MODELING ANNOTATED DATA

Reviewer: Saurabh Singh (ss1@uiuc.edu)

Problem

- Modeling of associated document items
  - Images & Annotations
  - Papers & Bibliographies
  - Genes & Functions
- Documents are considered as pairs of data streams.
- One type provides annotation for the other type.

Uses

- Retrieval, Clustering, Classification
- Automatic annotation
- Retrieval of un-annotated data.

This paper

Models Images (**r**) and Annotations (**w**)

Three primary tasks

- Joint distribution of an image and its caption (Clustering, Organization)
- Conditional distribution of words given an image. (Automatic annotation, text based retrieval)
- Conditional distribution of words given a region of an image. (Automatic labeling of regions)

Modeling

K factors or topics

- Each a distribution over words
- Each a distribution over image regions

Latent variables

- Topic assignments
- Distribution parameters (for components)

Features

Document: **(r, w),** N regions, M words

Distributions

$p(\mathbf{r}, \mathbf{w})$, $p(w \mid \mathbf{r})$, $p(w \mid \mathbf{r}, r_n)$

Text annotations

Vocabulary: 168 Terms (V)

Captions: 2-4 Words per Image

Multinomials on V conditioned on topics

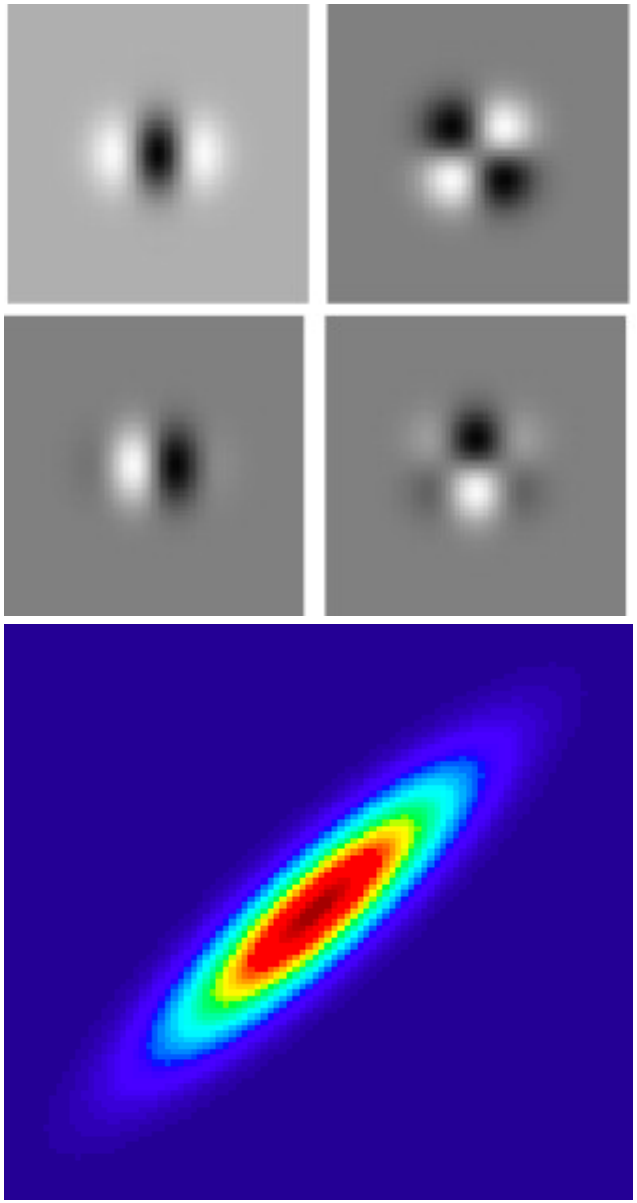
Images

Composed of 6-10 regions via N-cuts

Each region summarized as a feature vector ~40

- Size: Percentage of image
- Position: Center of mass [0, 1]
- Color: μ, σ of R,G,B, L, a, b etc.
- Texture: μ, σ of filter responses
- Shape: area/perimeter^2, moment of inertia etc.

