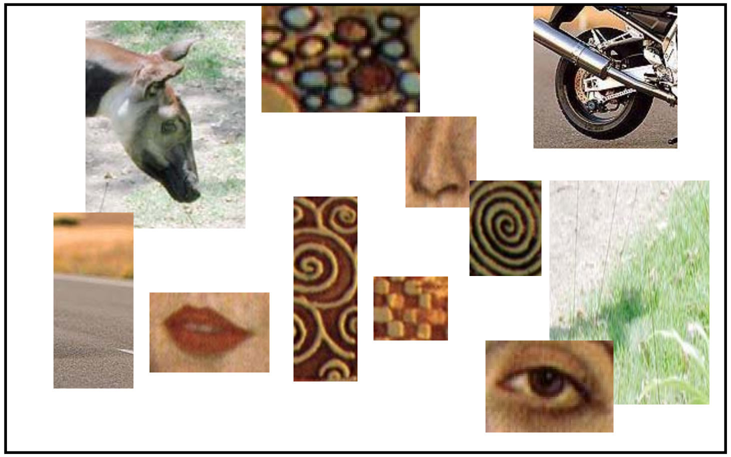
Generative learning methods for bags of features

- Model the probability of a bag of features given a class

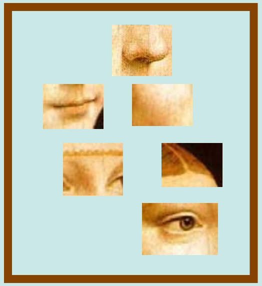


Many slides adapted from Fei-Fei Li, Rob Fergus, and Antonio Torralba

Generative methods

- We will cover two models, both inspired by text document analysis:
  - Naïve Bayes
  - Probabilistic Latent Semantic Analysis

The Naïve Bayes model



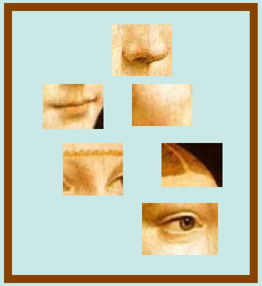
- Assume that each feature is conditionally independent *given the class*

$$p(w_1, \ldots, w_N \mid c) = \prod_{i=1}^{N} p(w_i \mid c)$$

w_i: *i*th feature in the image
N: number of features in the image

Csurka et al. 2004

The Naïve Bayes model



- Assume that each feature is conditionally independent *given the class*

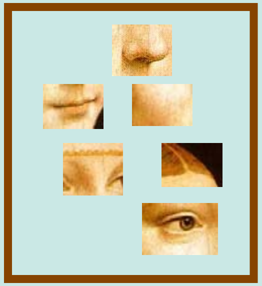
$$p(w_1, \ldots, w_N \mid c) = \prod_{i=1}^{N} p(w_i \mid c) = \prod_{w=1}^{W} p(w \mid c)^{n(w)}$$

w_i: *i*th feature in the image
N: number of features in the image

W: size of visual vocabulary
$n(w)$: number of features with index w in the image

Csurka et al. 2004

The Naïve Bayes model



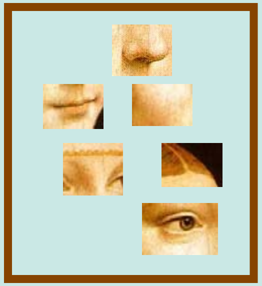
- Assume that each feature is conditionally independent *given the class*

$$p(w_1,\ldots,w_N \mid c) = \prod_{i=1}^{N} p(w_i \mid c) = \prod_{w=1}^{W} p(w \mid c)^{n(w)}$$

$$p(w \mid c) = \frac{\text{No. of features of type } w \text{ in training images of class } c}{\text{Total no. of features in training images of class } c}$$

