Hidden Variables, the EM Algorithm, and Mixtures of Gaussians

Computer Vision

CS 143, Brown

James Hays

Many slides from Derek Hoiem

Today’s Class

- Examples of Missing Data Problems
  - Detecting outliers
- Background
  - Maximum Likelihood Estimation
  - Probabilistic Inference
- Dealing with “Hidden” Variables
  - EM algorithm, Mixture of Gaussians
  - Hard EM

Slide: Derek Hoiem

Missing Data Problems: Outliers

You want to train an algorithm to predict whether a photograph is attractive. You collect annotations from Mechanical Turk. Some annotators try to give accurate ratings, but others answer randomly.

Challenge: Determine which people to trust and the average rating by accurate annotators.

Annotator Ratings

10
8
9
2
8

Photo: Jam343 (Flickr)

Missing Data Problems: Object Discovery

You have a collection of images and have extracted regions from them. Each is represented by a histogram of "visual words".

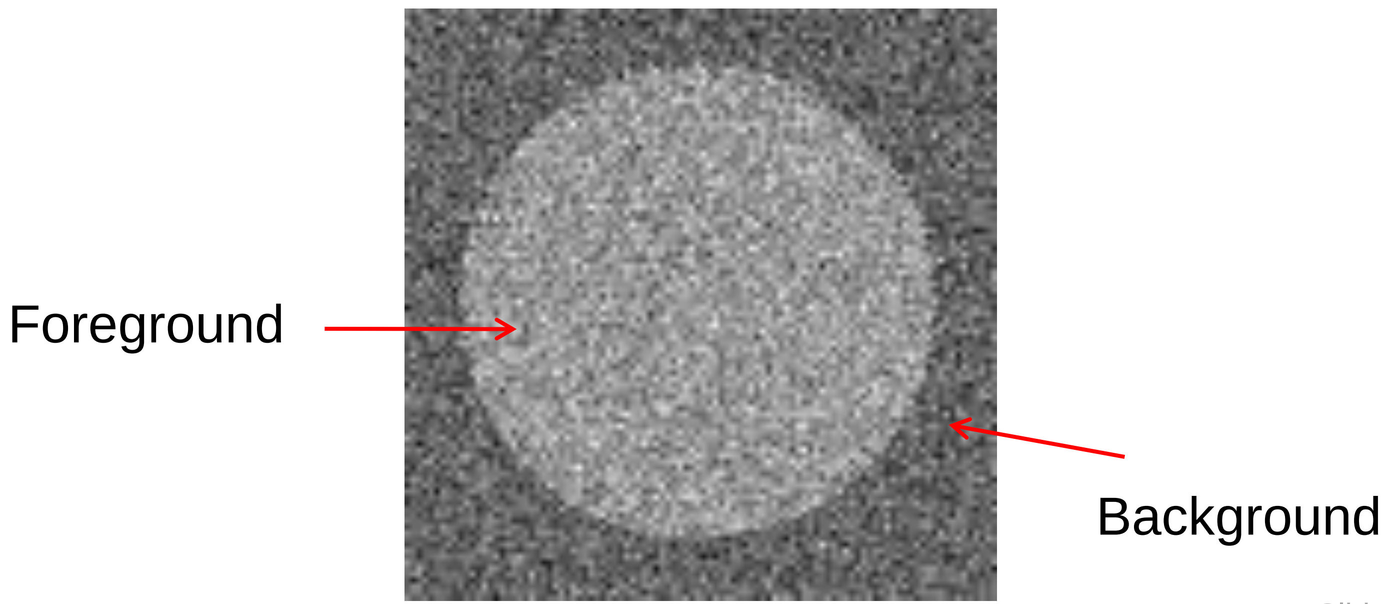
Challenge: Discover frequently occurring object categories, without pre-trained appearance models.

http://www.robots.ox.ac.uk/~vgg/publications/papers/russell06.pdf

Missing Data Problems: Segmentation

You are given an image and want to assign foreground/background pixels.

Challenge: Segment the image into figure and ground without knowing what the foreground looks like in advance.



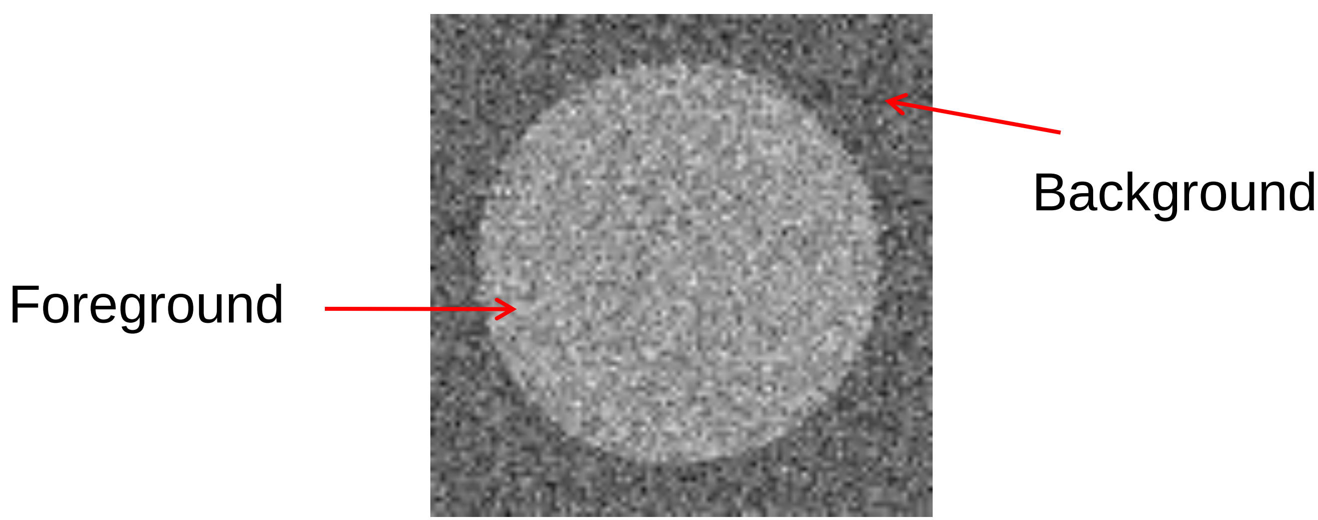
Slide: Derek Hoiem

Missing Data Problems: Segmentation

Challenge: Segment the image into figure and ground without knowing what the foreground looks like in advance.

Three steps:

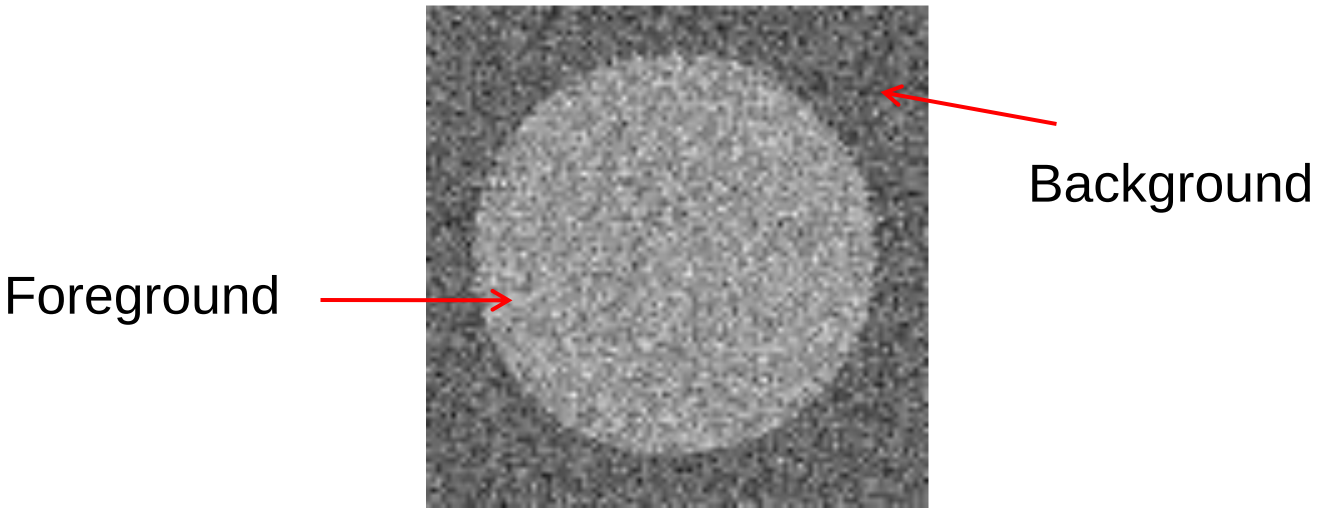
1. If we had labels, how could we model the appearance of foreground and background?
2. Once we have modeled the fg/bg appearance, how do we compute the likelihood that a pixel is foreground?
3. How can we get both labels and appearance models at once?



