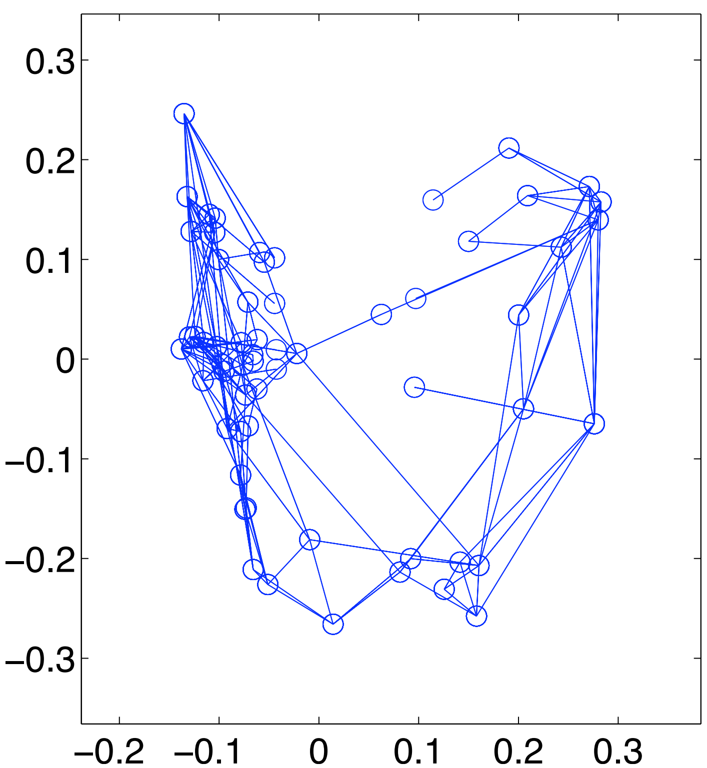
Review

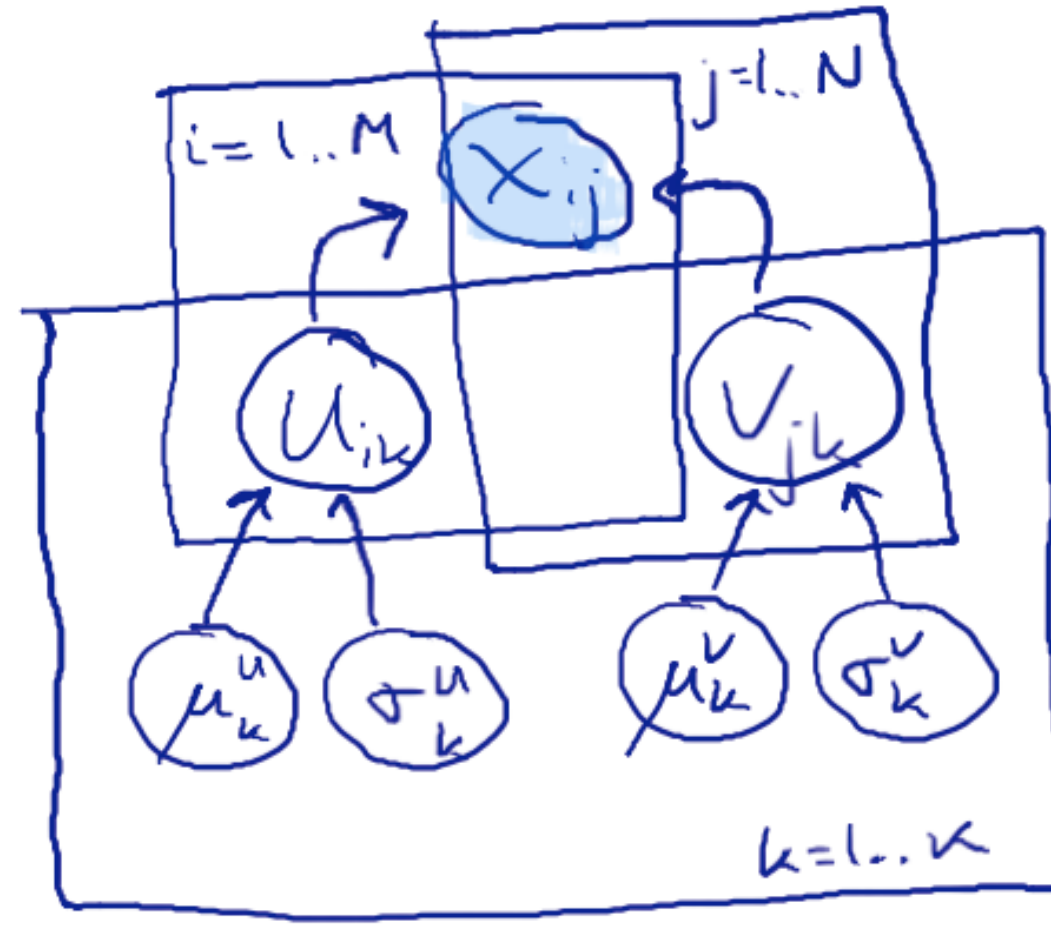
dolphin friendships



- Models that use SVD or eigen-analysis
  - PageRank: eigen-analysis of ***random surfer*** transition matrix
    - usually uses only *first* eigenvector
  - Spectral embedding: eigen-analysis (or equivalently SVD) of random surfer model in ***symmetric*** graph