Multivariate Gaussian over features: **μ**, Σ

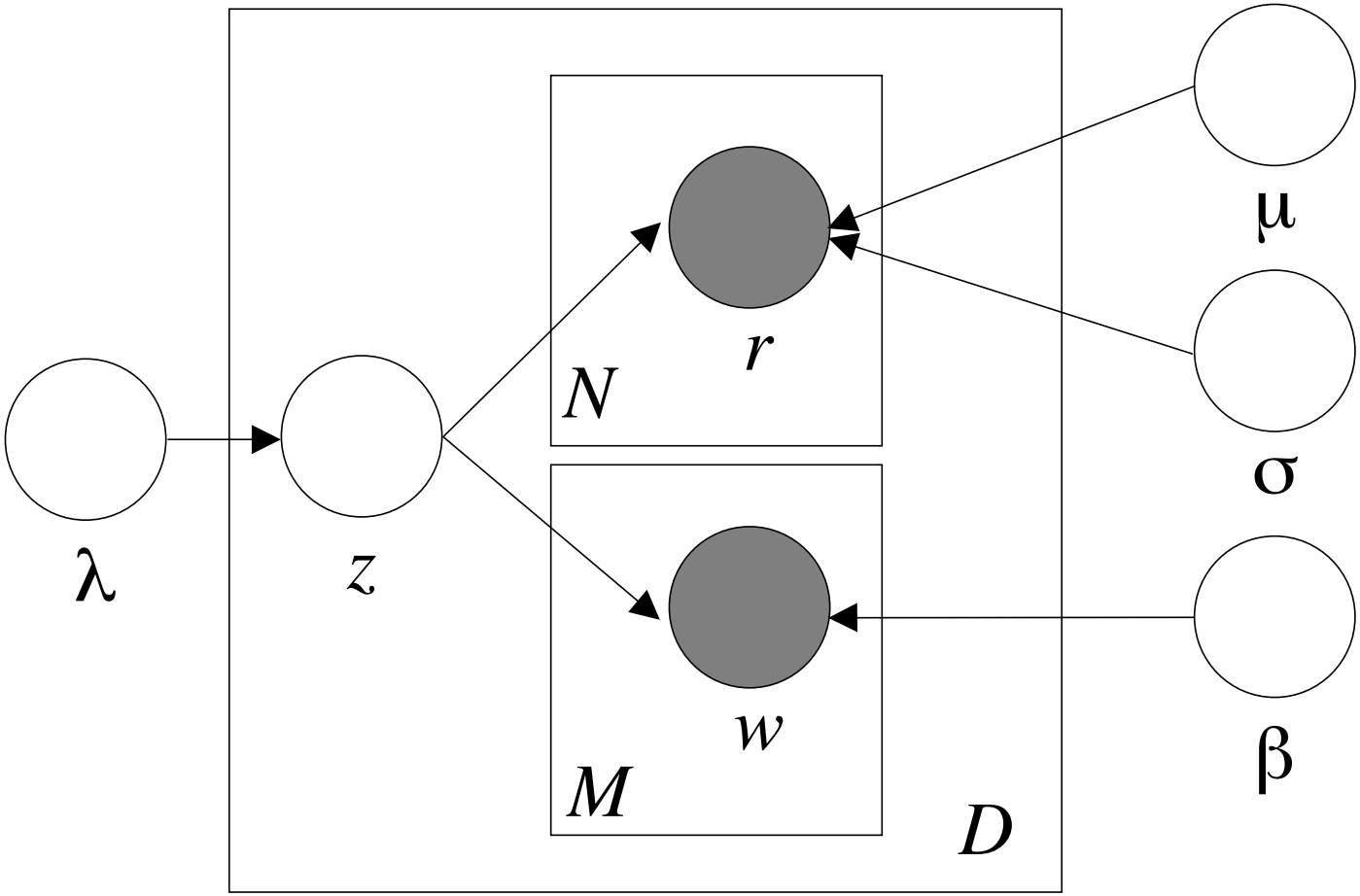


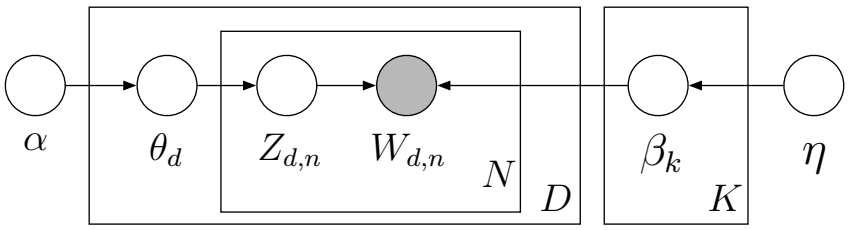
Models

Three hierarchical probabilistic models

1. Gaussian Multinomial mixture
2. Gaussian Multinomial LDA
3. Correspondence LDA

Gaussian Multinomial Mixture





Distributions

