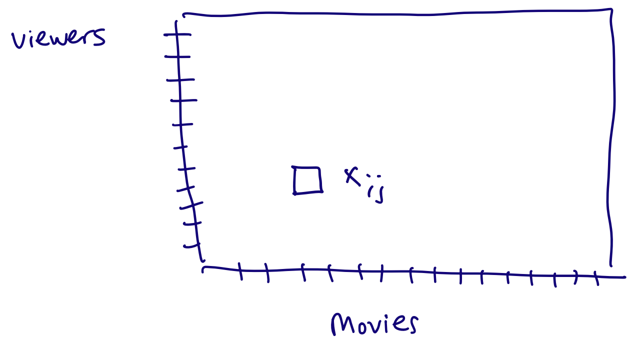
$f(\log a, \log b) \to \log(a+b) \leftarrow$

$\text{log sum exp}(\log x_1, \dots, \log x_p) \to \log(x_1 + \dots + x_p)$

This one trick will prevent all your NaNs

Matrix Factorization



interaction matrix

x_{ij} : rating viewer i gave to movie j

interaction data:

- viewers rating movies
- web users clicking on pages
- lawmakers voting on bills (roll call data)
- documents containing words
- cells contain genes, RNA expression, proteins
- people contain little DNA sequences.
- people bid on items
- social network data: x_{ij} is whether $i \leftrightarrow j$
- word co-occurrence x_{ij} is # times words co-occur

predict missing data
rank missing cells
understand latent structure.

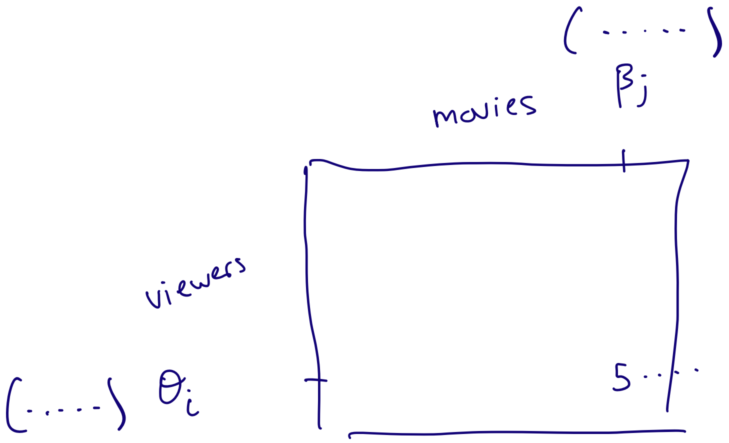
Rows: viewers Cols: Movies

K: # of components

$$x_{ij} \sim f(x \,;\, \underbrace{\theta_i \cdot \beta_j}_{K\text{-vector}})$$

θ_i : preferences

β_j : attributes



$x_{ij} \sim \mathcal{N}(\theta_i \cdot \beta_j, \sigma^2)$

For each viewer i

$$\theta_i \sim \mathcal{N}_k(0, \eta_\theta^2)$$

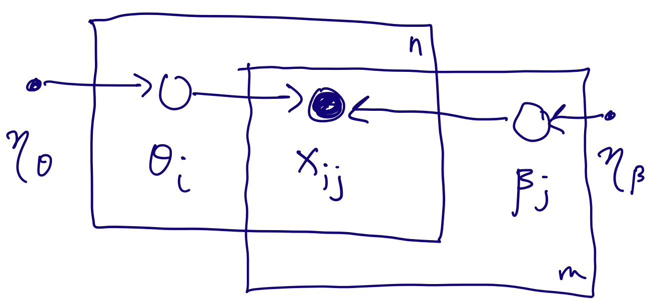
For each movie j

$$\beta_j \sim \mathcal{N}_k(0, \eta_\beta^2)$$

For each rated movie (i,j)

