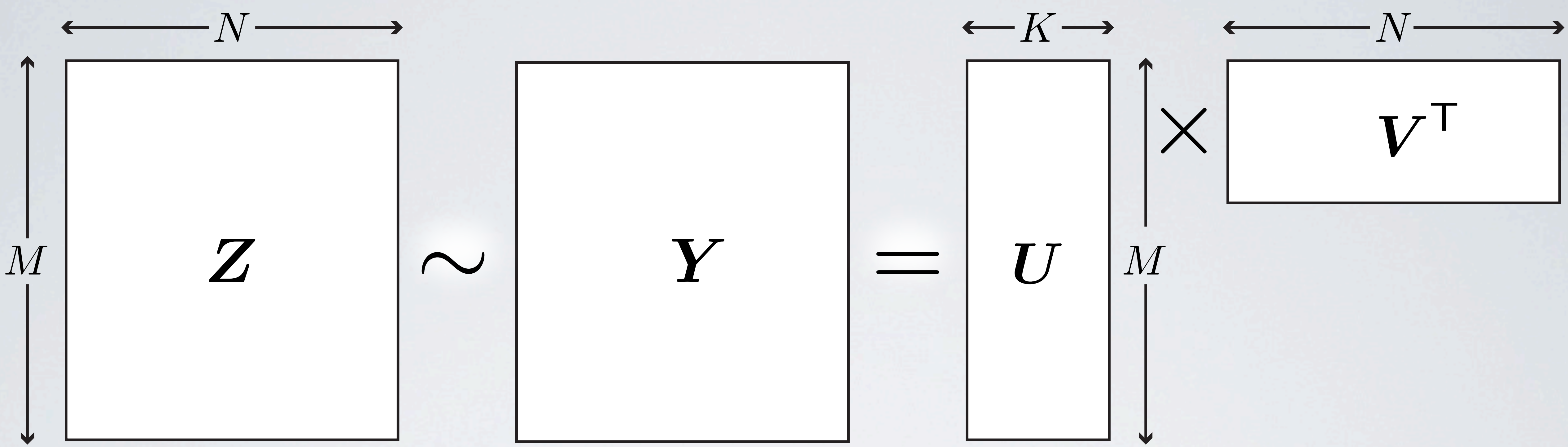
PROBABILISTIC MATRIX FACTORIZATION



Main Idea:

1. Make K much smaller than M or N.
2. Learn U and V from observed data.
3. Make predictions from learned U and V.

PROBABILISTIC MATRIX FACTORIZATION



$$Y_{m,n} = \boldsymbol{u}_m^{\mathsf{T}} \boldsymbol{v}_n$$

Features about row m,
e.g., "action-movie-ness"

Features about column n,
e.g., "affinity for action movies"

PMF FOR NBA BASKETBALL

FINAL 3:30 PM ET		1	2	3	4	Total
	BOS 93	18	30	24	21	93
	CLE 104	24	30	26	24	104

	BOS		CLE	
P	Allen	20	James	30
R	Rondo	8	Jamison	12
A	Rondo	6	James	7

Book Travel | Listen live with NBA Game Time — Complete Stats

FINAL 6:00 PM ET		1	2	3	4	Total
	PHI 91	21	28	24	18	91
	MIA 104	31	26	21	26	104

	PHI		MIA	
P	Kapono	17	Wade	38
R	Brand	10	Haslem	12
A	Iguodala	9	Arroyo	10

Book Travel | Watch live with NBA LEAGUE PASS or Listen live with NBA Game Time — Complete Stats

FINAL 6:00 PM ET		1	2	3	4	Total
	CHA 96	26	25	21	24	96
	ORL 89	25	28	15	21	89

	CHA		ORL	
P	Jackson	28	Howard	27
R	Ratliff	9	Howard	16
A	Felton	7	Carter	6

Book Travel | Watch live with NBA LEAGUE PASS or Listen live with NBA Game Time — Complete Stats

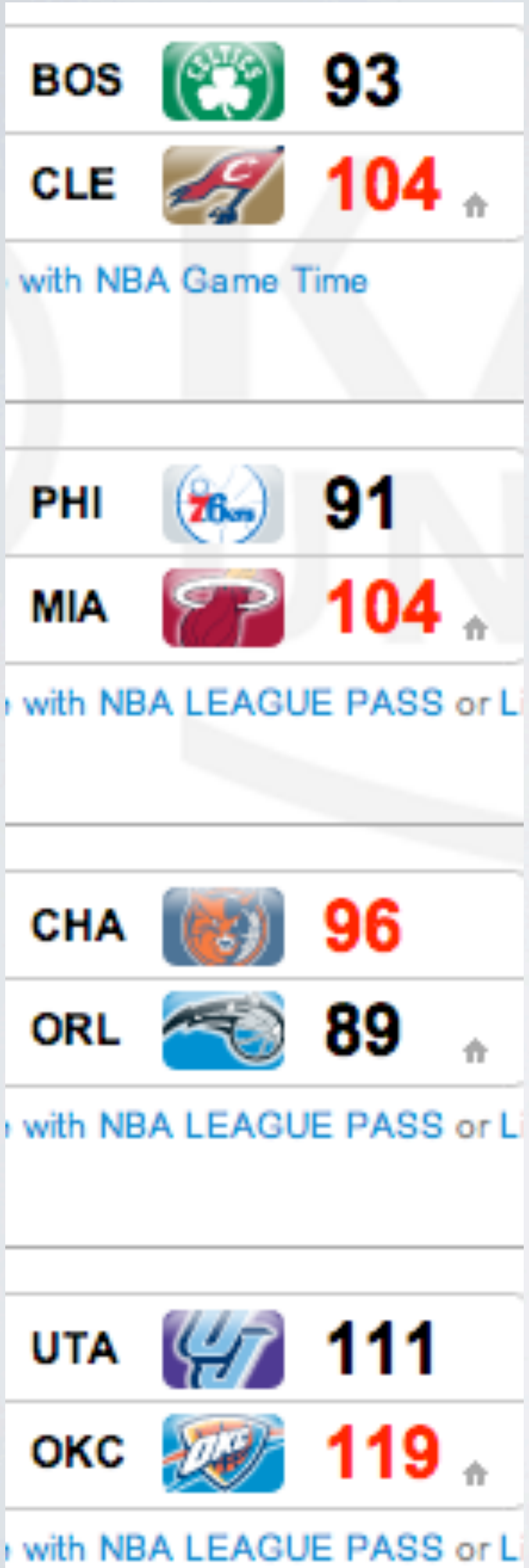
FINAL 7:00 PM ET		1	2	3	4	Total
	UTA 111	26	23	30	32	111
	OKC 119	29	30	31	29	119

	UTA		OKC	
P	Matthews	29	Durant	35
R	Boozer	11	Sefolosha	6
A	Williams	14	Westbrook	11