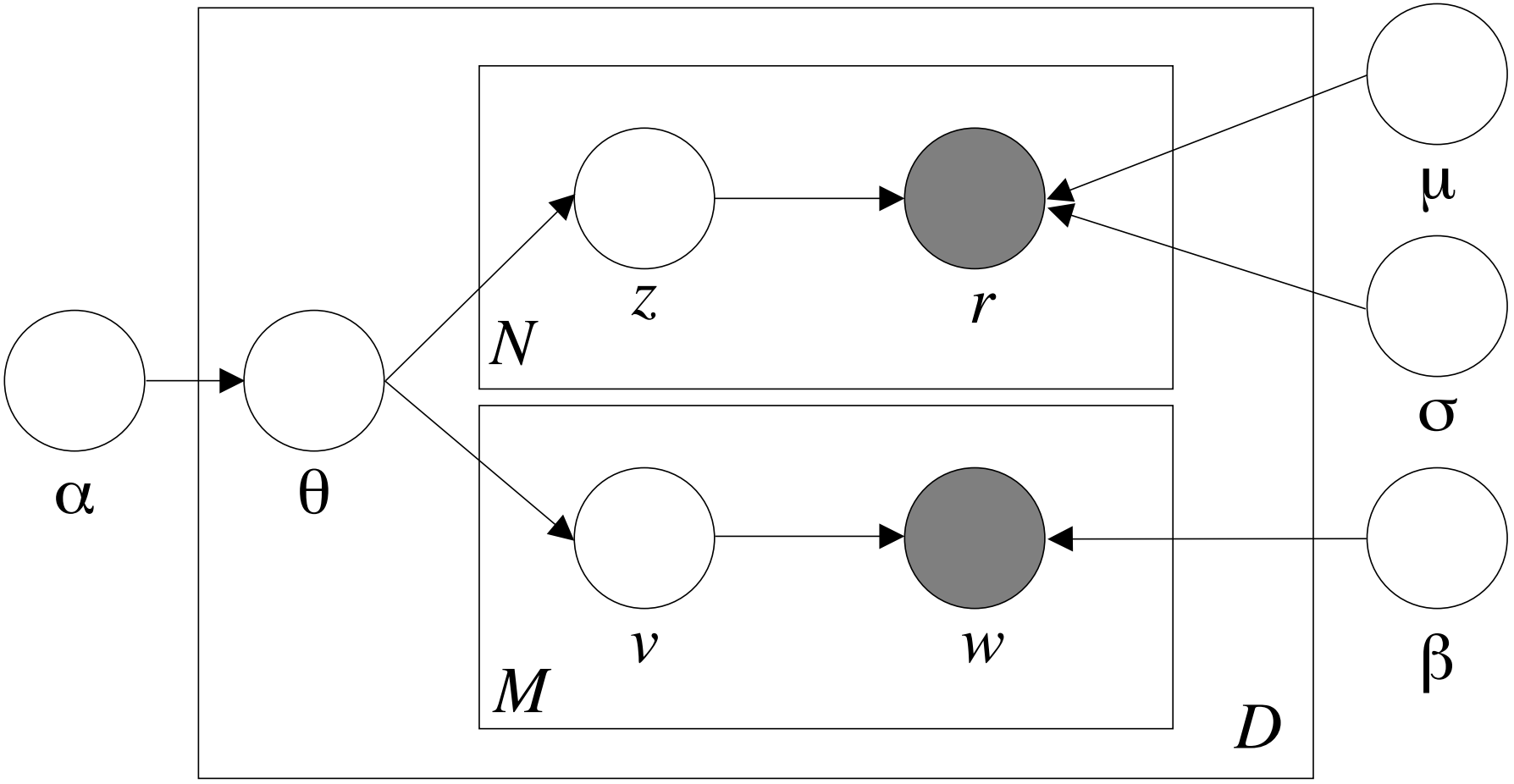
$$p(z, \mathbf{r}, \mathbf{w}) = p(z \mid \lambda) \prod_{n=1}^{N} p(r_n \mid z, \mu, \sigma) \cdot \prod_{m=1}^{M} p(w_m \mid z, \beta).$$

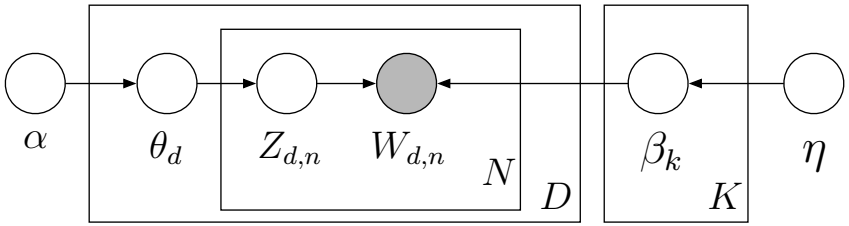
- p(**r**, **w**)
- p(w | **r**) = $\sum_z p(z \mid \mathbf{r}) p(w \mid z)$

But no

- p(w | **r**, r_n)

