E9 205 Machine Learning for Signal Processing

Mixture Gaussian Modeling and EM Algorithm

14-09-2016

Established 1911

INDIAN INSTITUTE OF SCIENCE

भारतीय विज्ञान संस्थान

Gaussian Distribution Summary

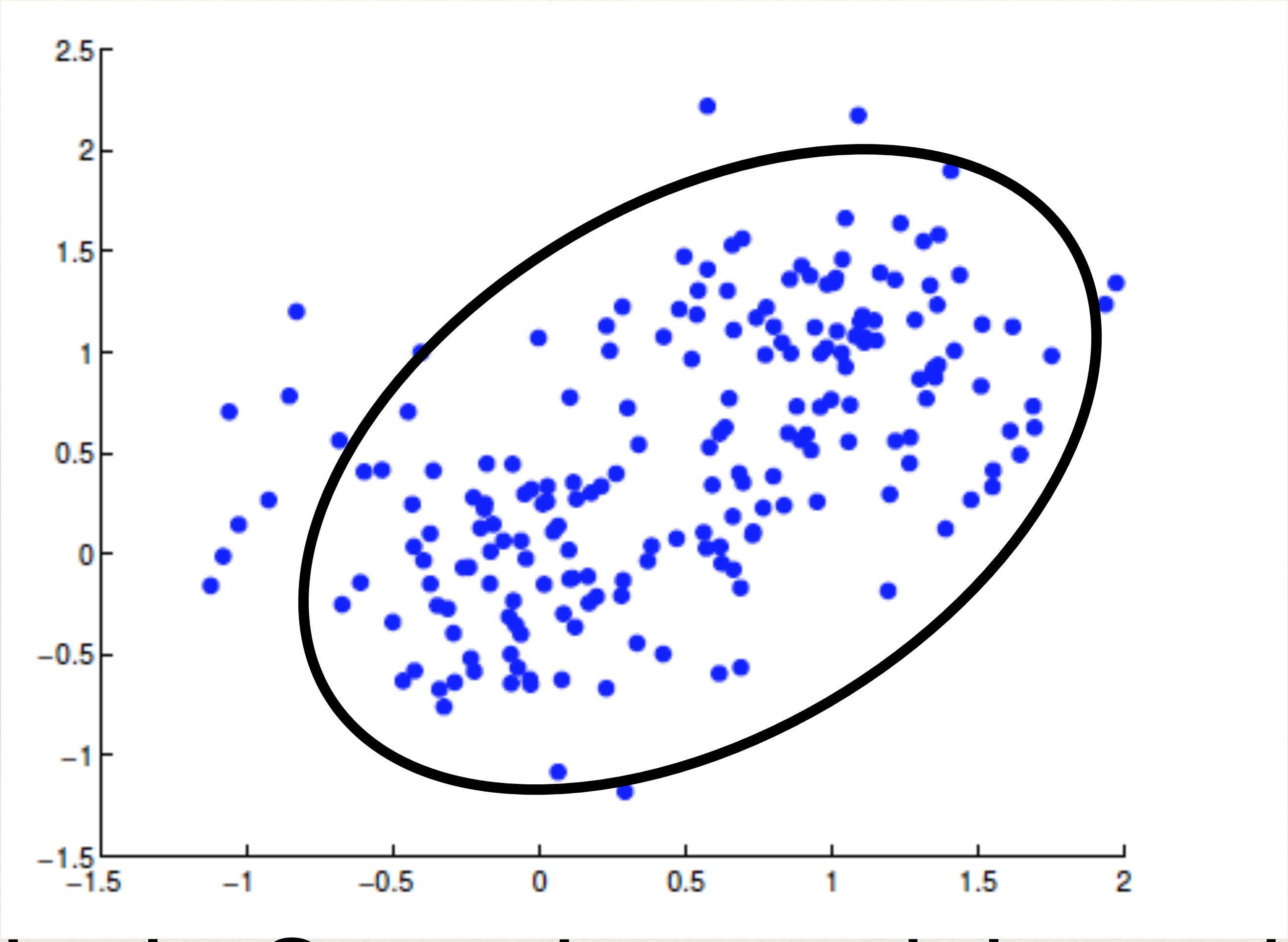
- The Gaussian model - parametric distributions
- Simple and useful properties.
- Can model unimodal (single peak distributions)
- MLE gives intuitive results
- Issues with Gaussian model
    - Multi-modal data
    - Not useful for complex data distributions
- Need for mixture models