Slide: Derek Hoiem

Maximum Likelihood Estimation

1. If we had labels, how could we model the appearance of foreground and background?



Slide: Derek Hoiem

Maximum Likelihood Estimation

data

$$\mathbf{x} = \{x_1 \ldots x_N\}$$

parameters

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\; p(\mathbf{x} \mid \theta)$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \prod_n p(x_n \mid \theta)$$

Slide: Derek Hoiem

Maximum Likelihood Estimation

$$\mathbf{x} = \{x_1 .. x_N\}$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ p(\mathbf{x} \mid \theta)$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \prod_n p(x_n \mid \theta)$$

Gaussian Distribution

$$p(x_n \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_n - \mu)^2}{2\sigma^2} \right)$$

Slide: Derek Hoiem

Maximum Likelihood Estimation

$$\mathbf{x} = \{x_1 .. x_N\}$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\; p(\mathbf{x} \mid \theta)$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \prod_n p(x_n \mid \theta)$$

Gaussian Distribution

$$p(x_n \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_n - \mu)^2}{2\sigma^2} \right)$$

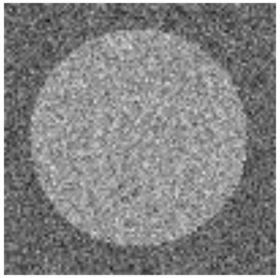
$$\hat{\mu} = \frac{1}{N} \sum_n x_n \qquad \hat{\sigma}^2 = \frac{1}{N} \sum_n (x_n - \hat{\mu})^2$$

Slide: Derek Hoiem

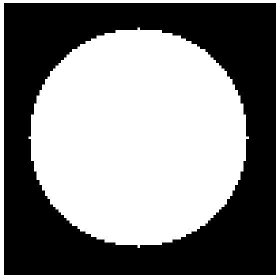
Example: MLE

Parameters used to Generate

```
fg: mu=0.6, sigma=0.1
bg: mu=0.4, sigma=0.1
```



im

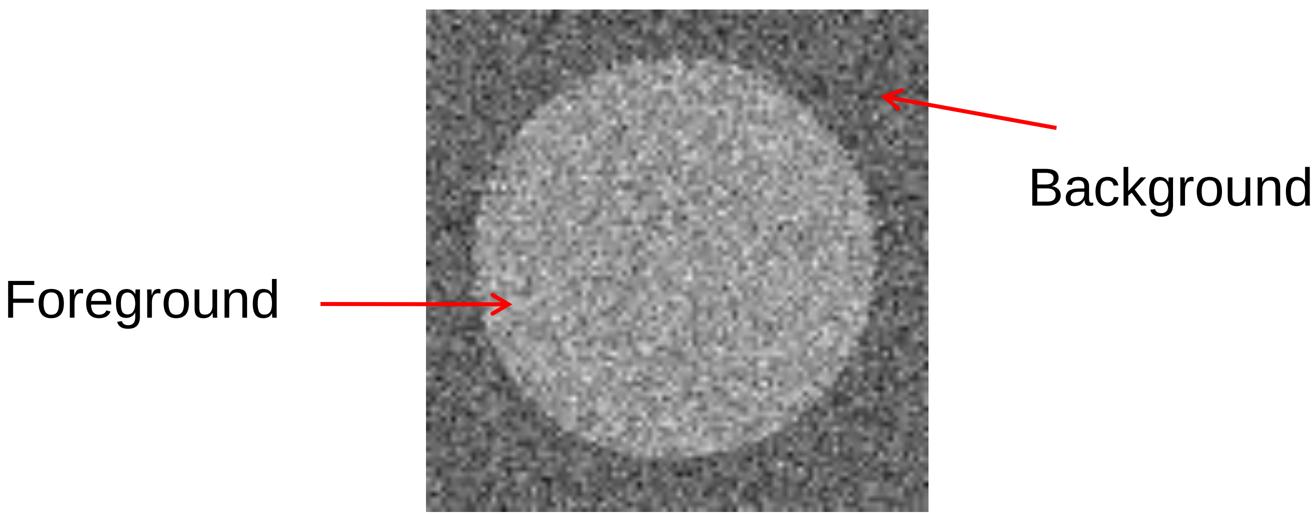


labels

```
>> mu_fg = mean(im(labels))
      mu_fg = 0.6012

>> sigma_fg = sqrt(mean((im(labels)-mu_fg).^2))
      sigma_fg = 0.1007

>> mu_bg = mean(im(~labels))
      mu_bg = 0.4007

>> sigma_bg = sqrt(mean((im(~labels)-mu_bg).^2))
      sigma_bg = 0.1007

>> pfg = mean(labels(:));
```

Probabilistic Inference

2. Once we have modeled the fg/bg appearance, how do we compute the likelihood that a pixel is foreground?



Slide: Derek Hoiem

Probabilistic Inference

Compute the likelihood that a particular model generated a sample

component or label

$$p(z_n = m \mid x_n, \theta)$$

Probabilistic Inference

Compute the likelihood that a particular model generated a sample

component or label

$$p(z_n = m \mid x_n, \theta) = \frac{p\left(z_n = m, x_n \mid \theta_m\right)}{p\left(x_n \mid \theta\right)}$$

Probabilistic Inference

Compute the likelihood that a particular model generated a sample

component or label

$$p(z_n = m \mid x_n, \theta) = \frac{p\left(z_n = m, x_n \mid \theta_m\right)}{p\left(x_n \mid \theta\right)}$$

$$= \frac{p\left(z_n = m, x_n \mid \theta_m\right)}{\sum_k p\left(z_n = k, x_n \mid \theta_k\right)}$$

Probabilistic Inference

Compute the likelihood that a particular model generated a sample

component or label

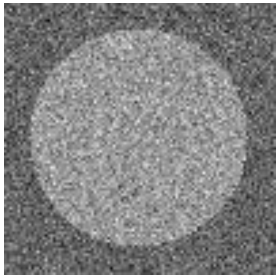
$$p(z_n = m \mid x_n, \theta) = \frac{p(z_n = m, x_n \mid \theta_m)}{p(x_n \mid \theta)}$$

$$= \frac{p(z_n = m, x_n \mid \theta_m)}{\sum_k p(z_n = k, x_n \mid \theta_k)}$$

$$= \frac{p(x_n \mid z_n = m, \theta_m)\, p(z_n = m \mid \theta_m)}{\sum_k p(x_n \mid z_n = k, \theta_k)\, p(z_n = k \mid \theta_k)}$$

Slide: Derek Hoiem

Example: Inference

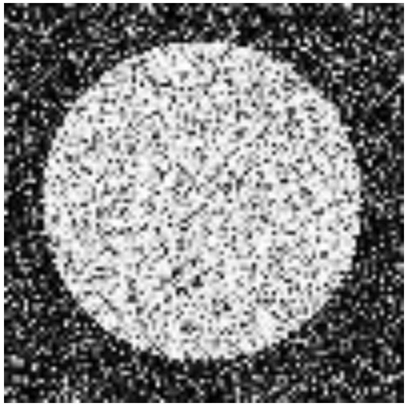


im

Learned Parameters

```
fg: mu=0.6, sigma=0.1
bg: mu=0.4, sigma=0.1
```

```
>> pfg = 0.5;
>> px_fg = normpdf(im, mu_fg, sigma_fg);
>> px_bg = normpdf(im, mu_bg, sigma_bg);
>> pfg_x = px_fg*pfg ./ (px_fg*pfg + px_bg*(1-pfg));
```



p(fg | im)