    - usually uses 2nd–Kth EVs (small K)
    - first EV is boring
  - Spectral clustering = spectral embedding followed by clustering

Review: PCA

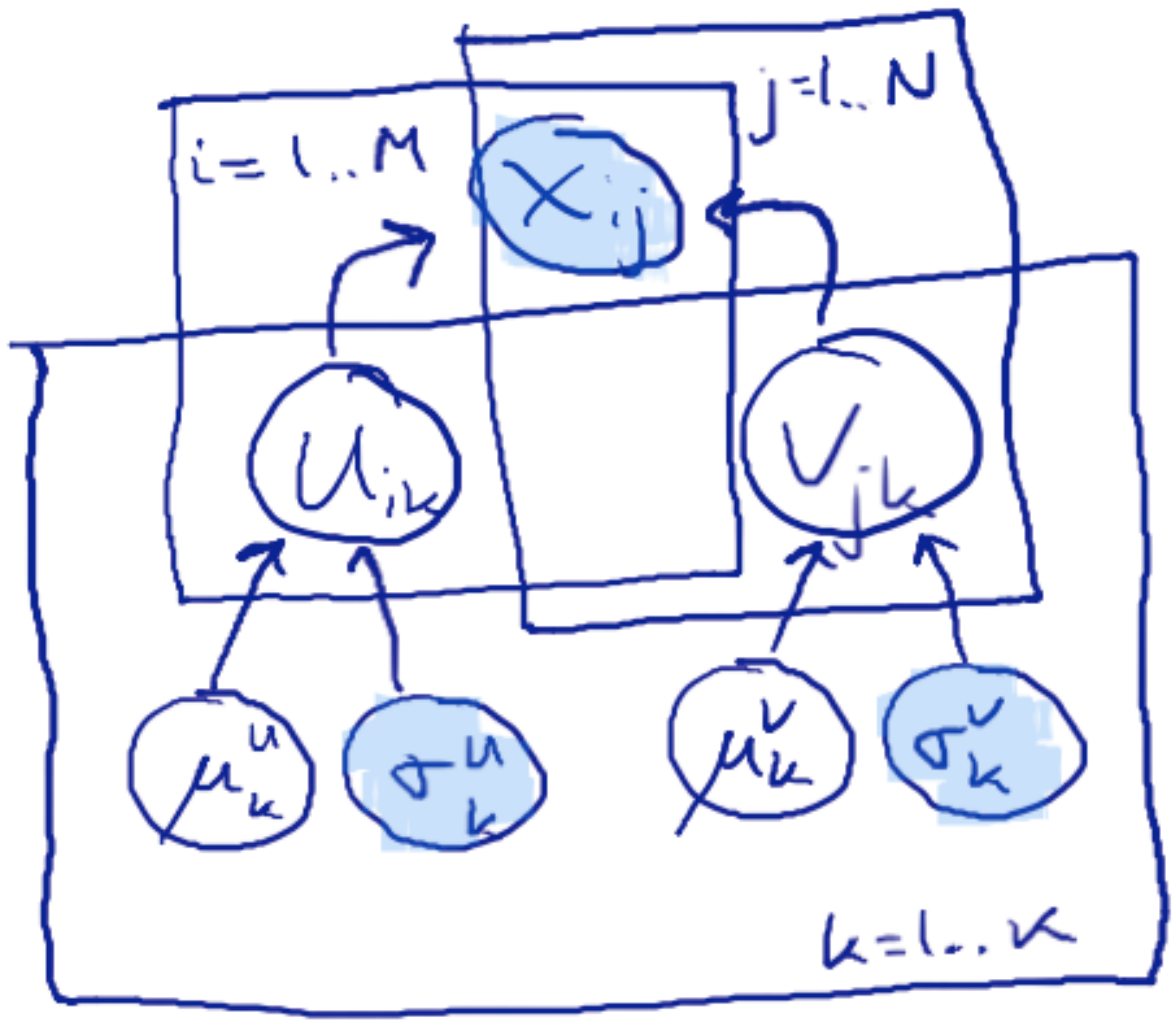
- The good: simple, successful
- The bad: linear, Gaussian
  - $E(X) = UV^T$
  - X, U, V ~ Gaussian
- The ugly: failure to generalize to new entities
  - Partial answer: ***hierarchical PCA***