Csurka et al. 2004

The Naïve Bayes model



- Assume that each feature is conditionally independent *given the class*

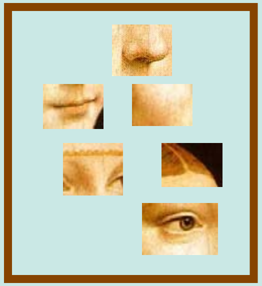
$$p(w_1,\ldots,w_N \mid c) = \prod_{i=1}^{N} p(w_i \mid c) = \prod_{w=1}^{W} p(w \mid c)^{n(w)}$$

$$p(w \mid c) = \frac{\text{No. of features of type } w \text{ in training images of class } c + 1}{\text{Total no. of features in training images of class } c + W}$$

(Laplace smoothing to avoid zero counts)

Csurka et al. 2004

The Naïve Bayes model



- MAP decision:

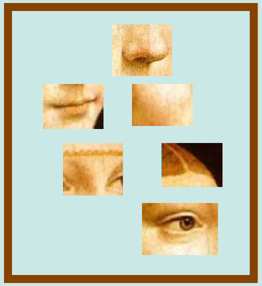
$$c^* = \arg\max_c \; p(c) \prod_{w=1}^{W} p(w \mid c)^{n(w)}$$

$$= \arg\max_c \; \log p(c) + \sum_{w=1}^{W} n(w) \log p(w \mid c)$$

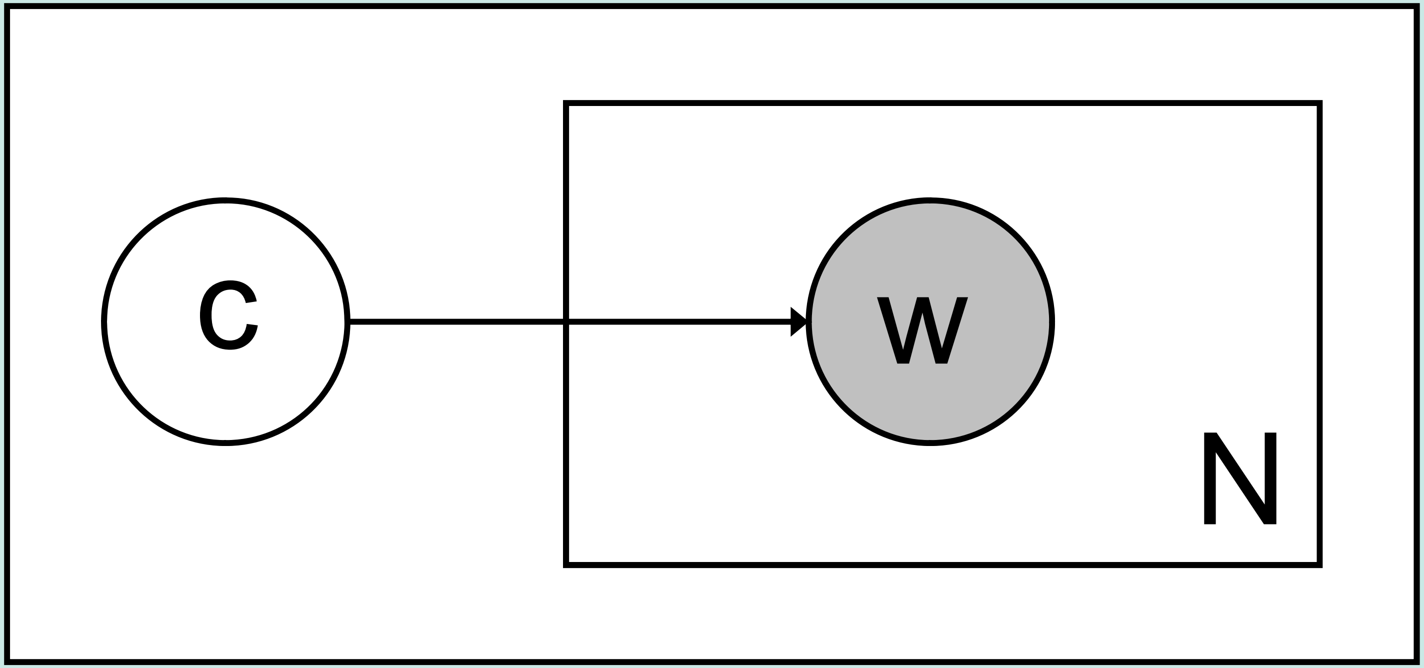
(you should compute the log of the likelihood instead of the likelihood itself in order to avoid underflow)