Gaussian Multinomial LDA





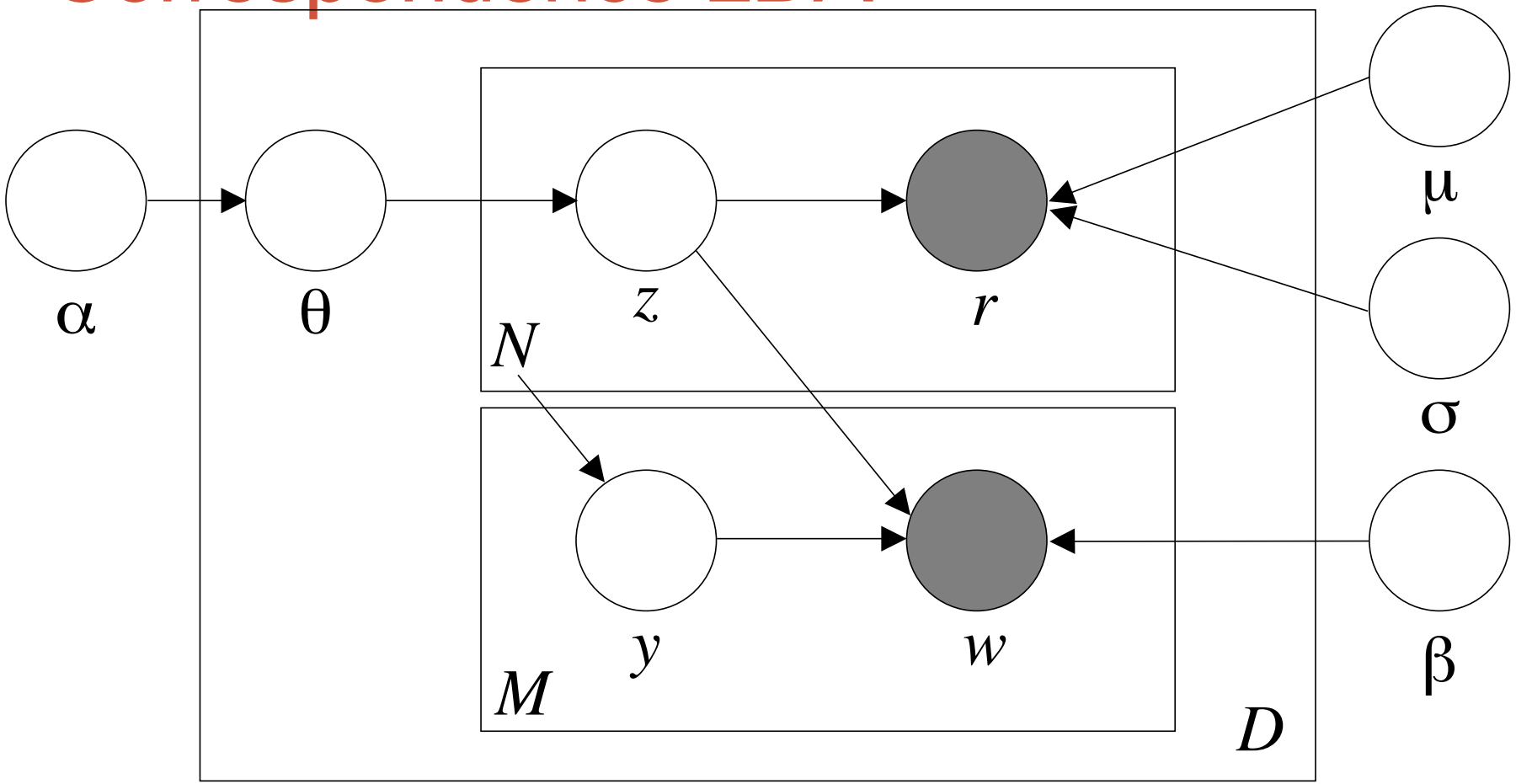
Distributions

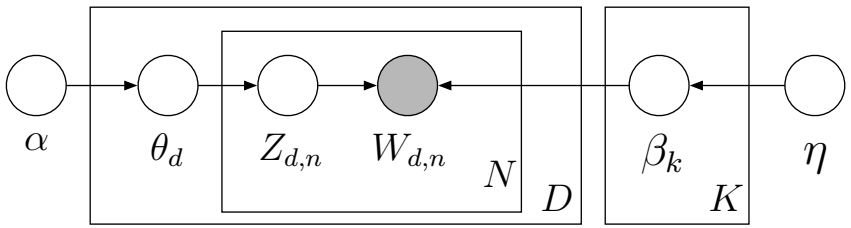
$$p(\mathbf{r}, \mathbf{w}, \theta, \mathbf{z}, \mathbf{v}) = p(\theta \mid \alpha) \left(\prod_{n=1}^{N} p(z_n \mid \theta) p(r_n \mid z_n, \mu, \sigma) \right) \cdot \left(\prod_{m=1}^{M} p(v_m \mid \theta) p(w_m \mid v_m, \beta) \right)$$

All

- p(**r**, **w**)
- p(w | **r**)
- p(w | **r**, r_n)

Correspondence LDA





Distributions

$$p(\mathbf{r}, \mathbf{w}, \theta, \mathbf{z}, \mathbf{y}) = p(\theta \mid \alpha) \left(\prod_{n=1}^{N} p(z_n \mid \theta) p(r_n \mid z_n, \mu, \sigma) \right) \cdot \left(\prod_{m=1}^{M} p(y_m \mid N) p(w_m \mid y_m, \mathbf{z}, \beta) \right)$$

All

- p(**r**, **w**)
- p(w | **r**)
- p(w | **r**, r_n)