What about the second rating for a new user?

- MLE/MAP of $U_{i\cdot}$ from one rating:
  - knowing μ^U:
  - result:
- How should we fix?
- Note: often have only a few ratings per user

MCMC for PCA



Need:

- Can do Bayesian inference by Gibbs sampling—for simplicity, assume σs known

Recognizing a Gaussian

- Suppose $X \sim N(X \mid \mu, \sigma^2)$
- $L = -\log P(X{=}x \mid \mu, \sigma^2) =$
  - $dL/dx =$
  - $d^2L/dx^2 =$
- So: if we see $d^2L/dx^2 = a$, $dL/dx = a(x - b)$
  - $\mu =$ $\sigma^2 =$

Gibbs step for an element of μ^U

- L = $\text{const.} + \sum_{ij} (x_{ij} - \sum_{k'} u_{ik'} v_{jk'})^2 / 2\sigma^2$

$+ \sum_{ik} (u_{ik} - \mu^u_k)^2 / 2(\sigma^u_k)^2 + \sum_{jk} (v_{jk} - \mu^v_k)^2 / 2(\sigma^v_k)^2$

Gibbs: element of U

- L = $\text{const.} + \sum_{ij} (X_{ij} - \sum_{k'} U_{ik'} V_{jk'})^2 / 2\sigma^2$
$+ \sum_{ik} (U_{ik} - \mu^u_k)^2 / 2(\sigma^u_k)^2 + \sum_{jk} (V_{jk} - \mu^v_k)^2 / 2(\sigma^v_k)^2$