Mixture of Gaussian* Example: Matting

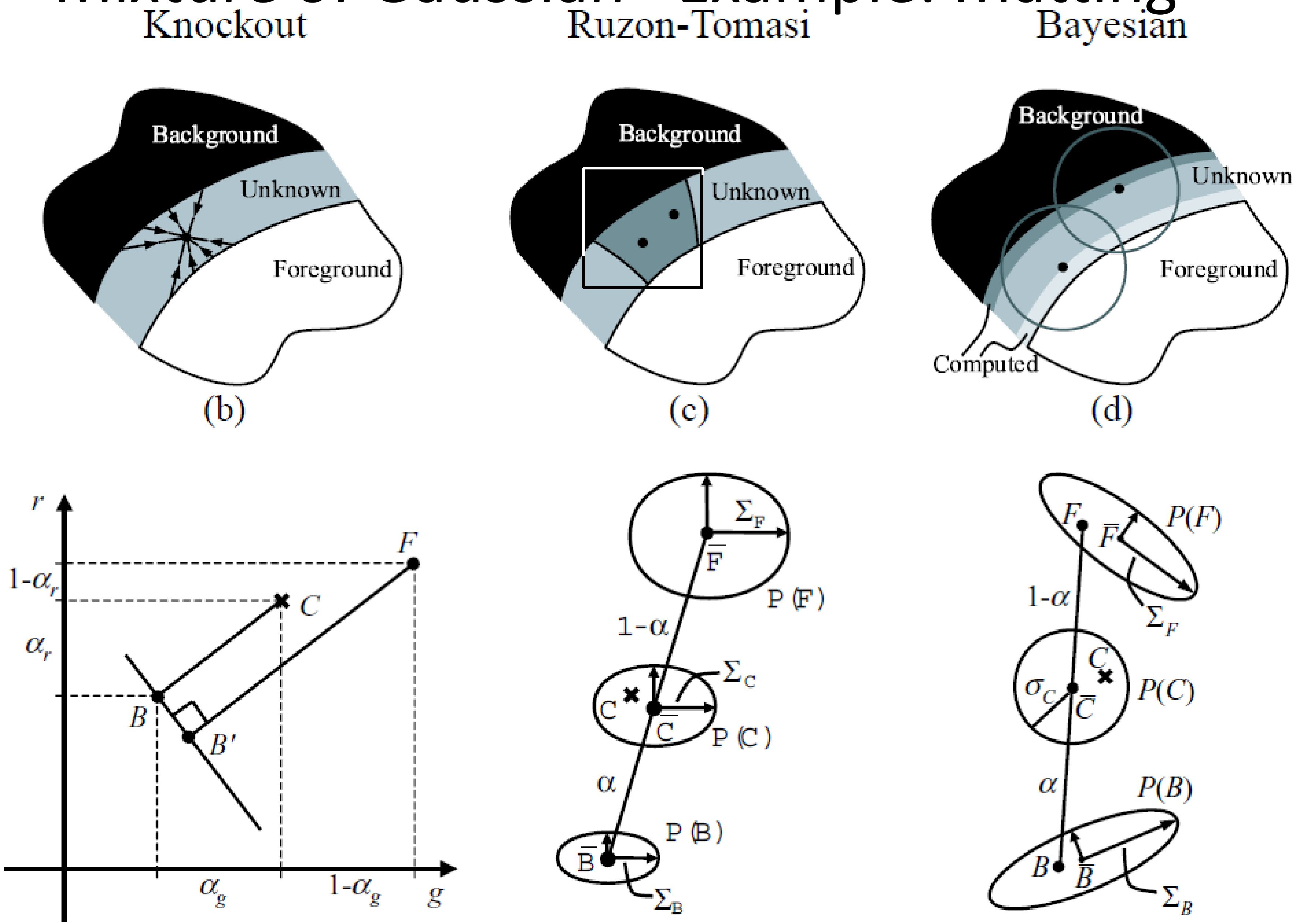
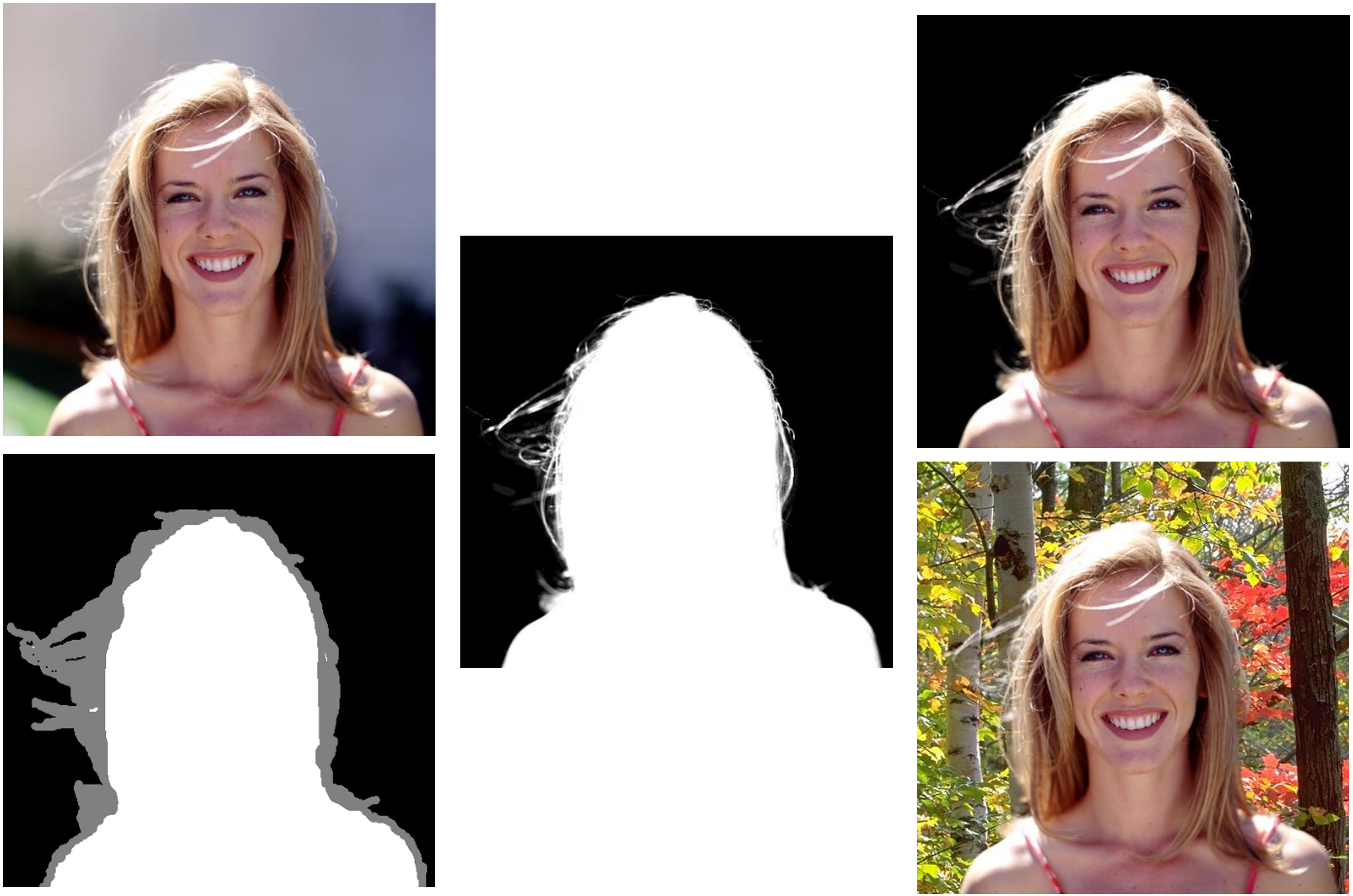