$$x_{ij} \sim \mathcal{N}(\theta_i \cdot \beta_j, 1)$$

<u>Prob. Matrix Factorization</u>



E.B. PMF

$$p(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}) = \left(\prod_{j=1}^{m} p(\beta_j ; \eta_\beta)\right)\left(\prod_{i=1}^{n} p(\theta_i ; \eta_\theta)\right)\left(\prod_{(i,j)\in \boldsymbol{x}} p(x_{ij} \mid \theta_i, \beta_j)\right)$$

$$\log p(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \boldsymbol{x}) = -\sum_{j=1}^{m} \|\beta_j\|^2 / 2\eta_\beta^2 - \sum_{i=1}^{n} \|\theta_i\|^2 / 2\eta_\theta^2 - \sum_{i,j \in \boldsymbol{x}} \frac{1}{2}(x_{ij} - \theta_i \cdot \beta_j)^2 + \text{cnst.}$$

$$\|\beta_j\|^2 = \sum_{k=1}^{k} \beta_{jk}^2$$

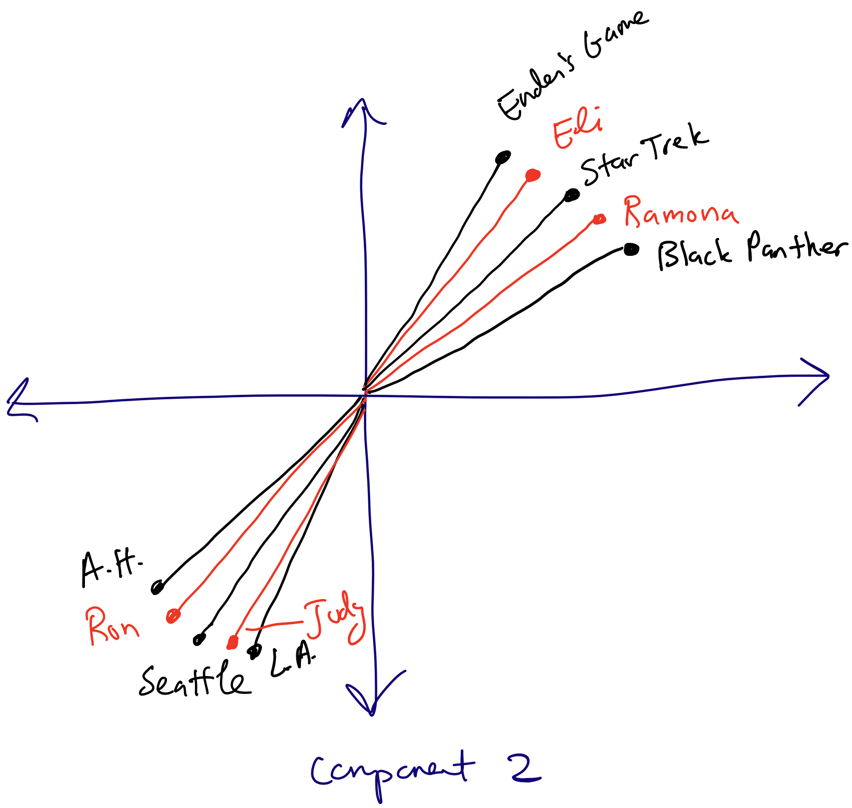
	Star Trek	Ender's Game	Black Panth	Sleep	Love	An
Ramona	+		+	−		
Eli	+	+				
Ron	−	−		+	+	+
Judy		−		+	+	+

K = 2

K=2

θ

β



$$\theta_i = (\theta_{i1}, \theta_{i2})$$

Ideal point models

roll call data x_{ij} vote of lawmaker i on bill j

$K = 1$

$$\alpha_j \sim N(0, \eta_\alpha^2) \quad \text{for each bill } j$$
$$\beta_j \sim N(0, \eta_\beta^2)$$

α_j : popularity
β_j : polarity

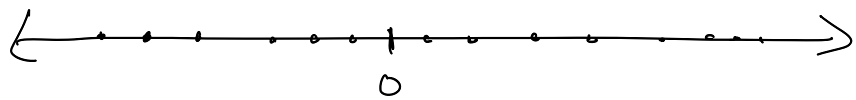
$$\theta_i \sim N(0, \eta_\theta^2) \quad \text{for each lawmaker } i$$

θ_i : ideal point

$$x_{ij} \mid \theta_i, \beta_j, \alpha_j \sim \text{Bernoulli}(\sigma(\theta_i \beta_j + \alpha_j))$$

Nominate scores.