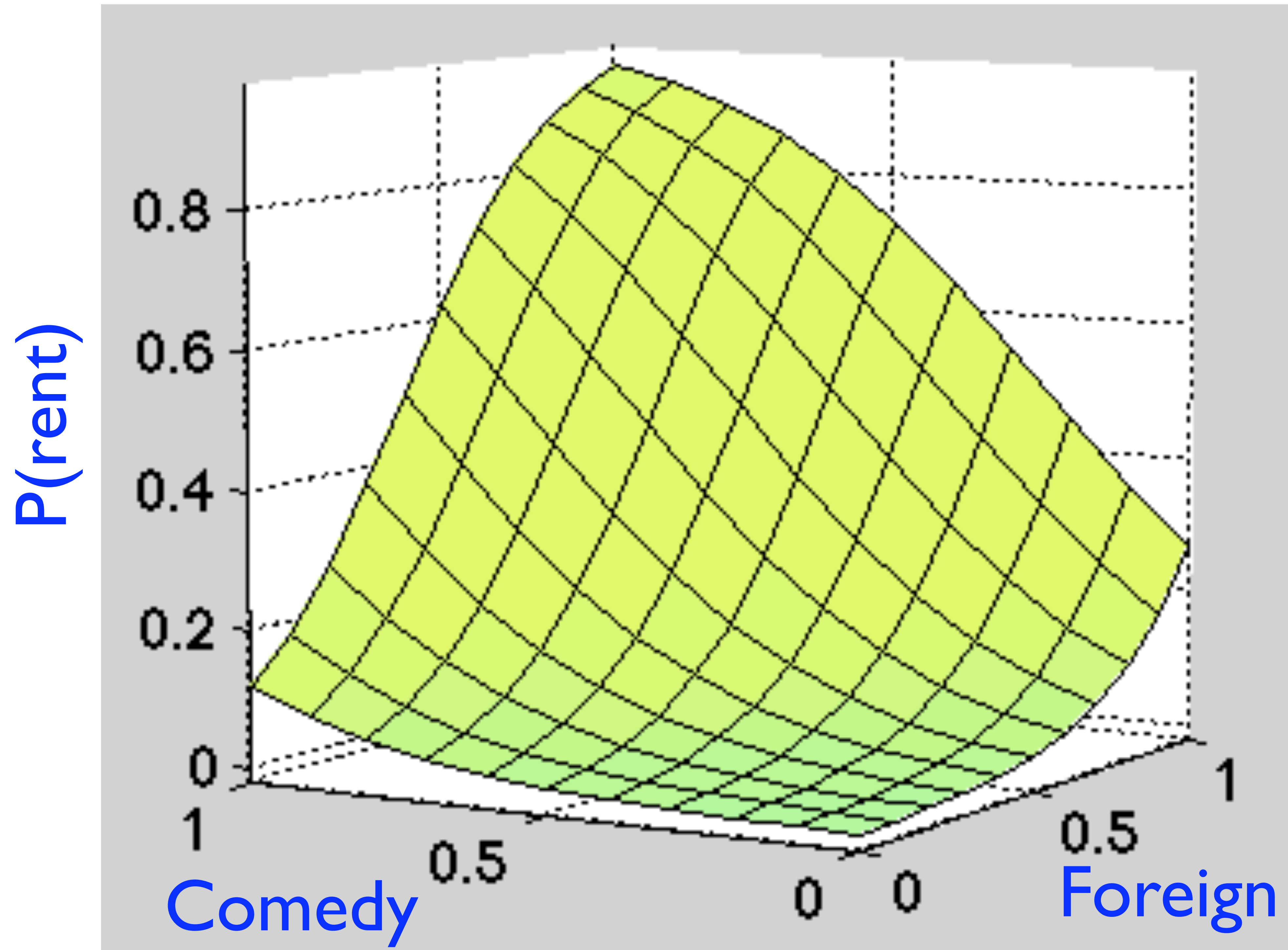
- dL / dU_{ik} =

- $dL^2 / (dU_{ik})^2$ =
  - post. mean = post. var. =

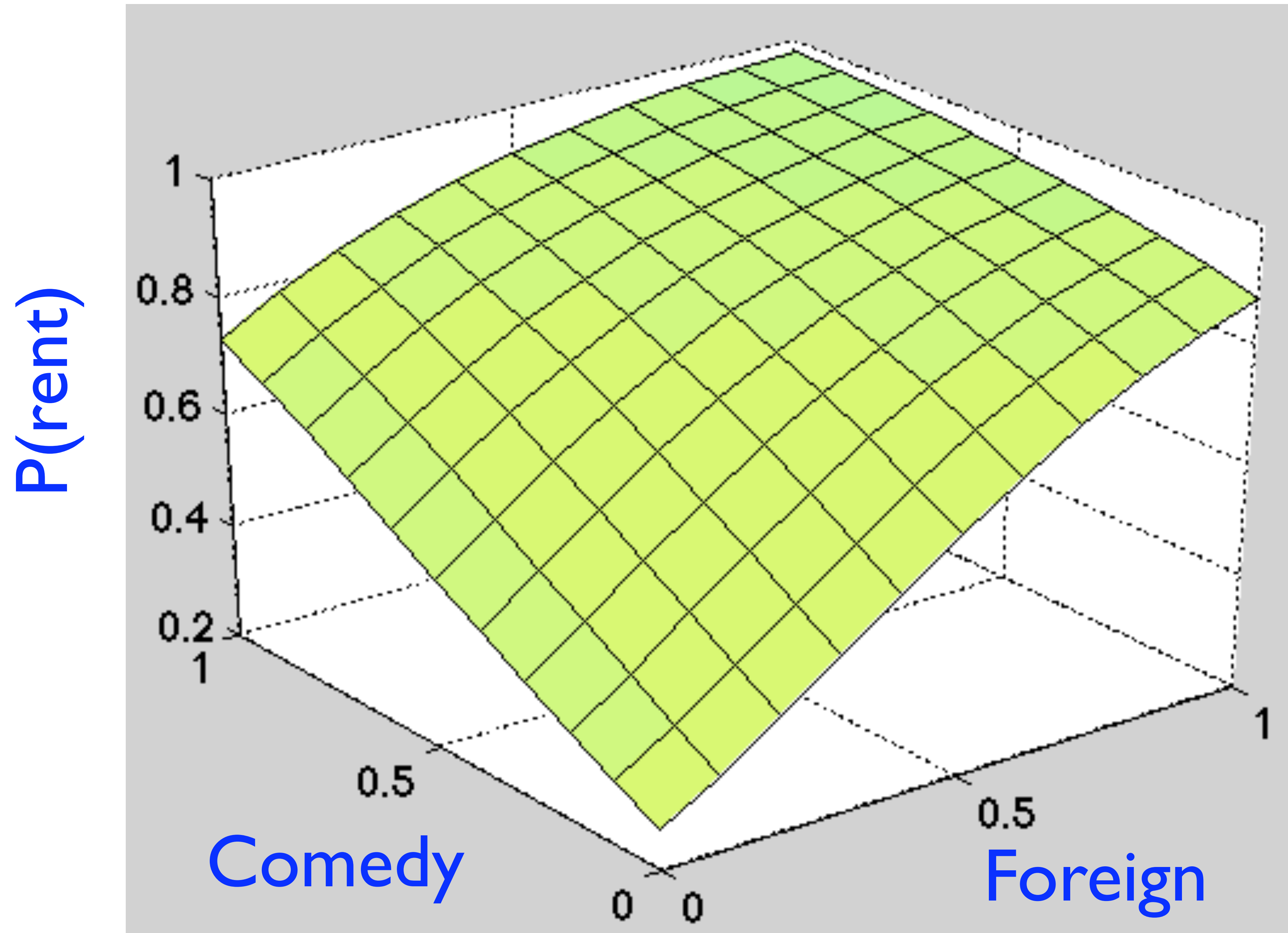
In reality

- Above, blocks are single elements of U or V
- Better: blocks are entire rows of U or V
  - take gradient, Hessian to get mean, covariance
  - formulas look a lot like linear regression (normal equations)
- And, want to fit σ^U, σ^V too
  - sample $1/\sigma^2$ from a ***Gamma*** (or Σ^{-1} from a ***Wishart***) distribution

Nonlinearity: conjunctive features



Disjunctive features



Non-Gaussian

- X, U, and V could each be non-Gaussian
  - e.g., binary!
    - rents(U, M), comedy(M), female(U)
- For X: predicting –0.1 instead of 0 is only as bad as predicting +0.1 instead of 0
- For U, V: might infer –17% comedy or 32% female