Figure from “Bayesian Matting”, Chuang et al. 2001

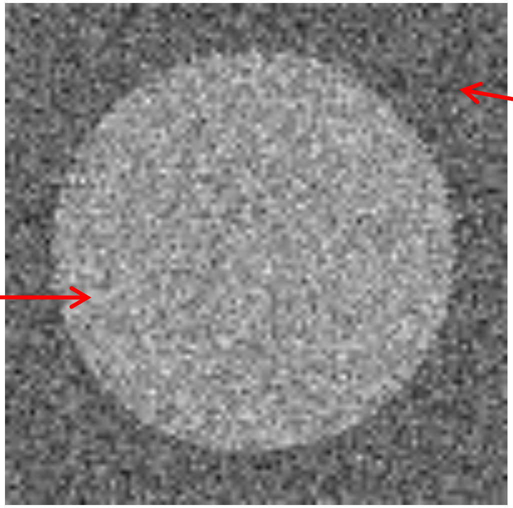
Mixture of Gaussian* Example: Matting



Result from “Bayesian Matting”, Chuang et al. 2001

Dealing with Hidden Variables

3. How can we get both labels and appearance models at once?

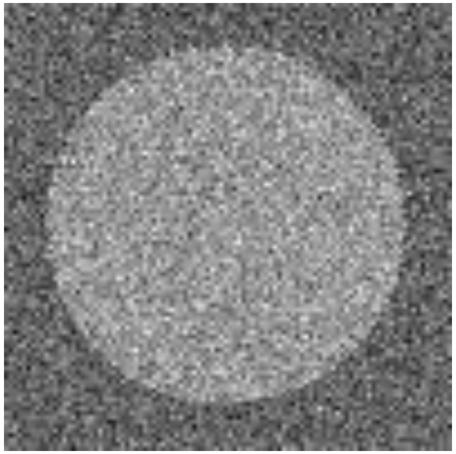


Slide: Derek Hoiem

Segmentation with Mixture of Gaussians

Pixels come from one of several Gaussian components

- We don’t know which pixels come from which components
- We don’t know the parameters for the components



Slide: Derek Hoiem

Simple solution

1. Initialize parameters

2. Compute the probability of each hidden variable given the current parameters

3. Compute new parameters for each model, weighted by likelihood of hidden variables

4. Repeat 2-3 until convergence

Mixture of Gaussians: Simple Solution

1. Initialize parameters

2. Compute likelihood of hidden variables for current parameters

$$\alpha_{nm} = p(z_n = m \mid x_n, \boldsymbol{\mu}^{(t)}, \boldsymbol{\sigma}^{2(t)}, \boldsymbol{\pi}^{(t)})$$

3. Estimate new parameters for each model, weighted by likelihood

$$\hat{\mu}_m^{(t+1)} = \frac{1}{\sum_n \alpha_{nm}} \sum_n \alpha_{nm} x_n \qquad \hat{\sigma}_m^{2(t+1)} = \frac{1}{\sum_n \alpha_{nm}} \sum_n \alpha_{nm} (x_n - \hat{\mu}_m)^2 \qquad \hat{\pi}_m^{(t+1)} = \frac{\sum_n \alpha_{nm}}{N}$$

Slide: Derek Hoiem

Expectation Maximization (EM) Algorithm

Goal: $\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \log\left(\sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z} \mid \theta) \right)$

1. E-step: compute

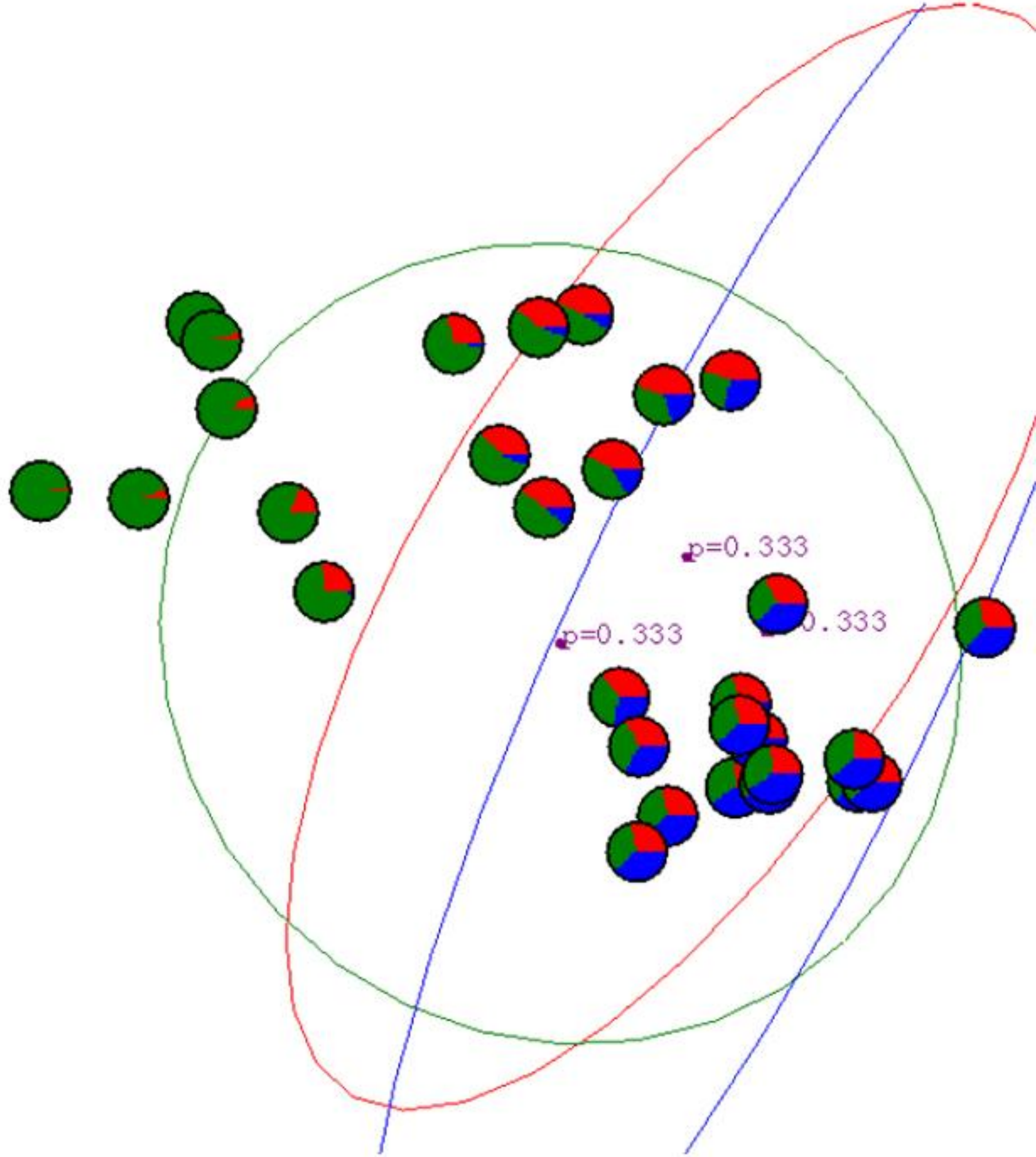
$$\mathrm{E}_{z|x,\theta^{(t)}}\left[\log\left(p(\mathbf{x}, \mathbf{z} \mid \theta)\right)\right] = \sum_{\mathbf{z}} \log\left(p(\mathbf{x}, \mathbf{z} \mid \theta)\right) p(\mathbf{z} \mid \mathbf{x}, \theta^{(t)})$$

2. M-step: solve

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} \sum_{\mathbf{z}} \log\left(p(\mathbf{x}, \mathbf{z} \mid \theta)\right) p(\mathbf{z} \mid \mathbf{x}, \theta^{(t)})$$