Inference & Estimation

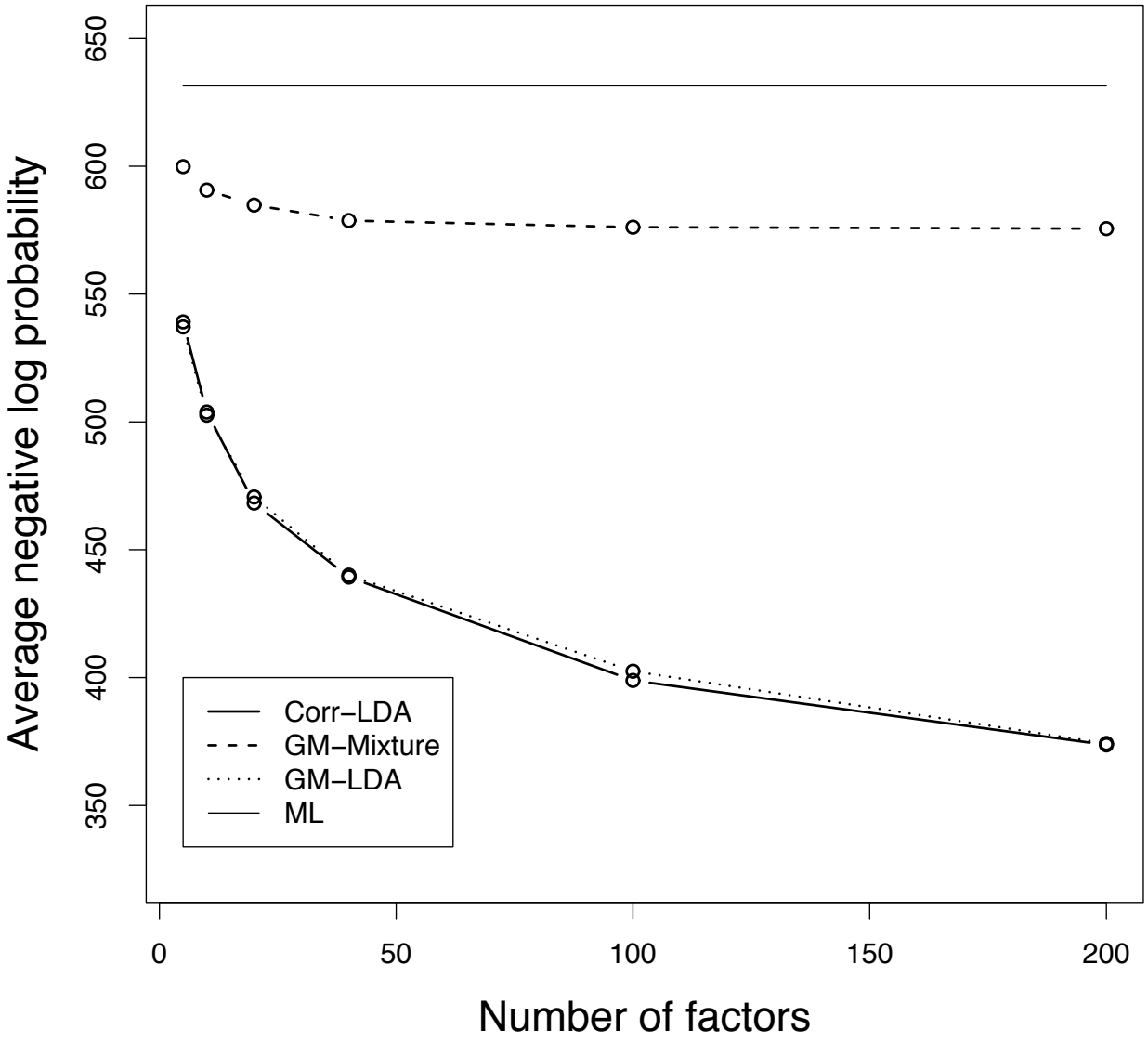
- Variational Inference
  - Exact intractable
  - Approximate assuming factorizable distribution
  - Minimize KL-Divergence via iterative updates to parameters
- Parameter Estimation
  - EM algorithm
  - E: Compute variational posterior.
  - M: MLE estimate of the model parameters.

Evaluation

- 7000 Images and their captions
- 75% Training & 25% Testing

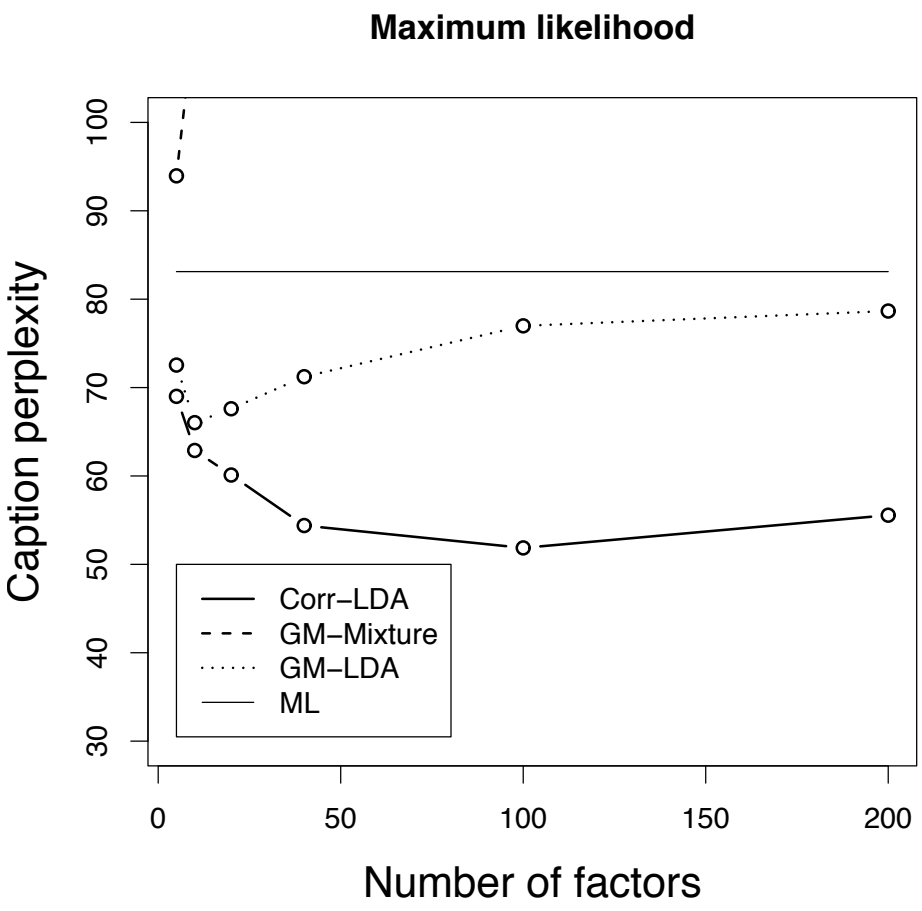
- Test set likelihood
- Automatic annotation
- Text based retrieval