Logistic PCA

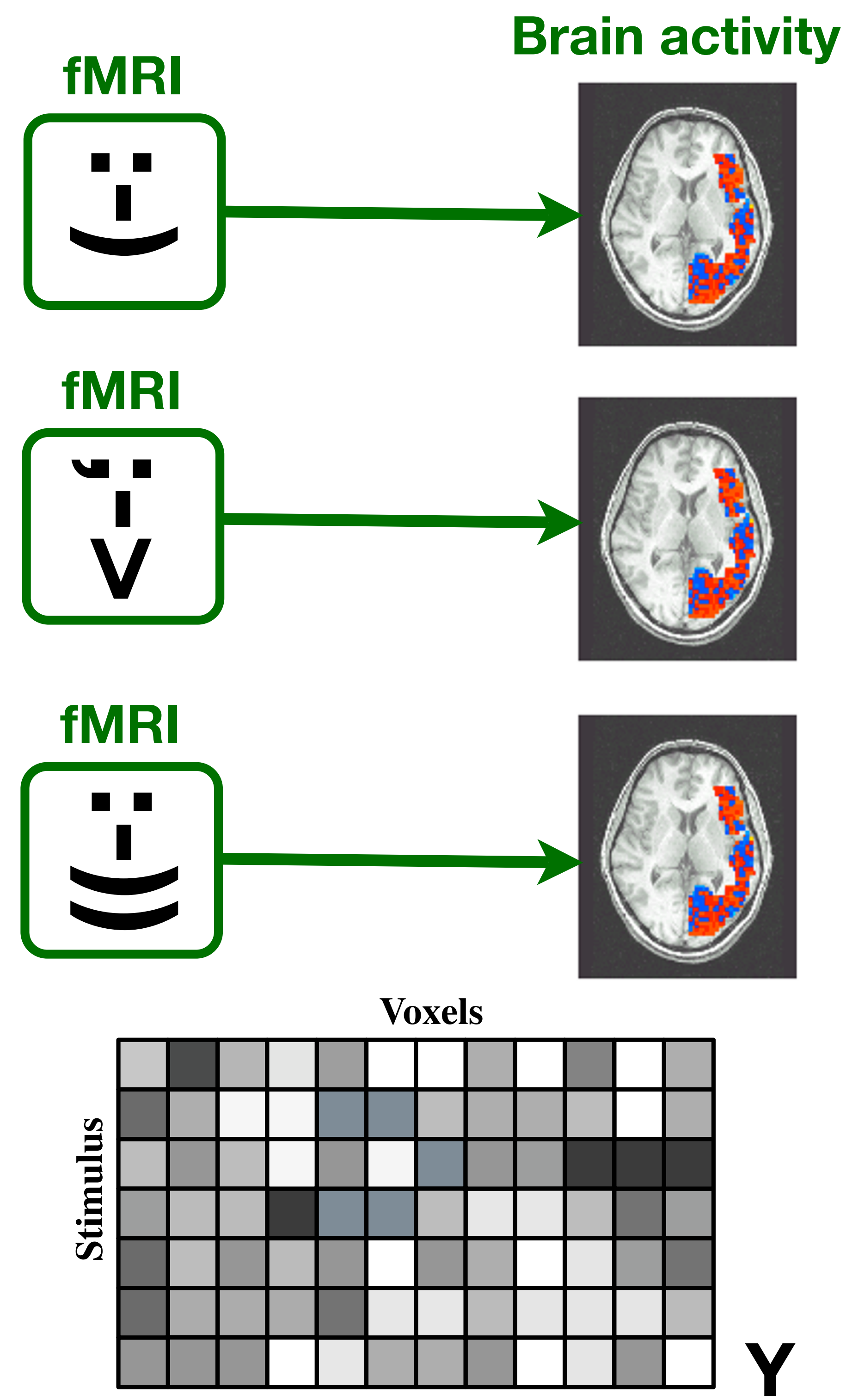
- Regular PCA: $X_{ij} \sim N(U_i \cdot V_j, \sigma^2)$
- Logistic PCA:
- Might expect learning, inference to be hard
  - but, MH works well, using $dL/d\theta$, $d^2L/d\theta^2$
- Generalization: ***exponential family PCA***
  - w/ optional hierarchy, Bayesianism

Application: fMRI

stimulus: “dog”