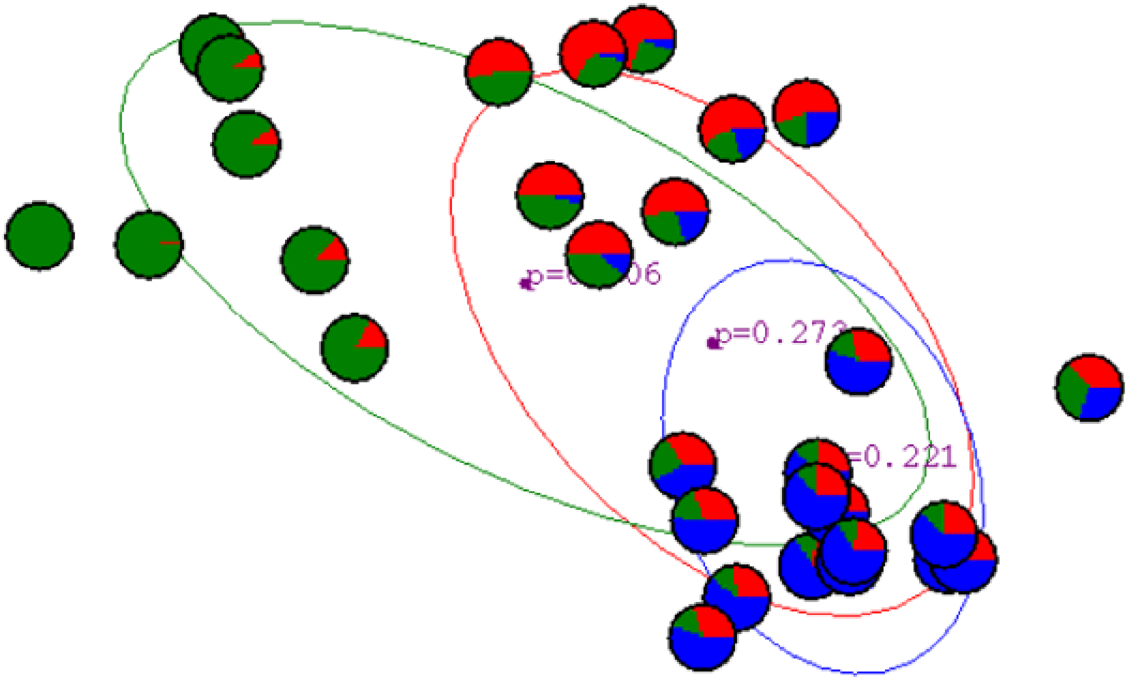
Slide: Derek Hoiem

Gaussian Mixture Example: Start

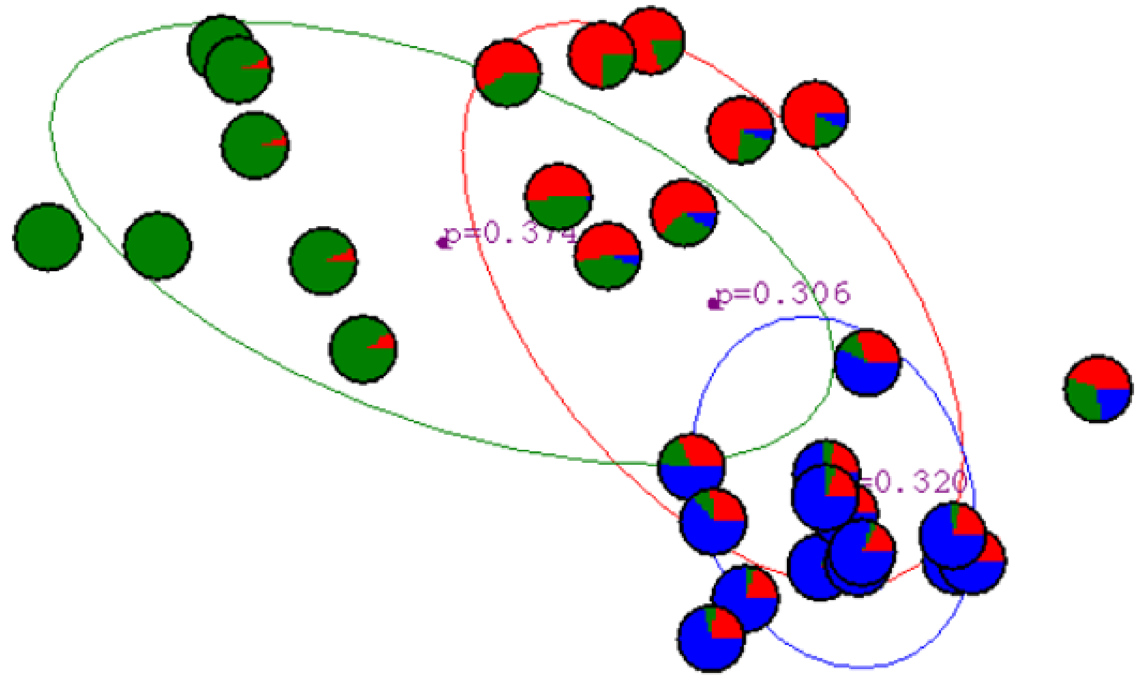


Advance apologies: in Black and White this example will be incomprehensible

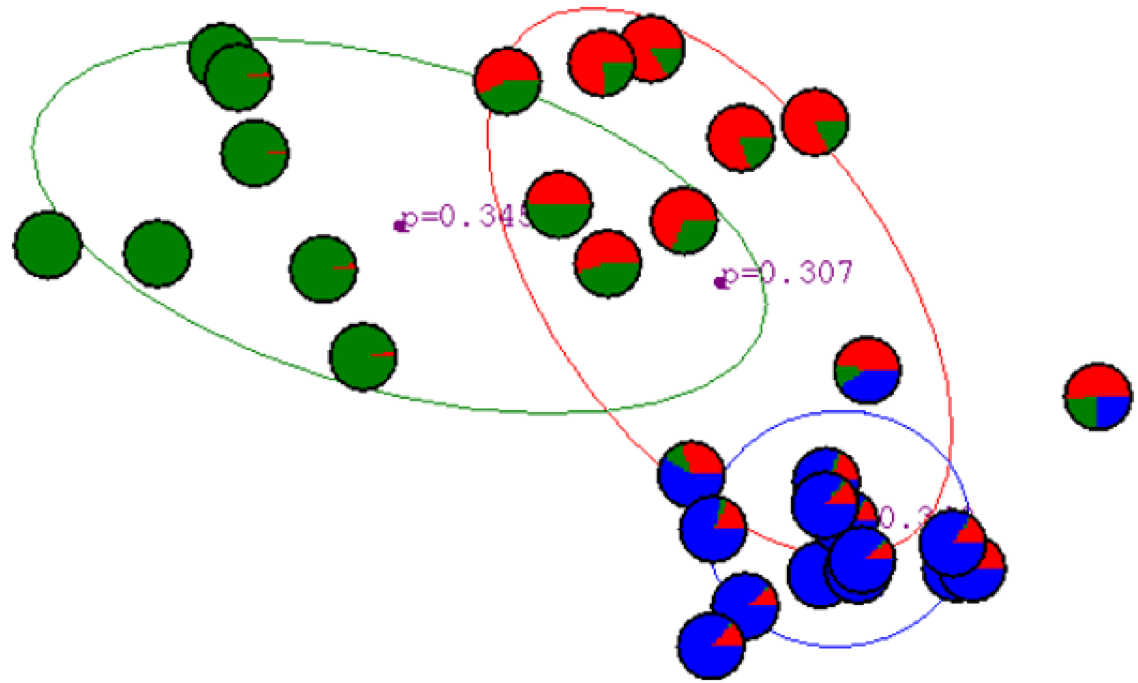
After first iteration



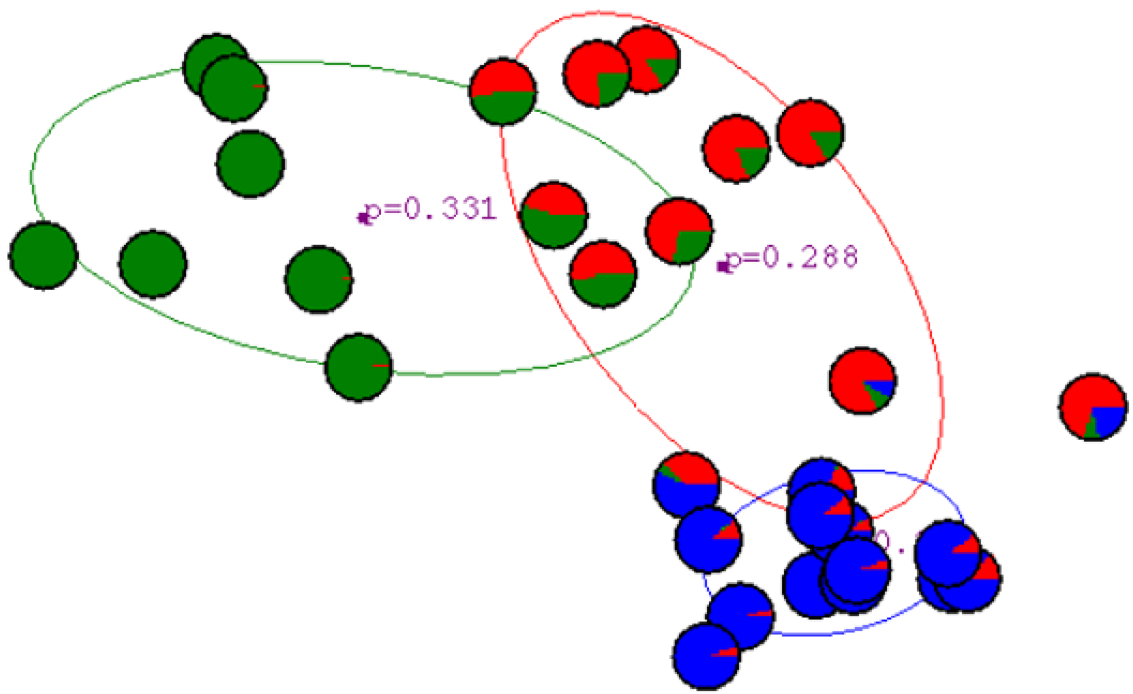
After 2nd iteration



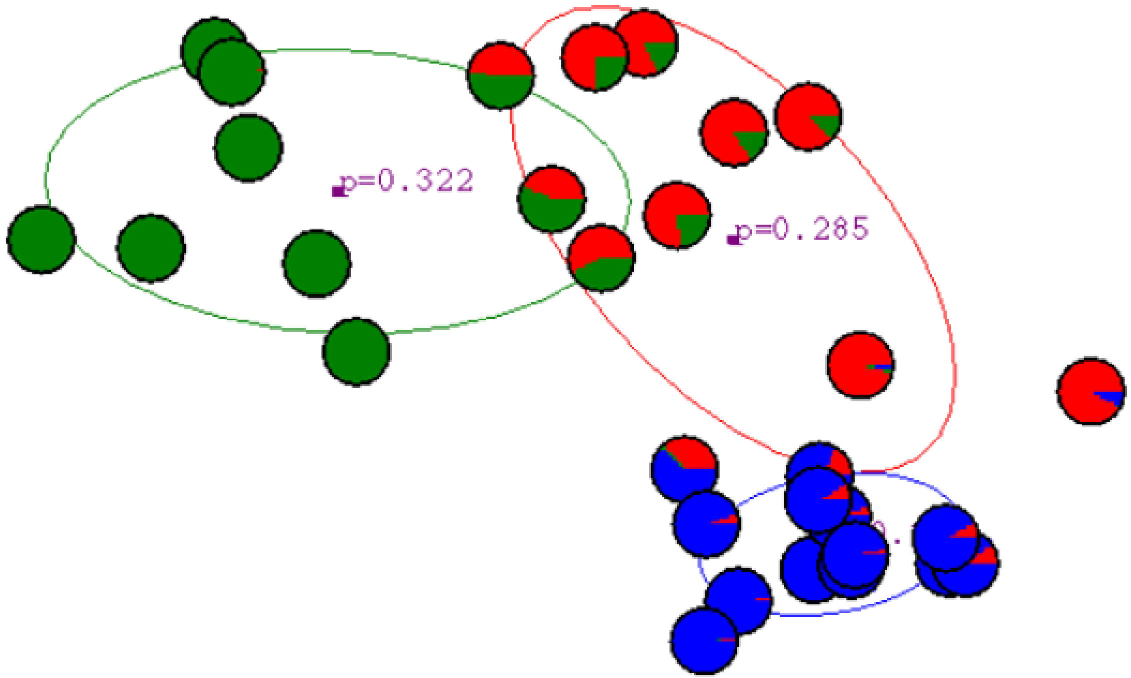
After 3rd iteration



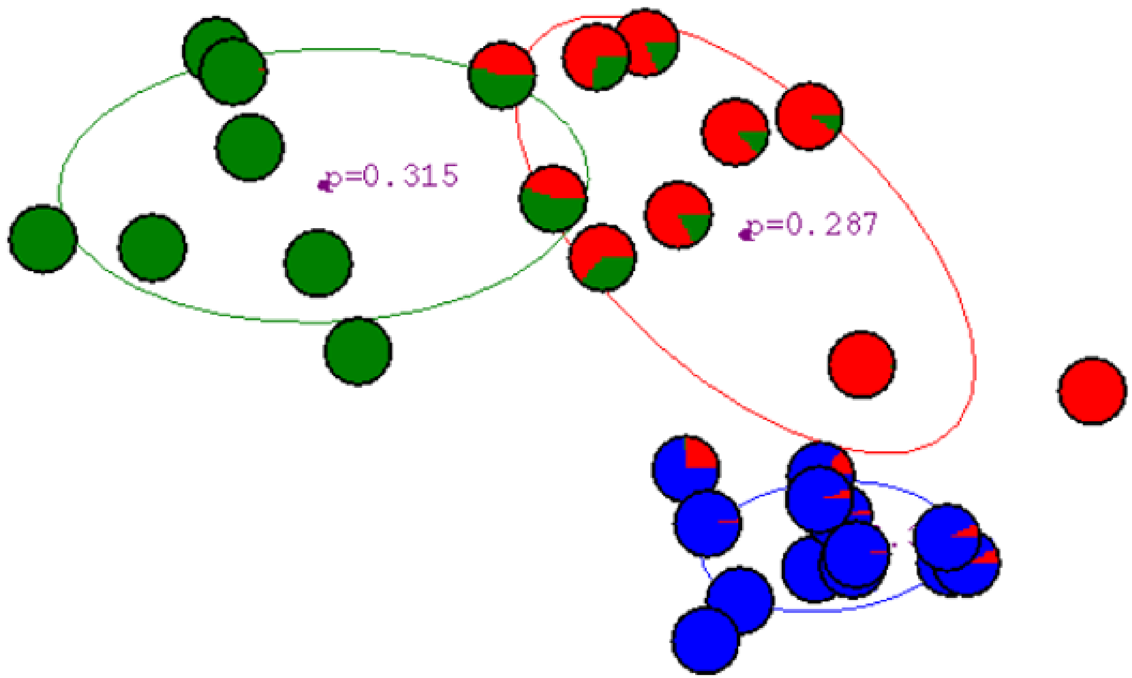
After 4th iteration



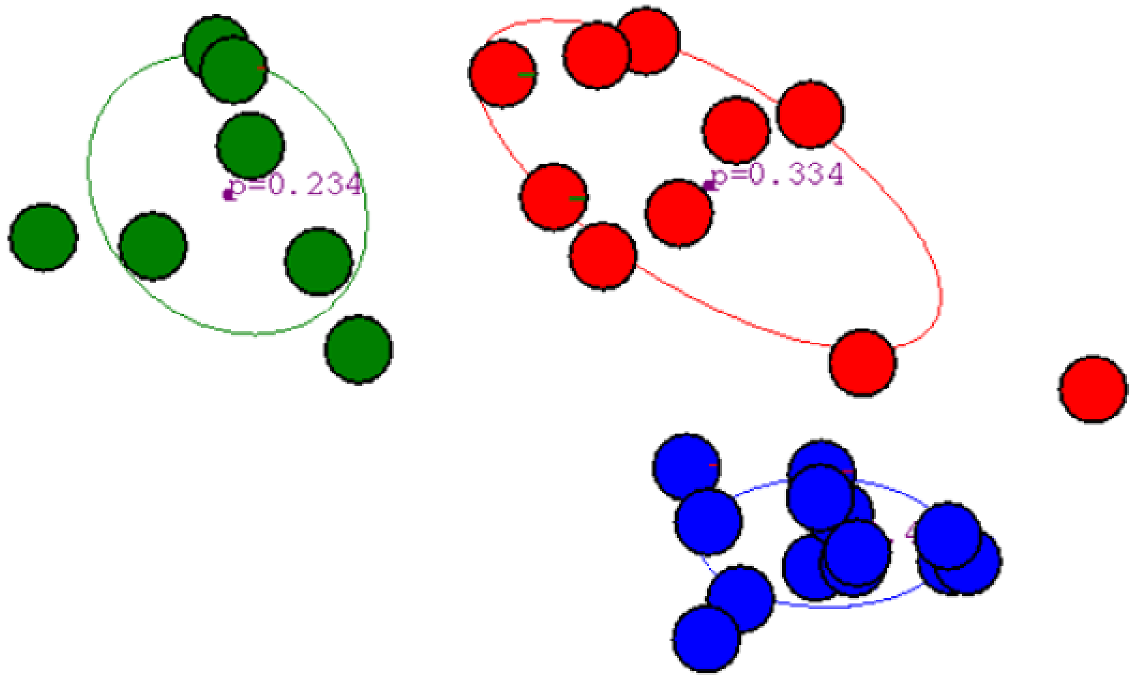
After 5th iteration