Book Travel | Watch live with NBA LEAGUE PASS or Listen live with NBA Game Time — Complete Stats

Source: http://www.nba.com/gameline/

PMF FOR NBA BASKETBALL

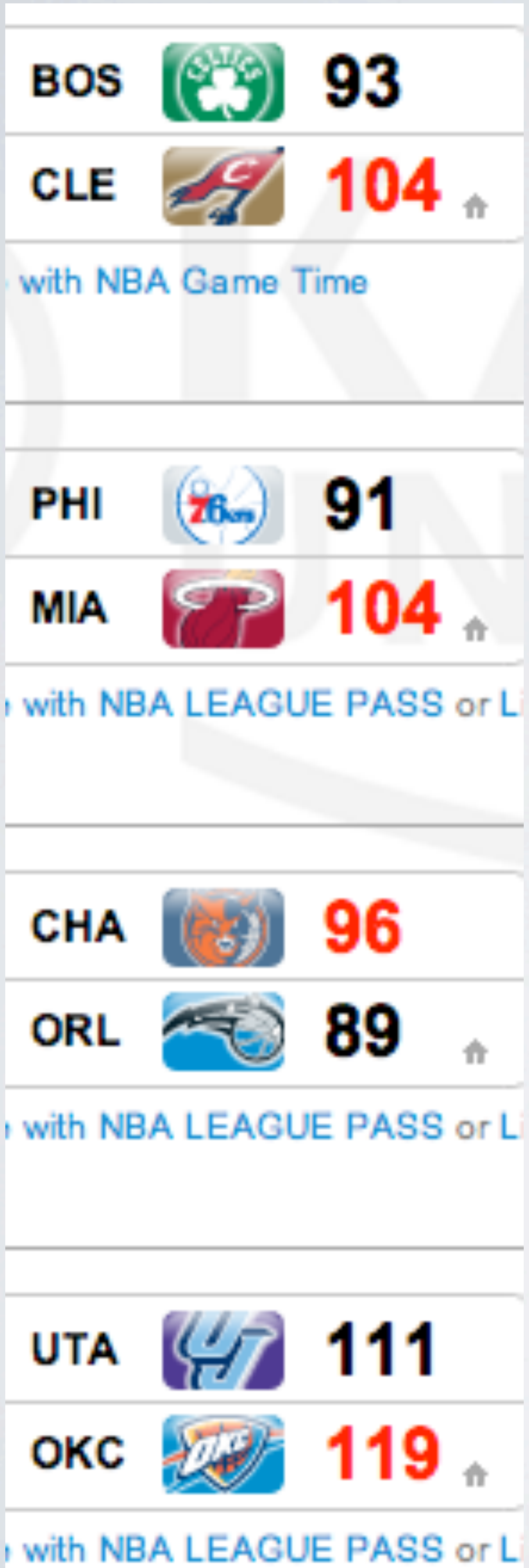


	BOS	CHA	CLE	MIA	OKC	ORL	PHI	UTA
BOS			93					
CHA						96		
CLE	104							
MIA							104	
OKC								119
ORL		89						
PHI				91				
UTA					111			

$Z_{m,n}$ = Score of team m against n.

$Z_{n,m}$ = Score of team n against m.

PMF FOR NBA BASKETBALL



	BOS	CHA	CLE	MIA	OKC	ORL	PHI	UTA
BOS			93					
CHA						96		
CLE	104							
MIA							104	
OKC								119
ORL		89						
PHI				91				
UTA					111			

$Z_{m,n}$ = Score of team m against n.

$Z_{n,m}$ = Score of team n against m.

PMF FOR NBA BASKETBALL

	BOS	CHA	CLE	MIA	OKC	ORL	PHI	UTA
BOS			93					
CHA						96		
CLE	104							
MIA							104	
OKC								119
ORL		89						
PHI				91				
UTA					111			

$Z_{m,n}$ = Score of team m against n.

$Z_{n,m}$ = Score of team n against m.

Offense

Defense

$$\underbrace{\boldsymbol{Z}}_{M \times N} \sim \underbrace{\boldsymbol{Y}}_{M \times N} = \underbrace{\boldsymbol{U}}_{M \times K} \times \underbrace{\boldsymbol{V}^{\mathsf{T}}}_{K \times N}$$

USING SIDE INFORMATION

In these PMF problems, there is often more information available than just the relational data.

Basketball:

- Date of the game
- Home team
- Starting lineup
- Referees

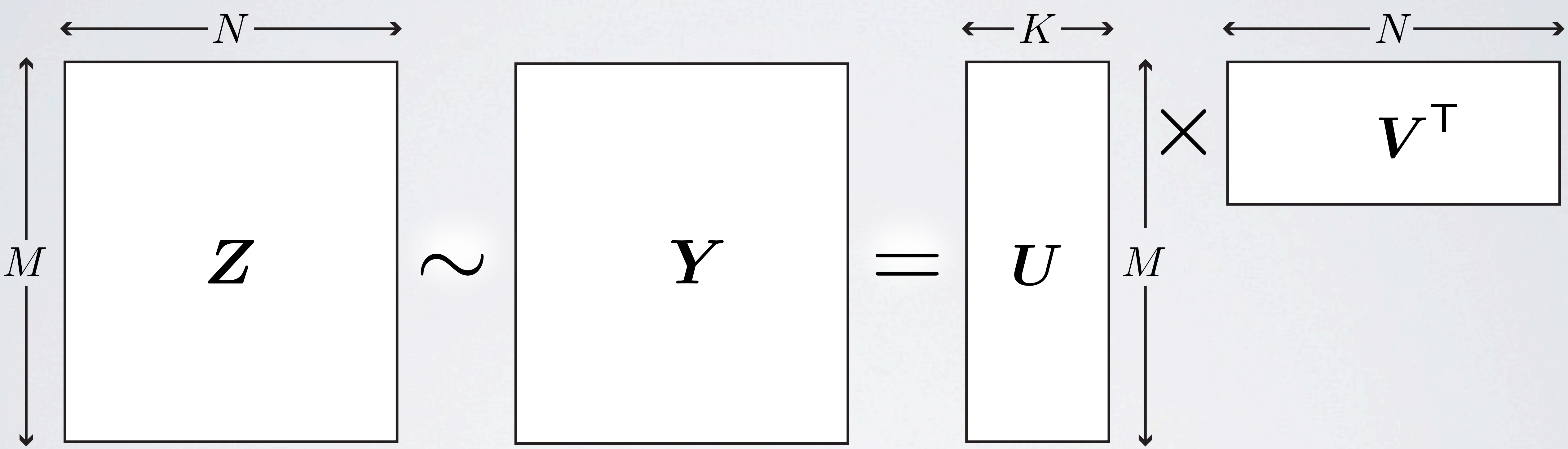
Film Ratings:

- Date of the rating
- User information
- Movie data
- Due to rental?