Csurka et al. 2004

The Naïve Bayes model

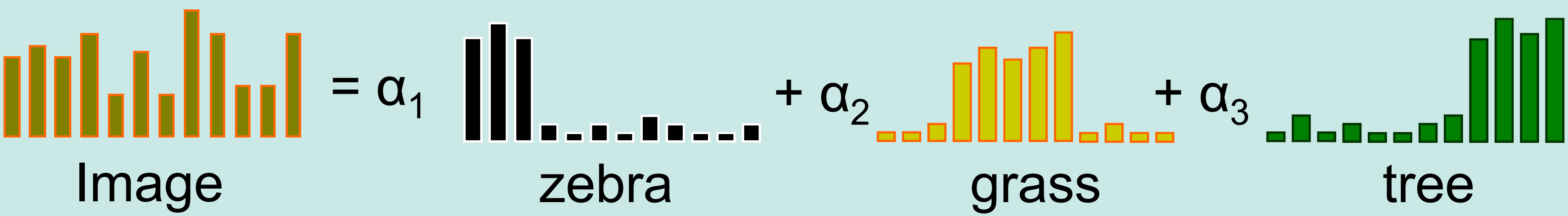


- "Graphical model":



Csurka et al. 2004

Probabilistic Latent Semantic Analysis

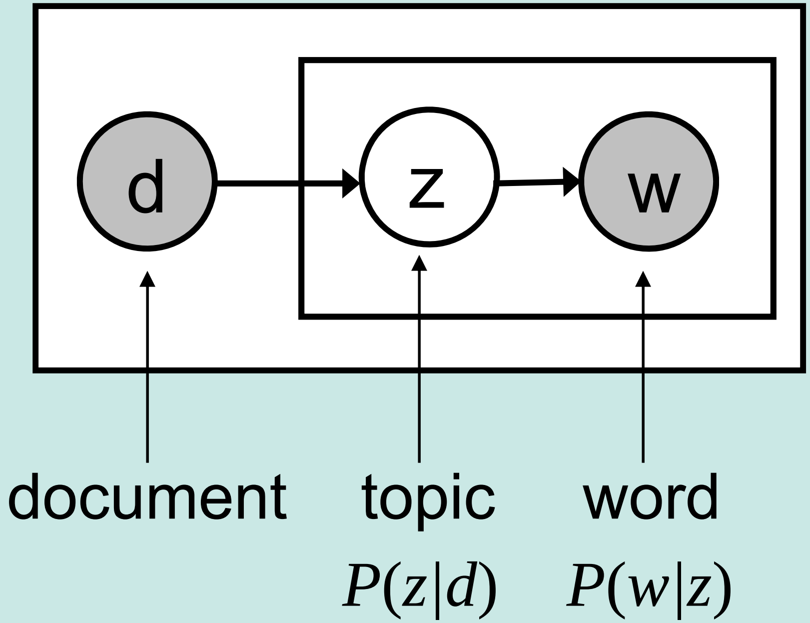


“visual topics”