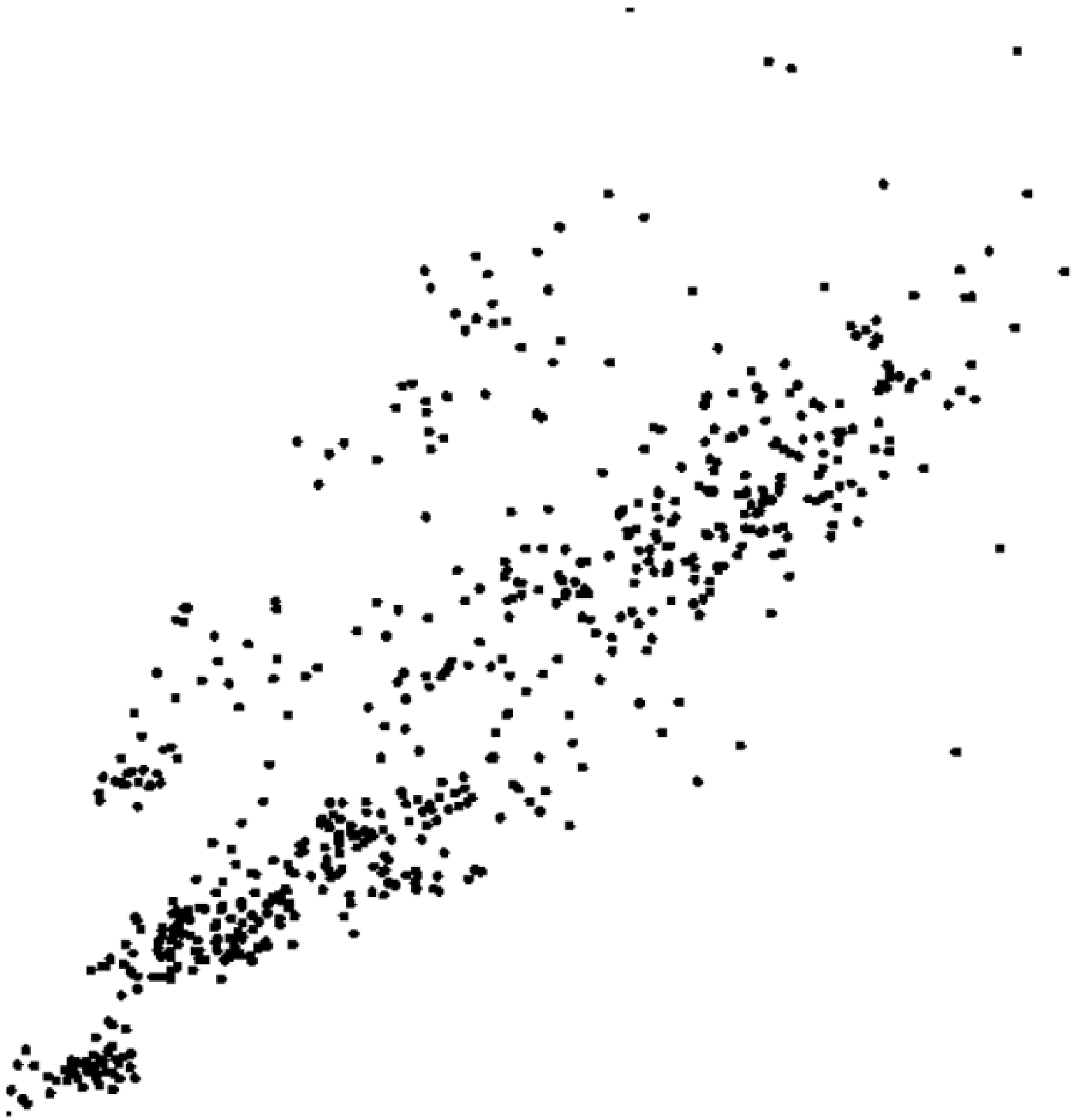
After 6th iteration



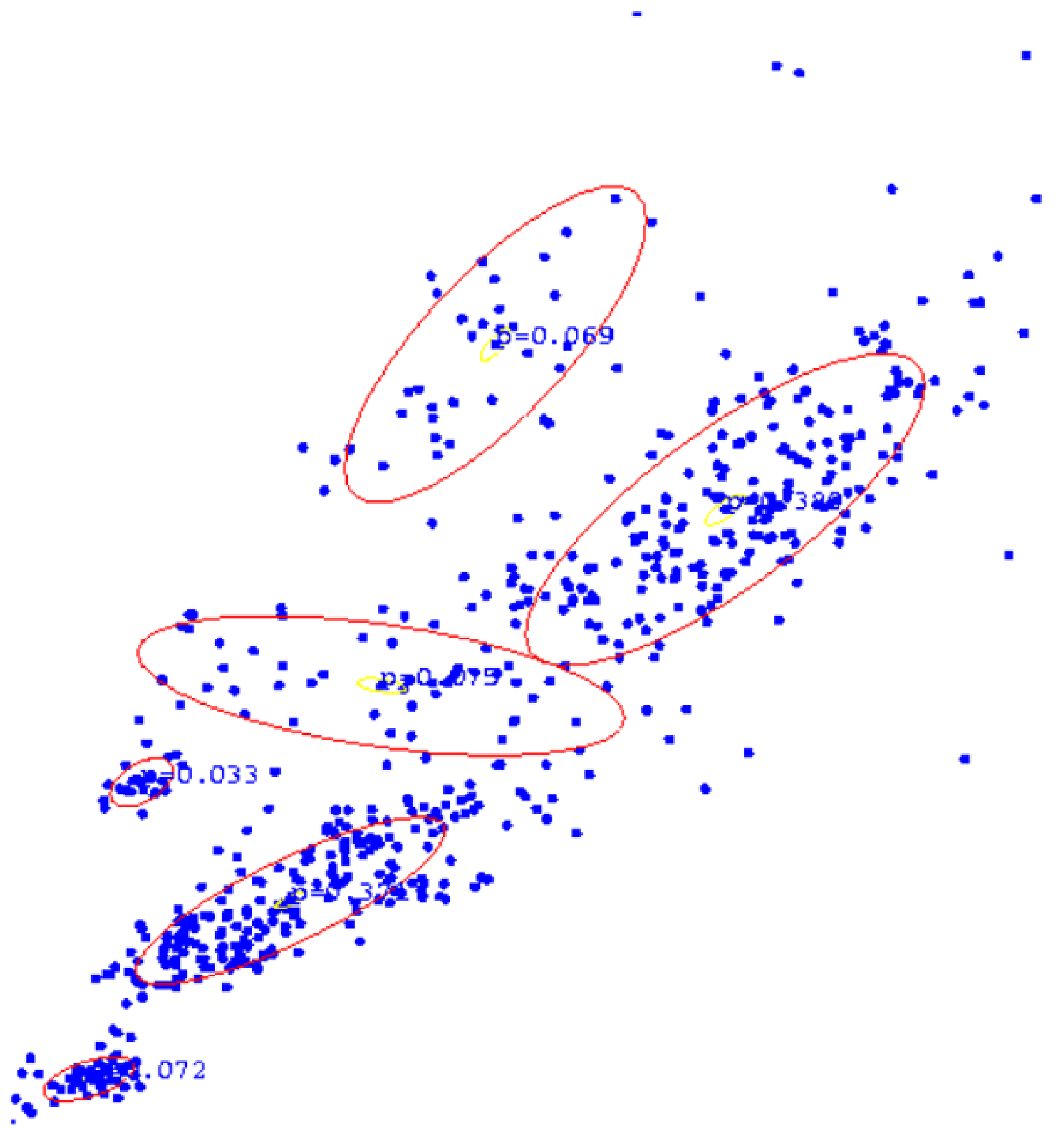
After 20th iteration



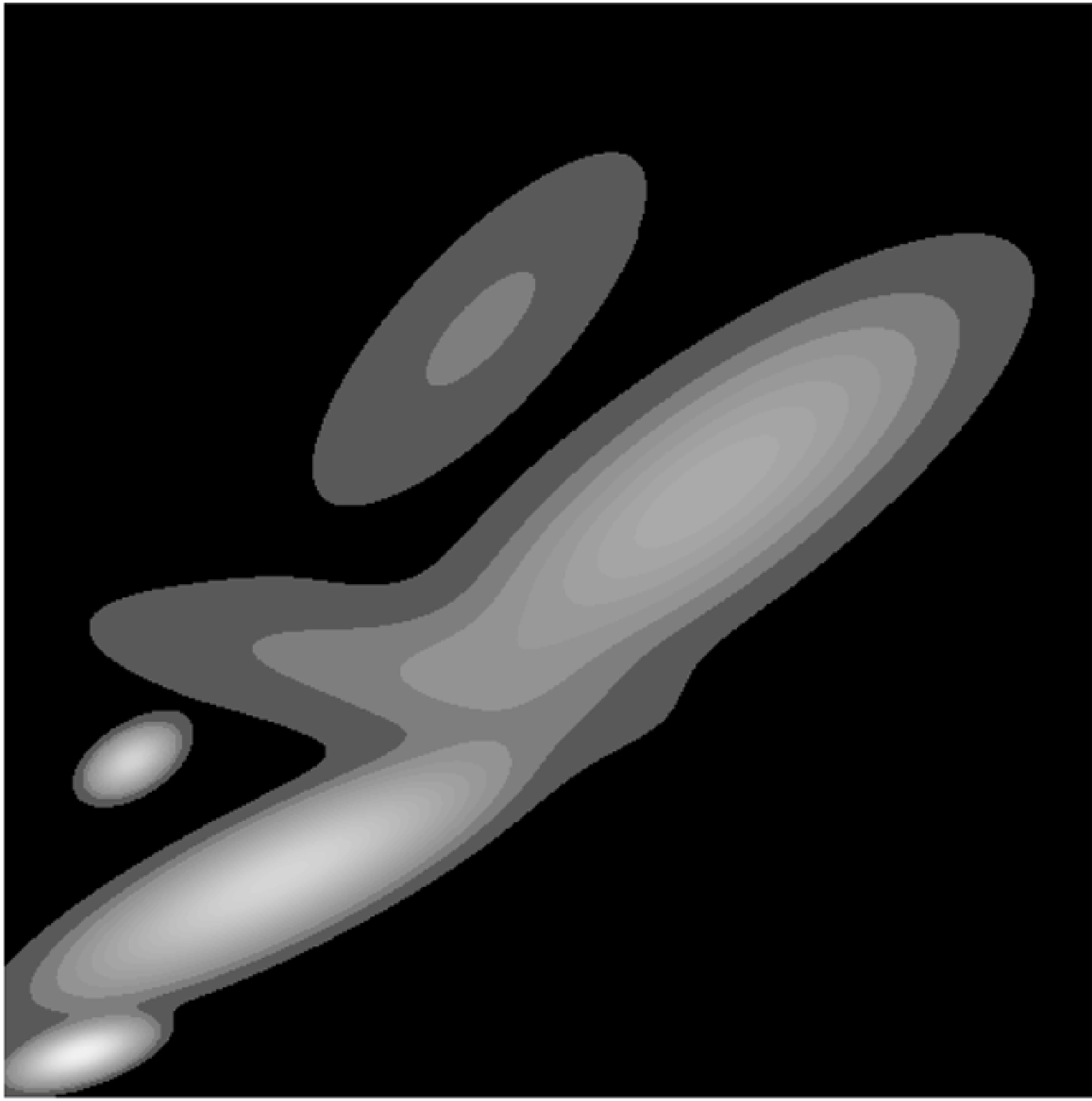
Some Bio Assay data



GMM clustering of the assay data



Resulting Density Estimator



Mixture of Gaussian demos

- http://www.cs.cmu.edu/~alad/em/
- http://lcn.epfl.ch/tutorial/english/gaussian/html/

Slide: Derek Hoiem

“Hard EM”

- Same as EM except compute z* as most likely values for hidden variables

- K-means is an example

- Advantages