How to include this side information?

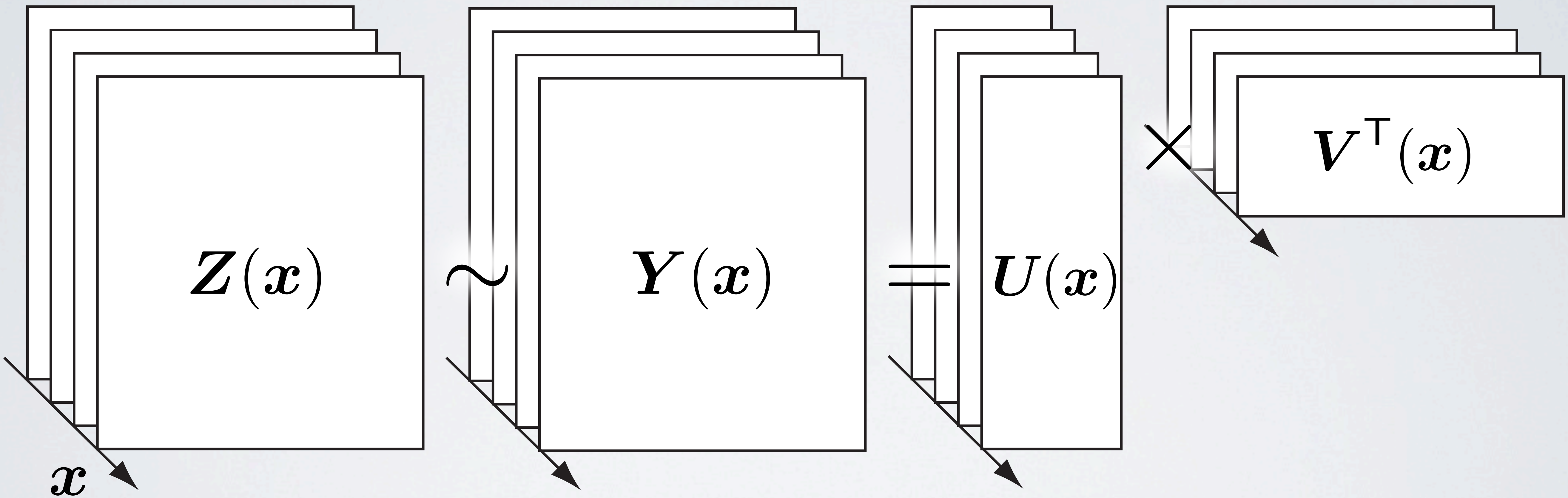
USING SIDE INFORMATION

Rather than a vector of latent features ...