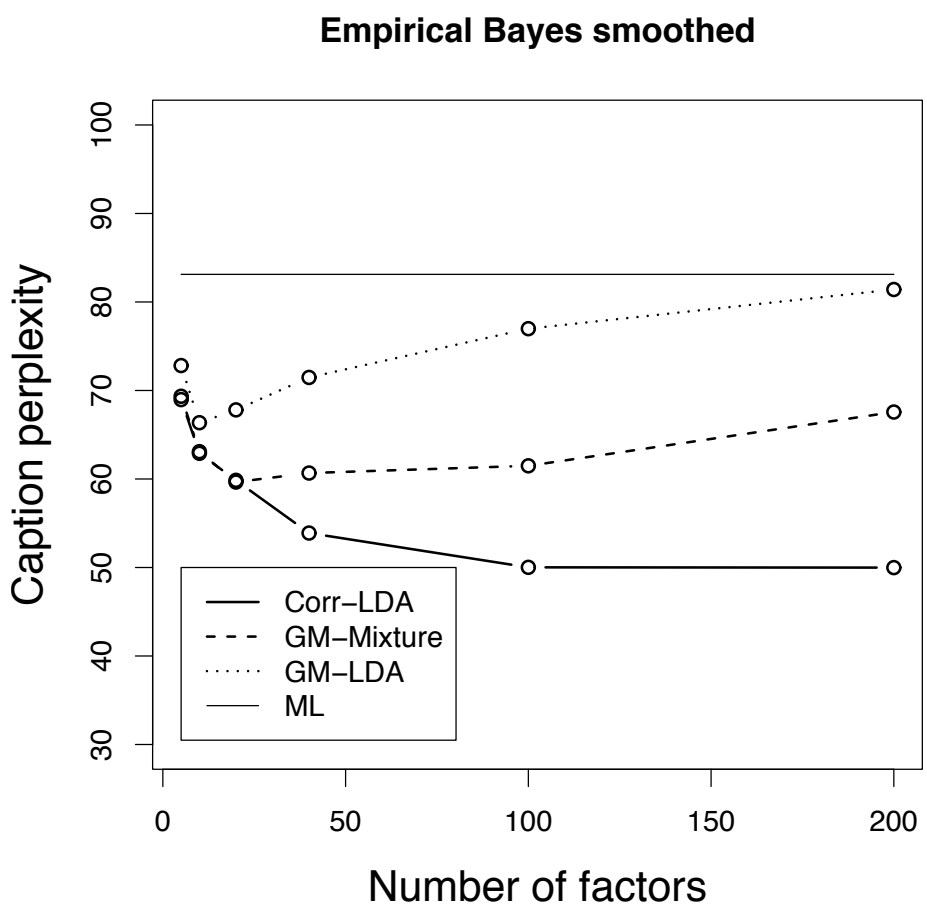
Eval: Test set likelihood



Eval: Automatic Annotation

$$\text{perplexity} = \exp\{-\sum_{d=1}^{D}\sum_{m=1}^{M_d}\log p(w_m \mid \mathbf{r}_d) / \sum_{d=1}^{D} M_d\}$$





Eval: Automatic Annotation (Qual.)