T. Hofmann, Probabilistic Latent Semantic Analysis, UAI 1999

Probabilistic Latent Semantic Analysis

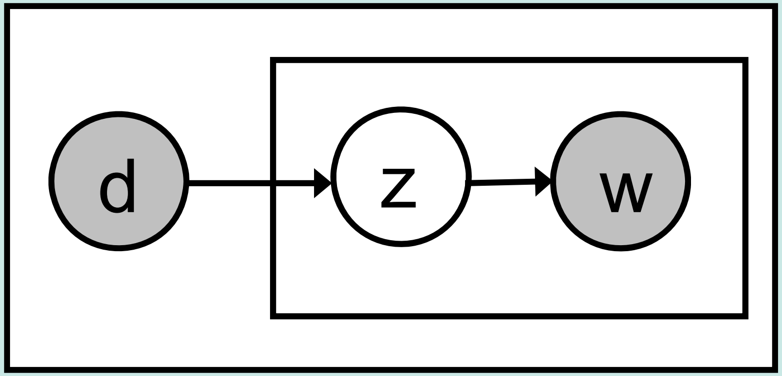
- Unsupervised technique
- Two-level generative model: a document is a mixture of topics, and each topic has its own characteristic word distribution



T. Hofmann, Probabilistic Latent Semantic Analysis, UAI 1999

Probabilistic Latent Semantic Analysis

- Unsupervised technique
- Two-level generative model: a document is a mixture of topics, and each topic has its own characteristic word distribution



$$p(w_i \mid d_j) = \sum_{k=1}^{K} p(w_i \mid z_k)\, p(z_k \mid d_j)$$

T. Hofmann, Probabilistic Latent Semantic Analysis, UAI 1999

The pLSA model

$$p(w_i \mid d_j) = \sum_{k=1}^{K} p(w_i \mid z_k)\, p(z_k \mid d_j)$$

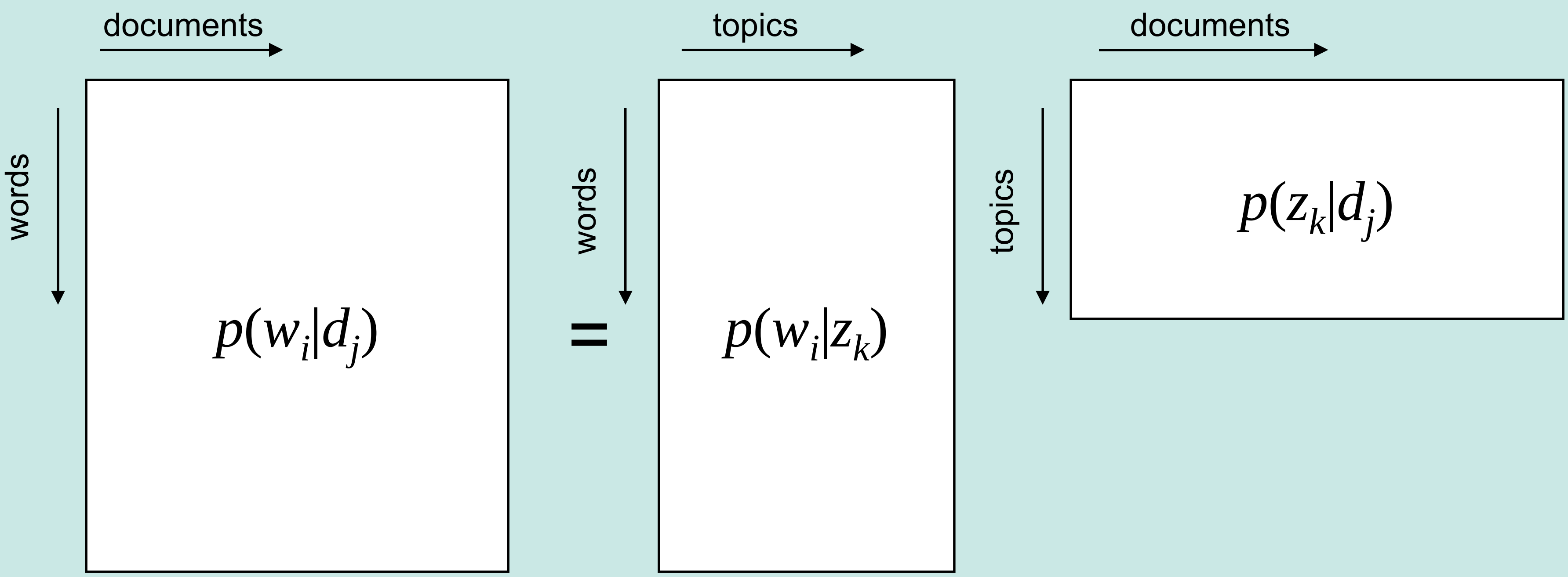
Probability of word i in document j (known)