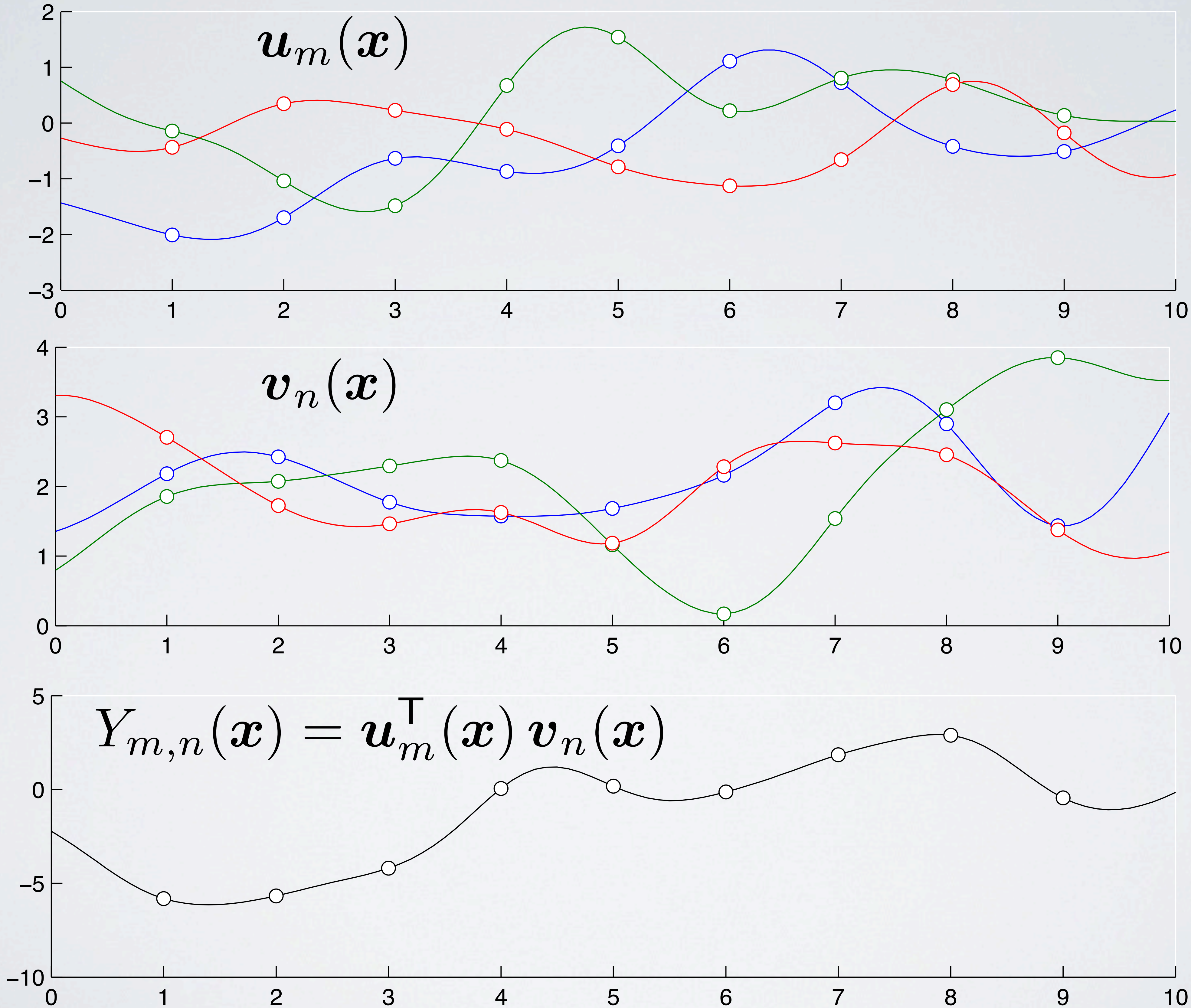


USING SIDE INFORMATION

... use **latent feature functions.**

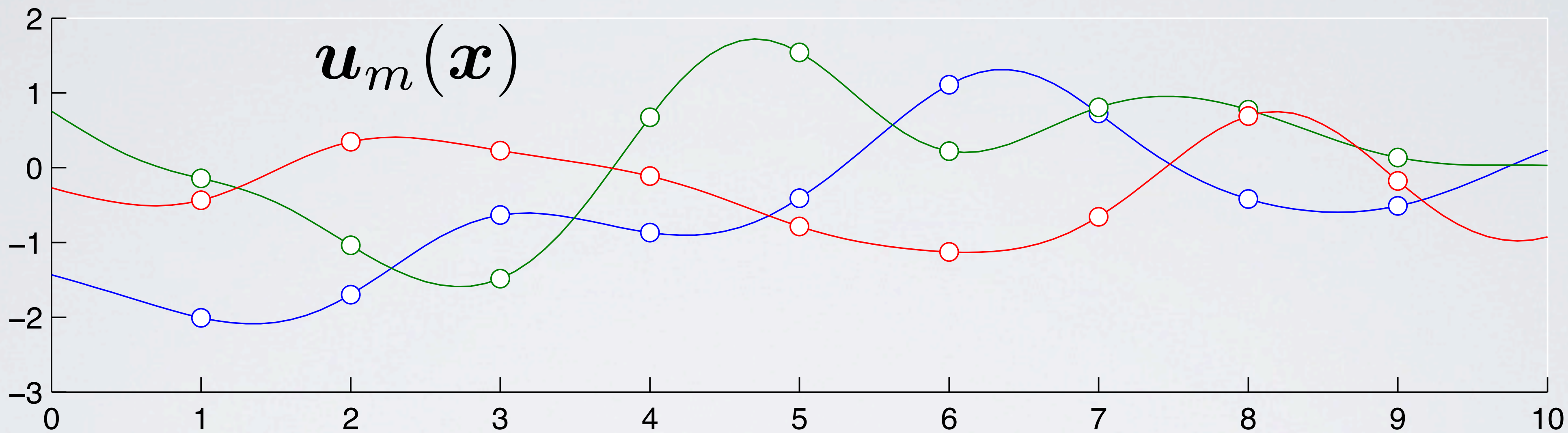


USING SIDE INFORMATION



SELECTING A PRIOR

We need to specify a reasonable prior on functions.

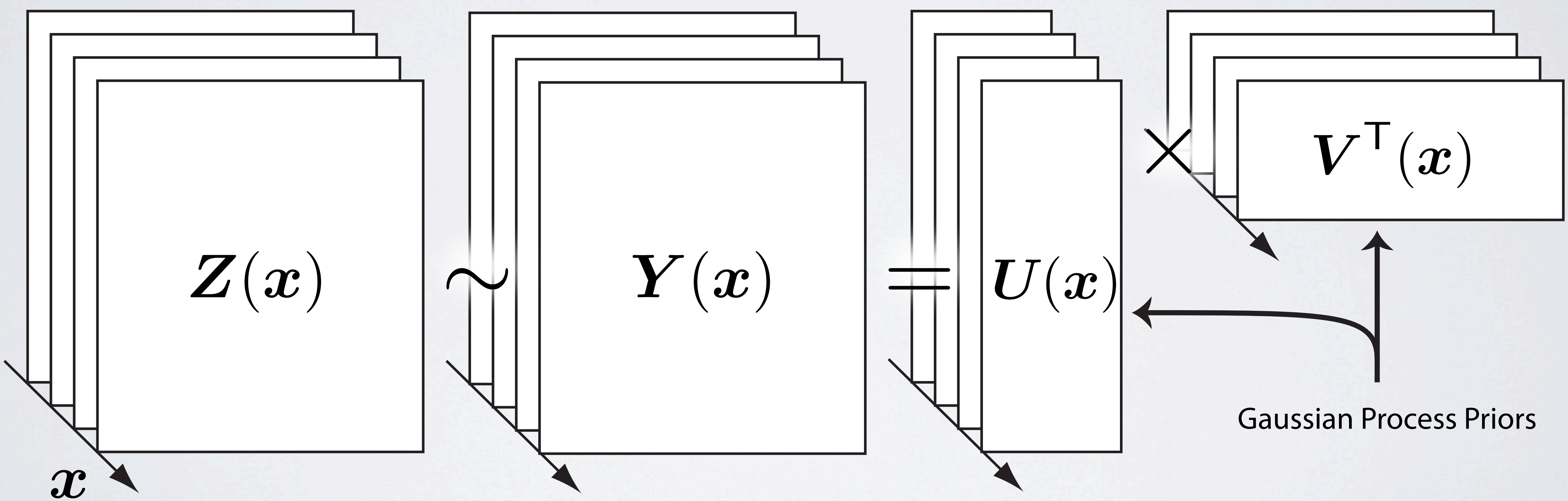


Desirable properties:

- Flexible.
- Enforce ideas of "smoothness".
- Specify sensitivity to input dimensions.
- Not have to choose a particular basis.