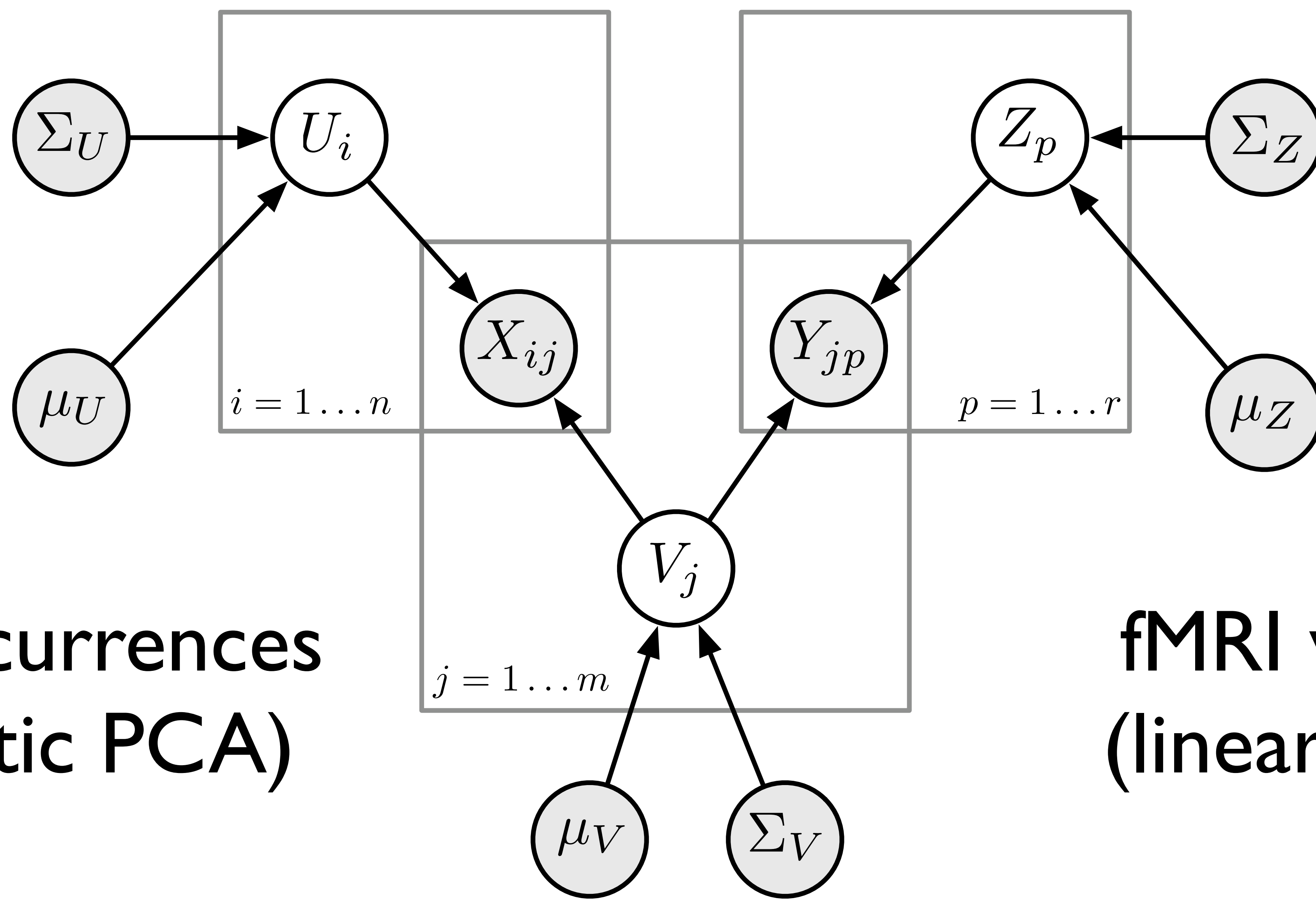
stimulus: “cat”

stimulus: “hammer”