True caption
scotland water

Corr–LDA
scotland water flowers hills tree

GM–LDA
tree water people mountain sky

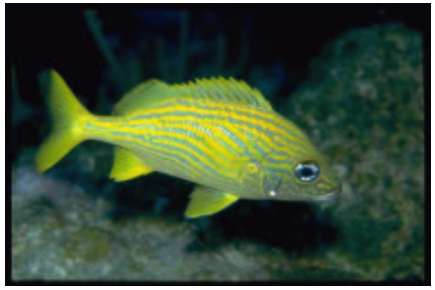
GM–Mixture
water sky clouds sunset scotland

True caption
clouds jet plane

Corr–LDA
sky plane jet mountain clouds

GM–LDA
sky water people tree clouds

GM–Mixture
sky plane jet clouds pattern



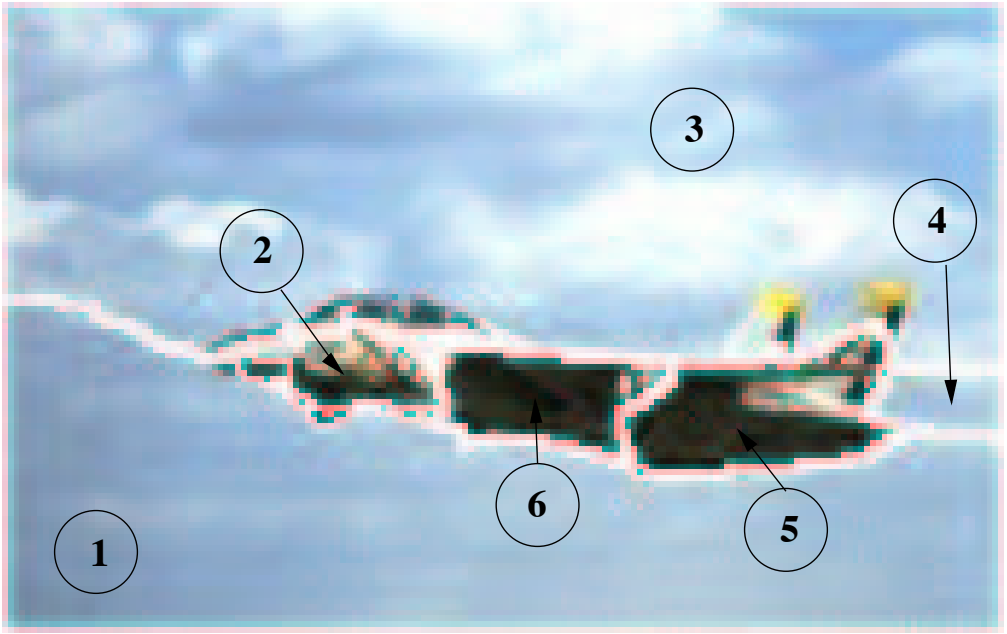
True caption
fish reefs water

Corr–LDA
fish water ocean tree coral

GM–LDA
water sky vegetables tree people

GM–Mixture
fungus mushrooms tree flowers leaves

Eval: Automatic Annotation (Qual.)



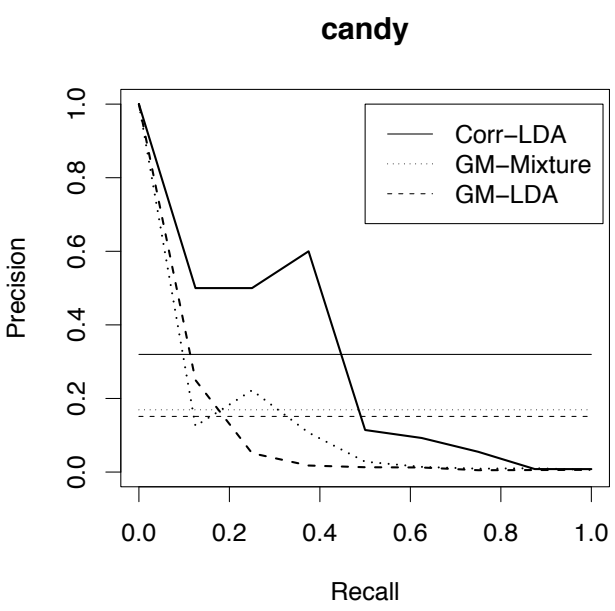
Corr–LDA:

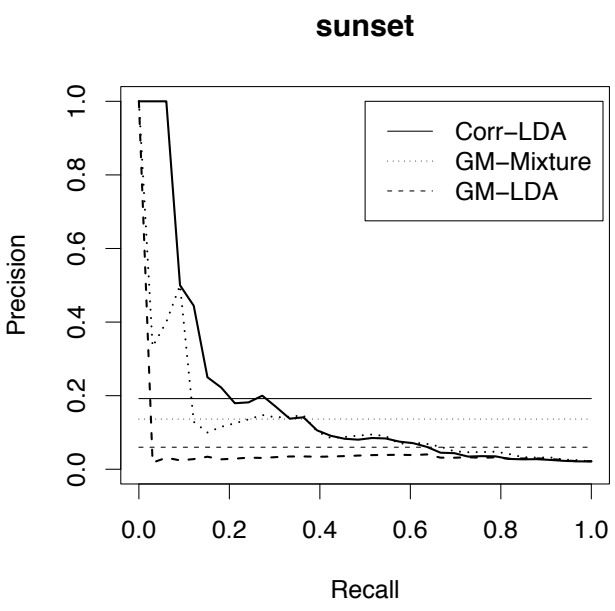
1. PEOPLE, TREE
2. SKY, JET
3. SKY, CLOUDS
4. SKY, MOUNTAIN
5. PLANE, JET
6. PLANE, JET