GAUSSIAN PROCESSES FOR PMF

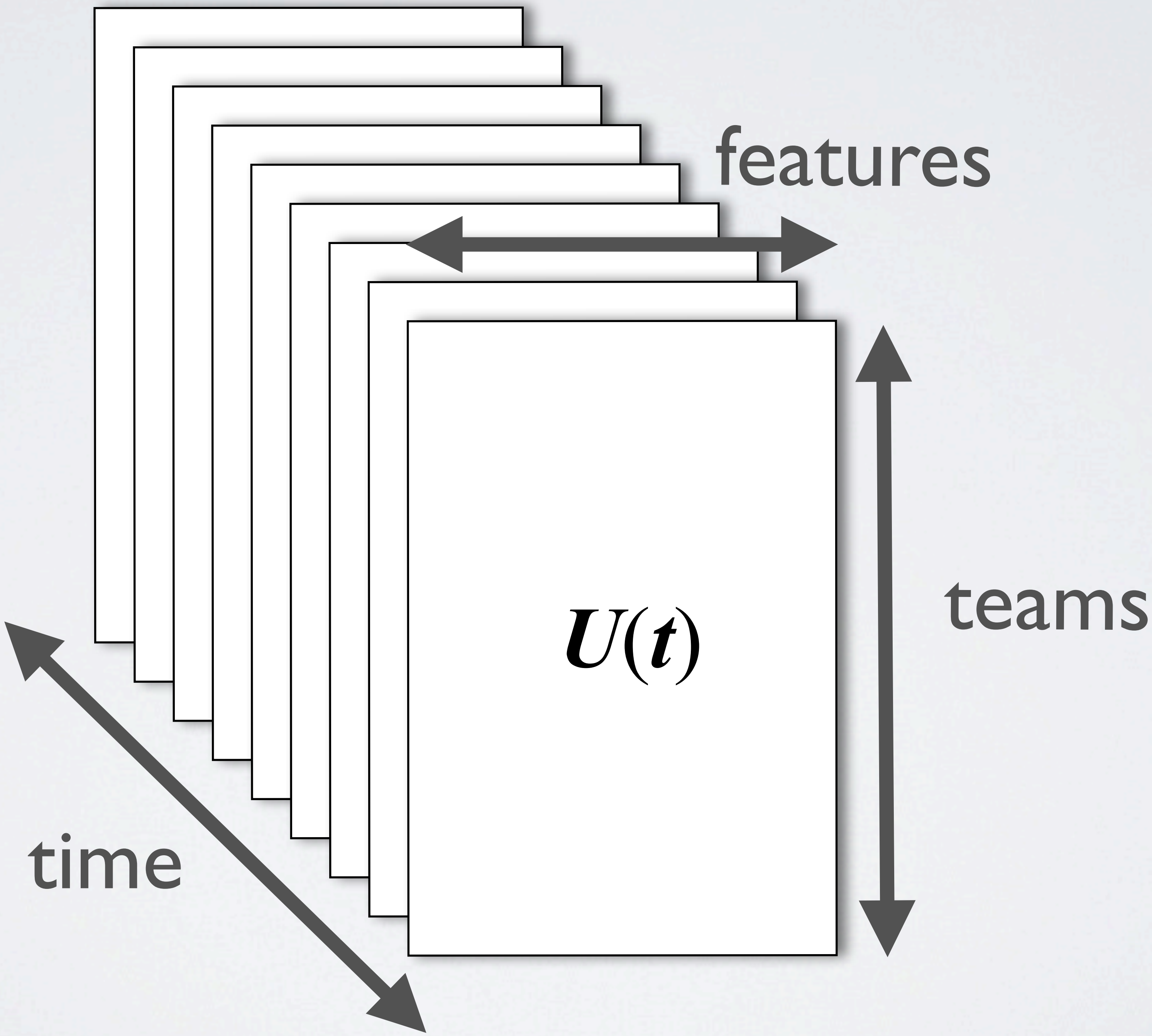
We use Gaussian processes to specify the general idea that *feature functions should be smooth in time* (or other side information) and *may depend differently on the different dimensions.*



Dependent Probabilistic Matrix Factorization

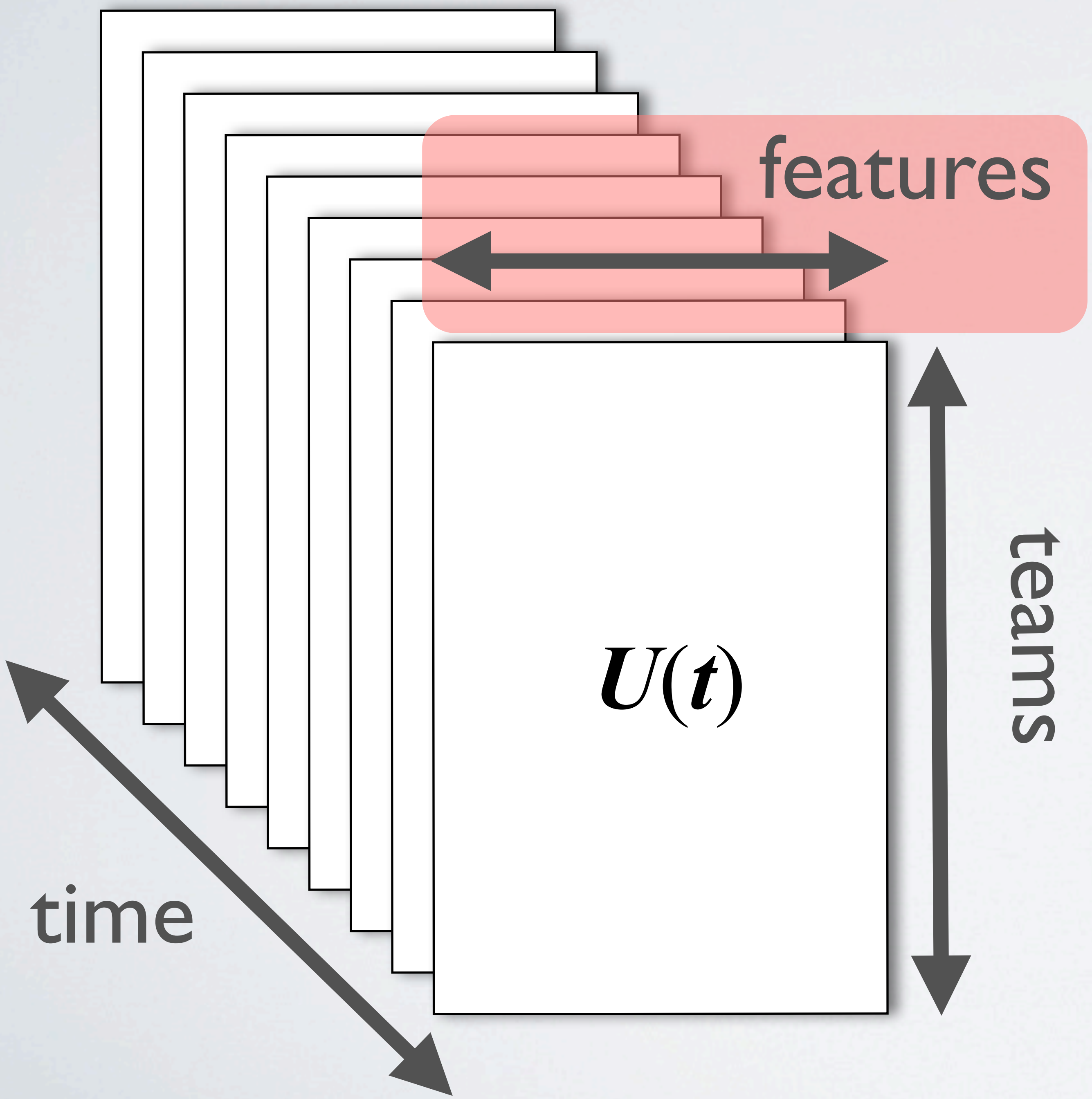
GAUSSIAN PROCESSES FOR PMF

Each team's offense and defense now has K features.



GAUSSIAN PROCESSES FOR PMF

Inter-Feature Sharing: a multi-task GP introduces dependencies between features.