credit: Ajit Singh

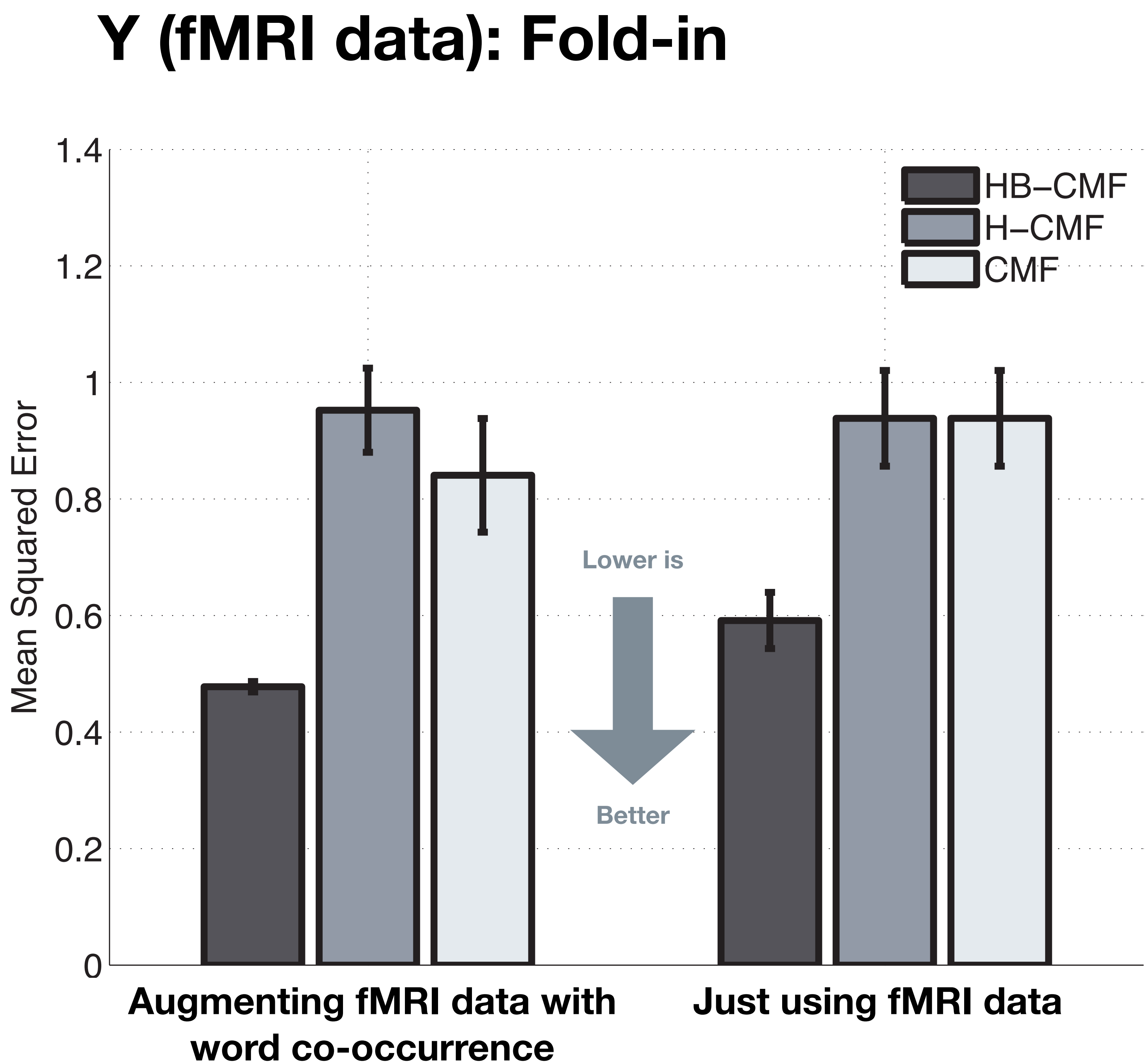
2-matrix model



co-occurrences
(logistic PCA)

fMRI voxels
(linear PCA)

Results (logistic PCA)



credit: Ajit Singh