Probability of word i given topic k (unknown)

Probability of topic k given document j (unknown)

The pLSA model

$$p(w_i \mid d_j) = \sum_{k=1}^{K} p(w_i \mid z_k)\, p(z_k \mid d_j)$$



Observed codeword distributions (M N)

Codeword distributions per topic (class) (M K)

Class distributions per image (K N)

Learning pLSA parameters

Maximize likelihood of data:

Observed counts of word *i* in document *j*

$$L = \prod_{i=1}^{M} \prod_{j=1}^{N} P(w_i|d_j)^{n(w_i,d_j)}$$

M … number of codewords

N … number of images

$$\sum_{k=1}^{K} P(z_k|d_j) P(w_i|z_k)$$

Slide credit: Josef Sivic

Inference

- Finding the most likely topic (class) for an image:

$$z^* = \arg\max_{z} p(z \mid d)$$

Inference

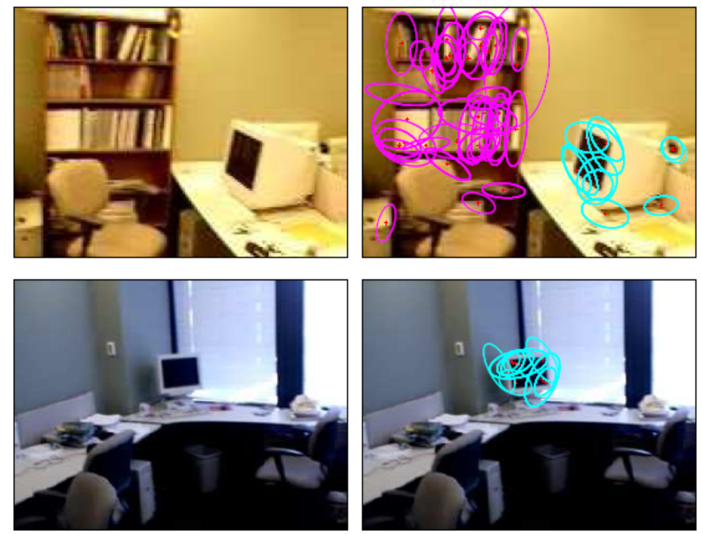
- Finding the most likely topic (class) for an image:

$$z^* = \arg\max_z p(z \mid d)$$

- Finding the most likely topic (class) for a visual word in a given image:

$$z^* = \arg\max_z p(z \mid w, d) = \arg\max_z \frac{p(w \mid z)\, p(z \mid d)}{\sum_{z'} p(w \mid z')\, p(z' \mid d)}$$

Topic discovery in images



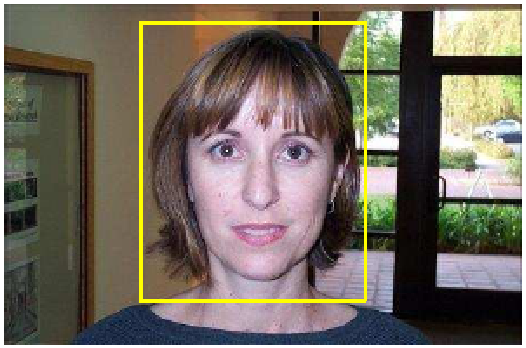
J. Sivic, B. Russell, A. Efros, A. Zisserman, B. Freeman, **Discovering Objects and their Location in Images**, *ICCV 2005*

From single features to "doublets"