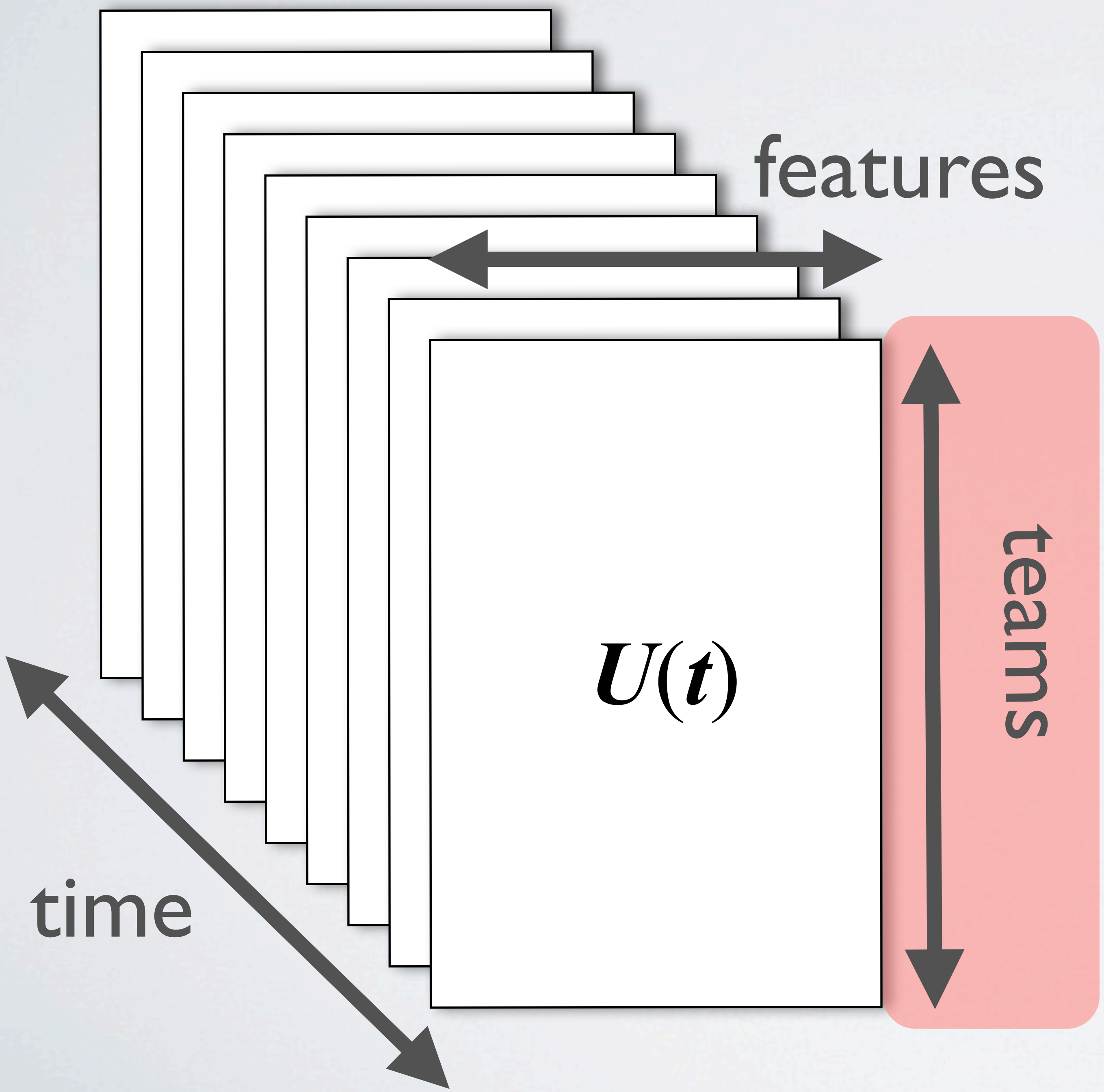


Captures ideas like:

"Good low-post players are often bad free throw shooters, e.g., Shaq, Dwight Howard."

GAUSSIAN PROCESSES FOR PMF

Intra-Feature Sharing: all teams share the same mean and covariance function for a given feature.



Captures ideas like:

"The effect of home-team advantage has similar properties for all teams."

INFERENCE IN THE DPMF

The posterior distribution is not available in closed form, so we use Markov chain Monte Carlo.

Slice sampling: MCMC without need for tuning.

We use two new methods for slice sampling Gaussian processes. Ask me about these offline if you're interested:

- "Elliptical slice sampling" - Murray, Adams, MacKay, AISTATS 2010, arXiv:1001.0175 [stat.CO]
- "Slice sampling covariance parameters of latent Gaussian models" - Murray and Adams: arXiv:0868.1062 [stat.CO]

DPMF FOR NBA BASKETBALL

We applied this new model to predicting regular-season NBA scores from 2002 to 2009.

Basketball is a nice machine learning problem:

- Medium-sized problem - about 10K observations.
- Natural rolling censored-data prediction problem.
- Expert human predictions are available.
- Side information is relevant for prediction.
- Properties of a team clearly vary over time.
- Hard, but fun!

NBA SCORE DATA

Properties of the NBA Score Data:

- 30 teams
- Each team plays 41 home games, 41 away.
- Regular season has 1230 games.
- We have 2002-03 through about half of 2009-10.
- Scores tend to be around 100 +/- 10.

A JOINT SCORE MODEL

	BOS	CHA	CLE	MIA	OKC	ORL	PHI	UTA
BOS			93					
CHA						96		
CLE	104							
MIA							104	
OKC								119
ORL		89						
PHI				91				
UTA					111			

$$\begin{bmatrix} Z_{m,n}(\boldsymbol{x}) \\ Z_{n,m}(\boldsymbol{x}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} Y_{m,n}(\boldsymbol{x}) \\ Y_{n,m}(\boldsymbol{x}) \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix} \right)$$

A JOINT SCORE MODEL

	BOS	CHA	CLE	MIA	OKC	ORL	PHI	UTA
BOS			93					
CHA						96		
CLE	104							
MIA							104	
OKC								119
ORL		89						
PHI				91				
UTA					111			

$$\begin{bmatrix} Z_{m,n}(\boldsymbol{x}) \\ Z_{n,m}(\boldsymbol{x}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} Y_{m,n}(\boldsymbol{x}) \\ Y_{n,m}(\boldsymbol{x}) \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix} \right)$$

EXPERIMENTAL SETUP

- Broke each season into four-week blocks.
- Ran inference to make predictions for each block.
- Could only see information in the past.
- Evaluated baseline Bayesian PMF against DPMF.
- Compared temporal, home/away and both.
- Tested latent factors K=1 through K=5.
- Compared predictive mean log probability of the test predictions with a Rao-Blackwellized estimate.
- Compared accuracy and RMSE to human experts.

EXPERT HUMAN PREDICTIONS?

Sports bookmakers provide two numbers:

Spread: The score offset that the favorite must win by in order for a bettor to yield an even-odds payout. "Lakers by 4.5"

Over/Under: An even-odds threshold threshold on the sum of the two scores. "Take the over at 205.5."