θ



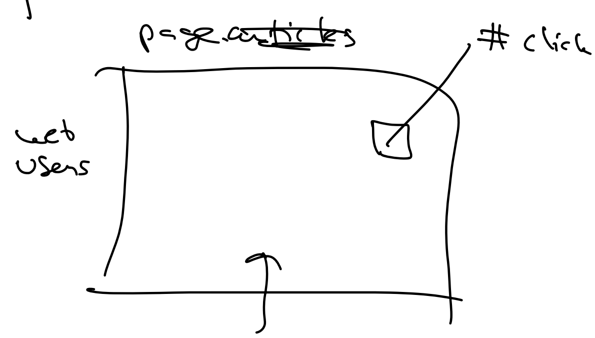
Poisson factorization

- $\underline{\beta_j} \sim \text{Gamma}_K(\lambda_\beta, \sigma_\beta)$
- $\theta_i \sim \text{Gamma}_K(\lambda_i, \sigma_\theta)$

$$x_{ij} \sim \text{Poisson}(\theta_i \cdot \beta_j)$$

interpretable.

positive count matrices.



"implicit data"

1. Fitting Gaussian MF
2. Use MF for recommendation
3. Other forms of MF
4. Interesting issues in recommendation systems — implicit data, cold start problem, attributes / content

Rec Sys.

Movie Lens

For each viewer

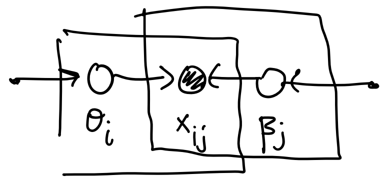
$$\theta_i \sim \mathcal{N}_k(0, \eta_\theta^2)$$

For each movie

$$\beta_j \sim \mathcal{N}_k(0, \eta_\beta^2)$$

For each viewing

$$x_{ij} \sim \mathcal{N}(\theta_i \cdot \beta_j, \sigma^2)$$



β_j , w_j

θ_j z_j

$w_j \rightarrow \beta_j$

MAP estimation of θ_i $i=1..n$ β_j $j=1..m$

$$\log p(-) = \sum_{j=1}^{m} \log p(\beta_j ; \eta_\beta) + \sum_{i=1}^{n} \log p(\theta_i ; \eta_\theta) + \sum_{(i,j)\in X} \log p(x_{ij} \mid \theta_i, \beta_j)$$

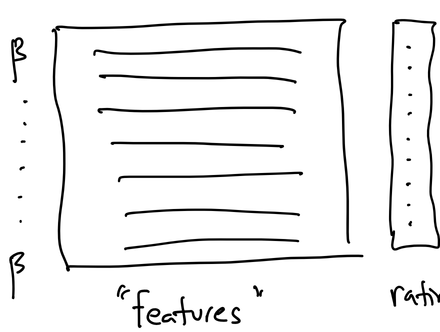
coordinate ascent:

For each i
optimize θ_i
For each j
optimize β_j
Repeat

$$\mathcal{L}(\theta_i) = \log p(\theta_i ; \eta_\theta) + \sum_{j\in \mathbf{X}_i} \log p(x_{ij} \mid \theta_i, \beta_j)$$

$$-\frac{1}{2\eta_\theta^2}\|\theta_i\|^2 - \frac{1}{2\sigma^2}\sum_{j\in \mathbf{X}_i}(x_{ij} - \theta_i\cdot\beta_j)^2 + c$$

viewer i :