  – Simpler: can be applied when cannot derive EM
  – Sometimes works better if you want to make hard predictions at the end
- But
  – Generally, pdf parameters are not as accurate as EM

Slide: Derek Hoiem

Missing Data Problems: Outliers

You want to train an algorithm to predict whether a photograph is attractive. You collect annotations from Mechanical Turk. Some annotators try to give accurate ratings, but others answer randomly.

Challenge: Determine which people to trust and the average rating by accurate annotators.

Annotator Ratings

10
8
9
2
8

Photo: Jam343 (Flickr)

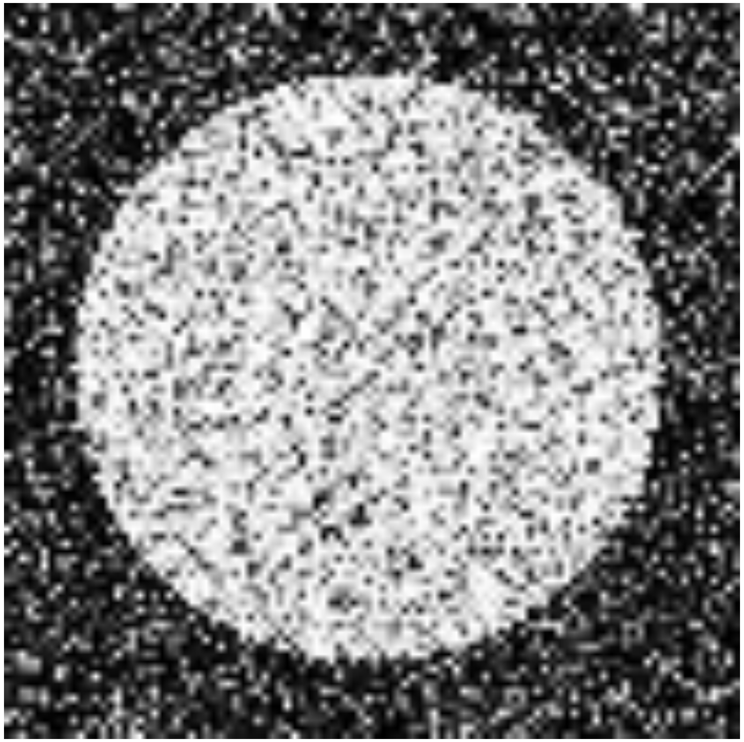
Missing Data Problems: Object Discovery

You have a collection of images and have extracted regions from them. Each is represented by a histogram of "visual words".

Challenge: Discover frequently occurring object categories, without pre-trained appearance models.

http://www.robots.ox.ac.uk/~vgg/publications/papers/russell06.pdf

What’s wrong with this prediction?



P(foreground | image)

Slide: Derek Hoiem

Next class

- MRFs and Graph-cut Segmentation