Solve a simple linear system to get scores:

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \text{away score} \\ \text{home score} \end{bmatrix} = \begin{bmatrix} \text{over/under} \\ \text{home spread} \end{bmatrix}$$

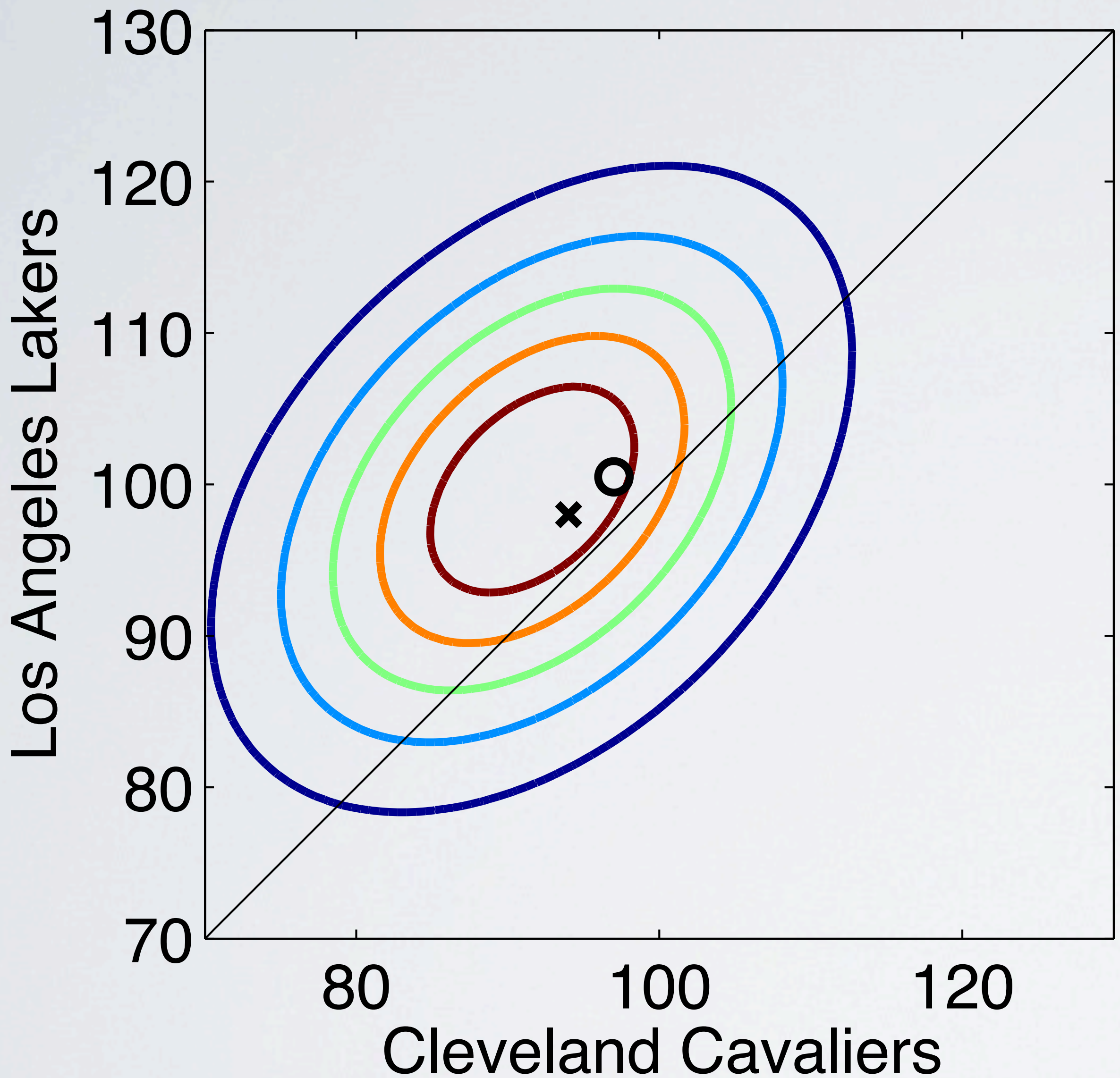
WINNER PREDICTION ERROR

	2002	2003	2004	2005	2006	2007	2008	2009
PMF	37.2	38.1	41.4	36.7	40.3	37.5	36.1	38.2
DPMF(h)	35.1	36.9	38.4	34.3	36.8	37.6	34.1	29.5
DPMF(t)	37.2	37.7	37.4	35.9	38.1	34.3	33.6	32.4
DPMF(t,h)	35.3	34.9	34.0	33.2	36.2	33.2	31.5	30.0
Human	30.9	32.2	29.7	31.4	33.3	30.6	29.8	29.4