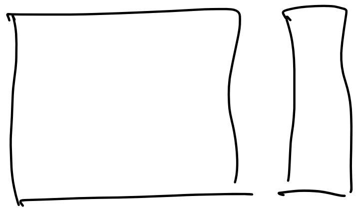


(β_ℓ, x_ℓ)

$$\nabla_{\theta_i}\mathcal{L} = -\frac{1}{\eta_\theta^2}\theta_i + \frac{1}{\sigma^2}\sum_{j\in X_i}(x_{ij} - \theta_i\cdot\beta_j)\beta_j$$

$$\mathcal{L}(\beta_j) = -\frac{1}{2\eta_\beta^2} \|\beta_j\| - \frac{1}{2\sigma^2} \sum_{i \in \mathbf{X}_j} (x_{ij} - \theta_i \cdot \beta_j)^2$$

$\equiv$



preferences of people who rated movie j

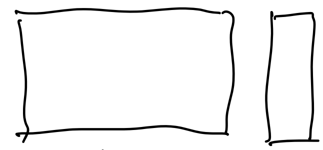
ratings of movie j

EASE model.

H. Steck.

each movie is a set of coefficients β_j

$\beta_j \in \mathbb{R}^{m-1}$

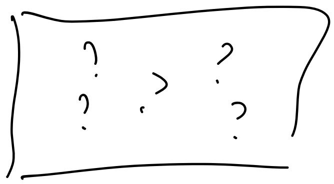


histories of the users.

users who rated item j

How to use a MF.

Matrix completion.



$$p(x_{ij} \mid \mathbf{X}) = \int p(x_{ij} \mid \theta_i, \beta_j)\, p(\theta_i, \beta_j \mid \mathbf{X})\, d\theta_i\, d\beta_j$$

$$\approx p(x_{ij} \mid \hat{\theta}_i, \hat{\beta}_j)$$

$$\mathbb{E}[X_{ij} \mid \mathbf{X}] \approx \hat{\theta}_i \cdot \hat{\beta}_j \triangleq \hat{x}_{ij}$$

---

other forms of factorization.

$$x_{ij} \sim f(x;\ \theta_i \cdot \beta_j)$$

ideal point model

Clinton + 2002?

$\alpha_j, \beta_j \sim p(\cdot)$ for each bill j

$\theta_i \sim p(\cdot)$ for each lawmaker i

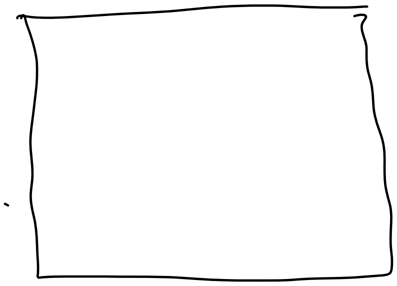
$x_{ij} \sim \mathrm{Bern}(\sigma(\theta_i \beta_j + \alpha_j))$

Poisson factorization

- $\beta_j \sim \text{Gamma}_K(\cdot,\cdot)$
- $\theta_i \sim \text{Gamma}_K(\cdot,\cdot)$
- $x_{ij} \sim \text{Poi}(\theta_i \cdot \beta_j)$

web pages
genes

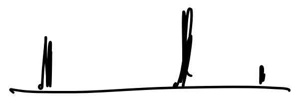
web users
cells
documents.



counts

sparse.

$p(\beta, \theta \mid x)$



β

θ

nonnegativity.

nonnegative matrix factorization

Seung and Lee (1999)

$(x_{ij} - \theta\beta)^2$

"KL objective"

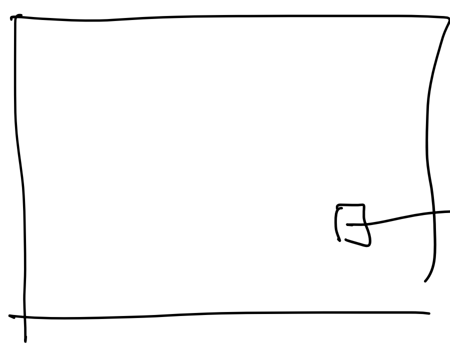
Latent Space Model

social. network analysis

Karate dataset.

entity θ

entity ~~people~~ θ



is i friends w/ j ?

$x_{ij} \sim \text{Bern}(\sigma(d(\theta_i, \theta_j)))$

Hoff + 2002

MCMC.

$d: \theta_i \cdot \theta_j$

$\|\theta_i - \theta_j\|^2$