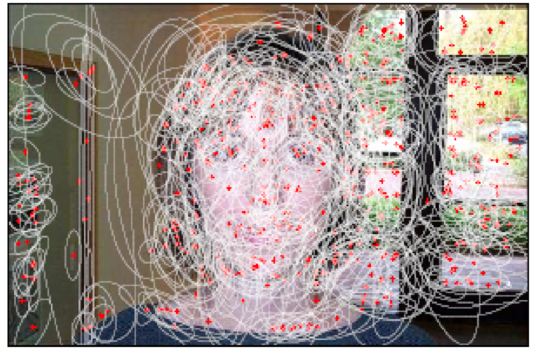
1. Run pLSA on a regular visual vocabulary
2. Identify a small number of top visual words for each topic
3. Form a "doublet" vocabulary from these top visual words
4. Run pLSA again on the augmented vocabulary

J. Sivic, B. Russell, A. Efros, A. Zisserman, B. Freeman, **Discovering Objects and their Location in Images**, *ICCV 2005*

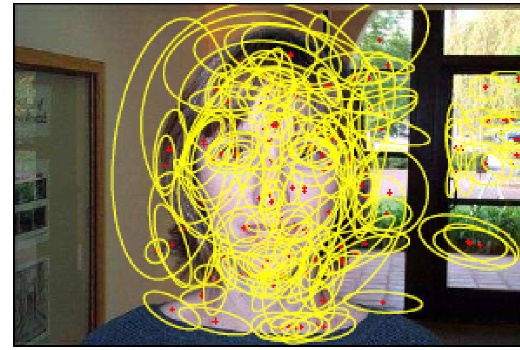
From single features to “doublets”



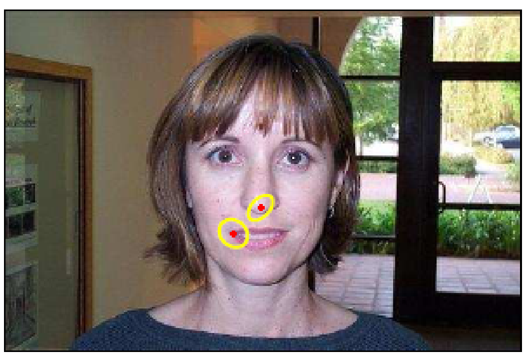
Ground truth



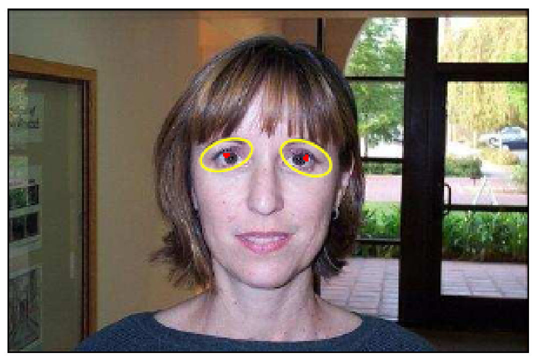
All features



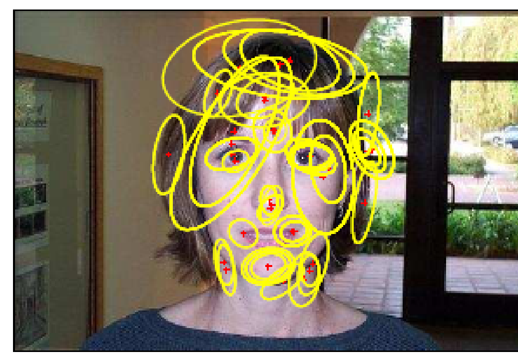
“Face” features initially found by pLSA



One doublet



Another doublet



“Face” doublets

J. Sivic, B. Russell, A. Efros, A. Zisserman, B. Freeman, **Discovering Objects and their Location in Images**, *ICCV 2005*

Summary: Generative models

- Naïve Bayes
  - *Unigram models* in document analysis
  - Assumes conditional independence of words given class
  - Parameter estimation: frequency counting
- Probabilistic Latent Semantic Analysis
  - Unsupervised technique
  - Each document is a mixture of topics (image is a mixture of classes)
  - Can be thought of as matrix decomposition
  - Parameter estimation: Expectation-Maximization