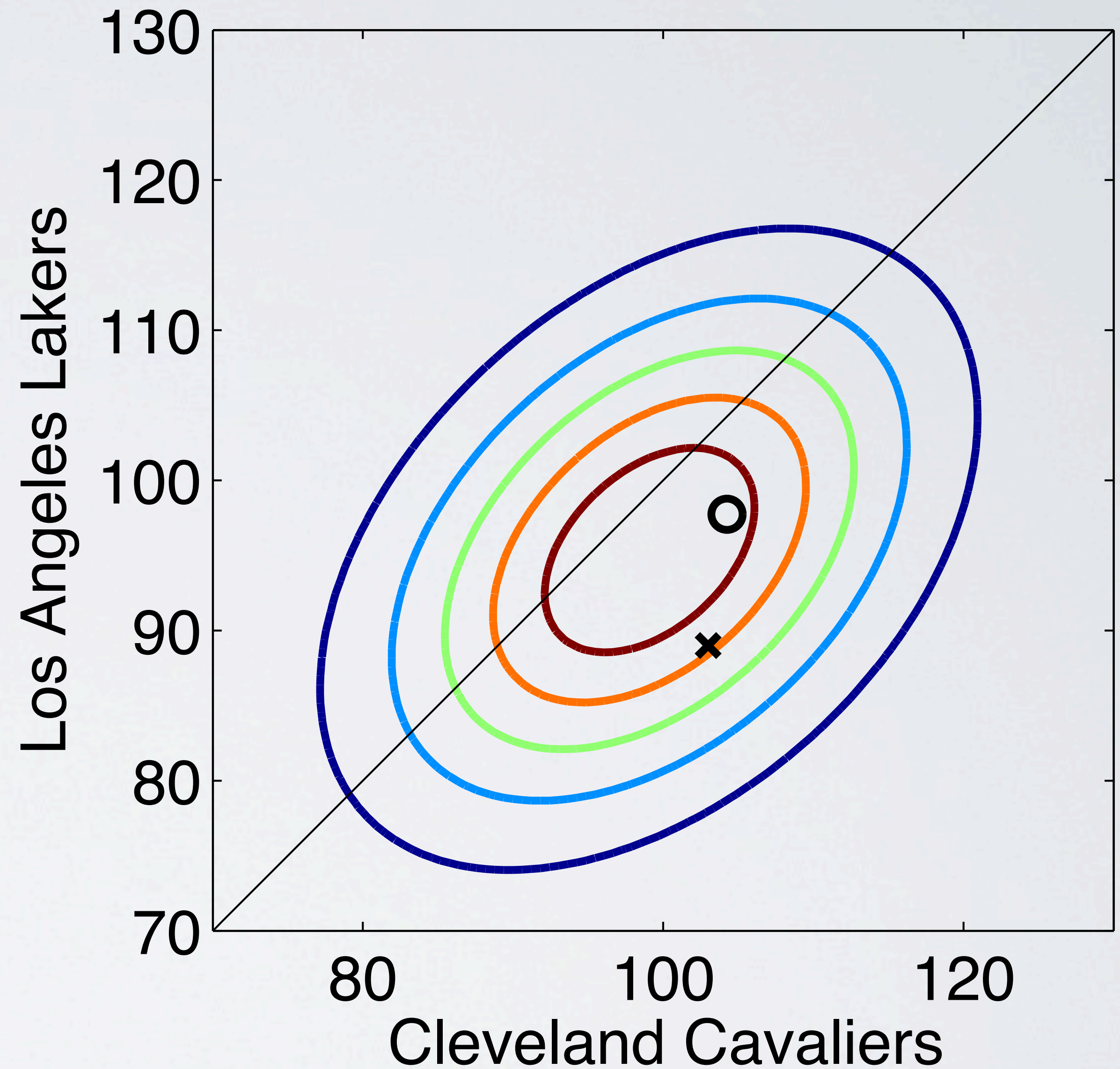
SCORE PREDICTION RMSE

	2002	2003	2004	2005	2006	2007	2008	2009
PMF	16.4	15.9	16.8	15.8	17.1	17.4	16.3	16.9
DPMF(h)	15.9	15.9	16.3	15.4	16.8	17.9	16.1	16.7
DPMF(t)	16.4	15.6	16.1	15.3	16.7	16.8	16.0	16.7
DPMF(t,h)	15.9	15.6	15.8	15.2	16.5	16.6	15.9	16.7
Human	14.9	14.4	14.6	14.7	15.4	15.2	15.0	15.5

LAKERS VS CAVALIERS

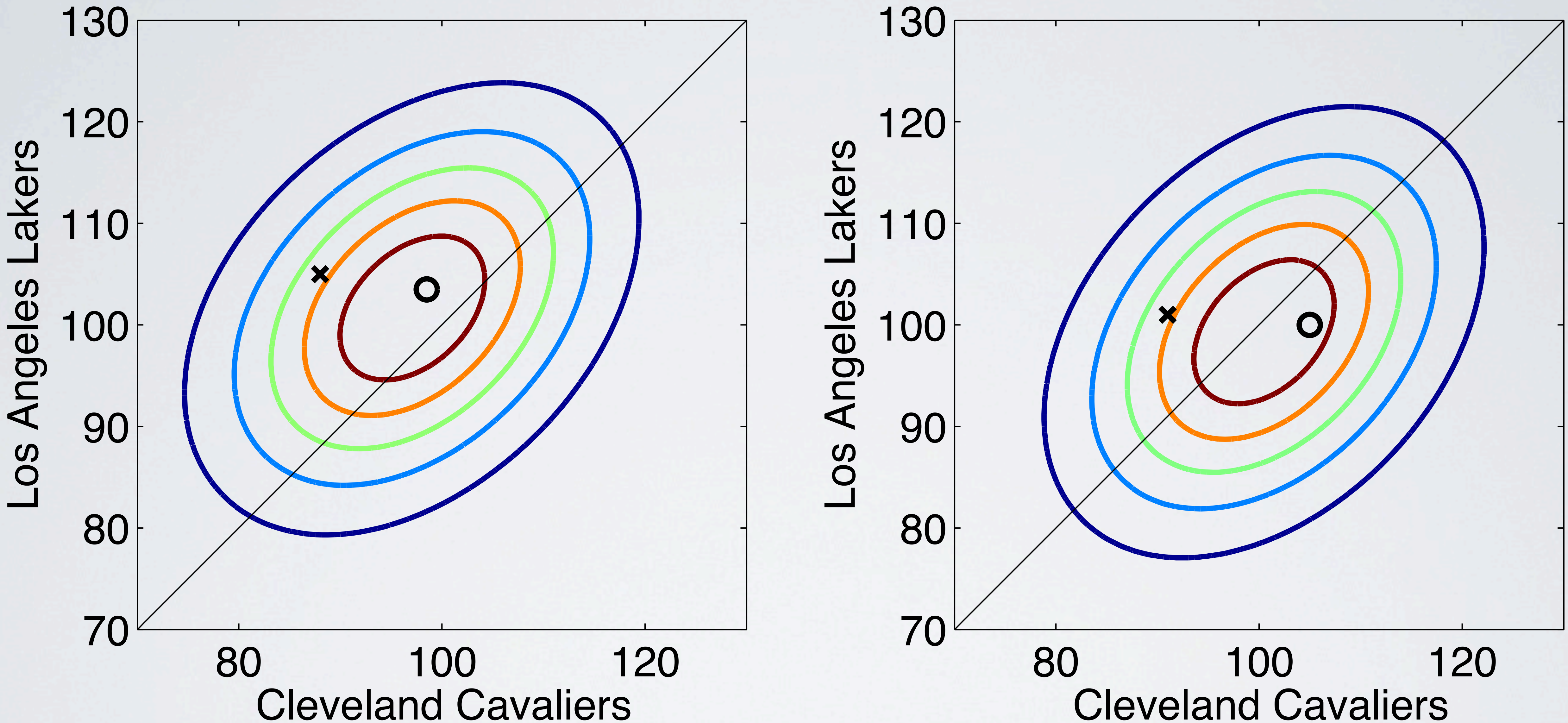


Lakers at home, 2004



Cavaliers at home, 2004

LAKERS VS CAVALIERS



Lakers at home, 2008

Cavaliers at home, 2008

SUMMARY

Probabilistic Matrix Factorization (PMF) is a powerful method for handling relational data.

There is frequently side information available that helps do the relational modeling, but incorporating it can be difficult.

We use Gaussian processes to create dependencies between PMF problems, hence "dependent probabilistic matrix factorization."

The NBA data application shows that this information can be successfully incorporated.

THANKS

- George Dahl (Co-author)
- Iain Murray (Co-author)
- Rich Zemel
- Danny Tarlow
- Geoff Hinton
- Amit Gruber
- Yoseph Barash

Code and Data at
http://www.cs.toronto.edu/~rpa/dpmf