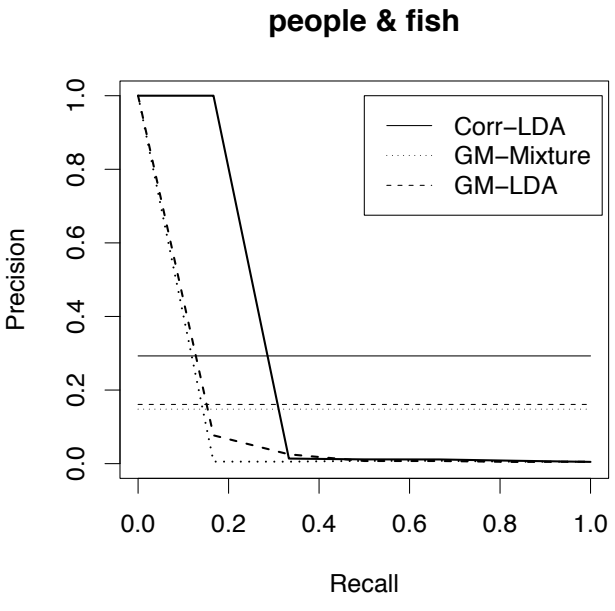
GM–LDA:

1. HOTEL, WATER
2. PLANE, JET
3. TUNDRA, PENGUIN
4. PLANE, JET
5. WATER, SKY
6. BOATS, WATER

Text Based Retrieval

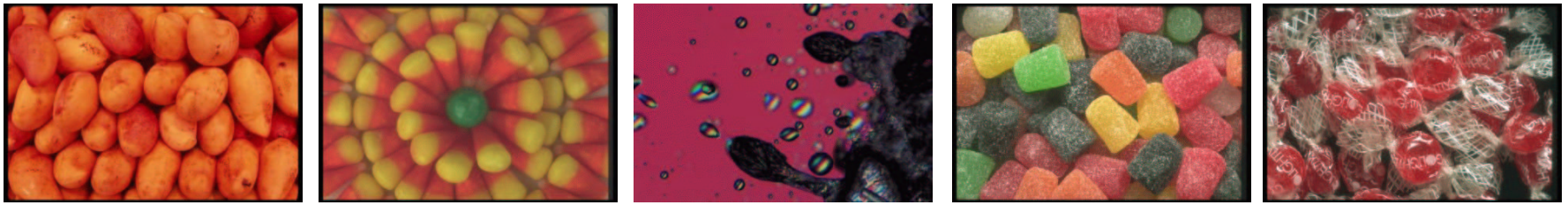




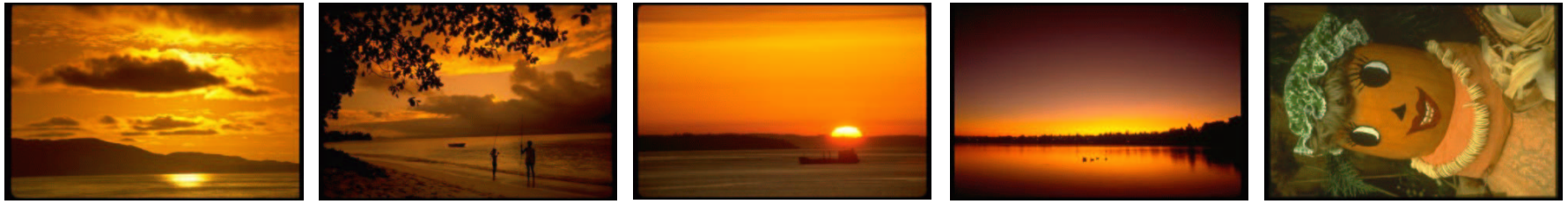


Text Based Retrieval (Qual.)

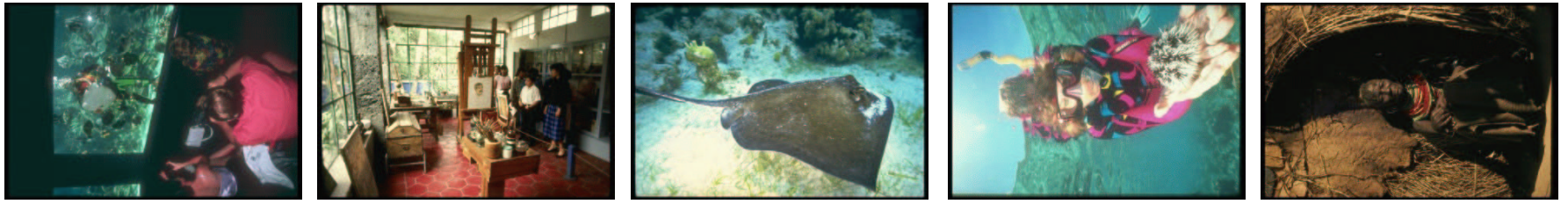
Candy



Sunset



People & Fish



Conclusion

If conditionals are needed, then model them explicitly