Established
1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

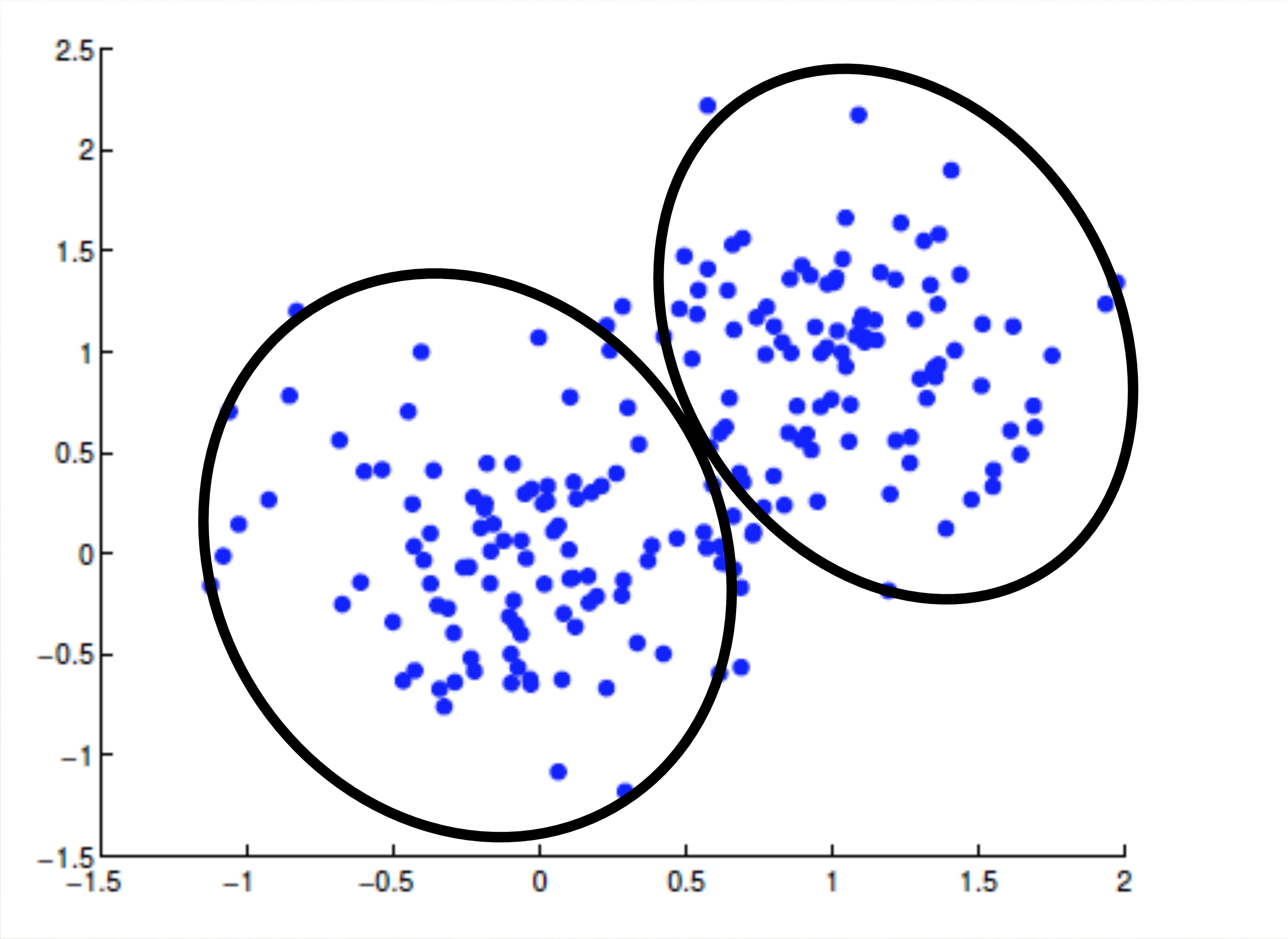
Gaussian Distribution

Often the data lies in clusters (2-D example)



Fitting a single Gaussian model may be too broad.

Gaussian Distribution

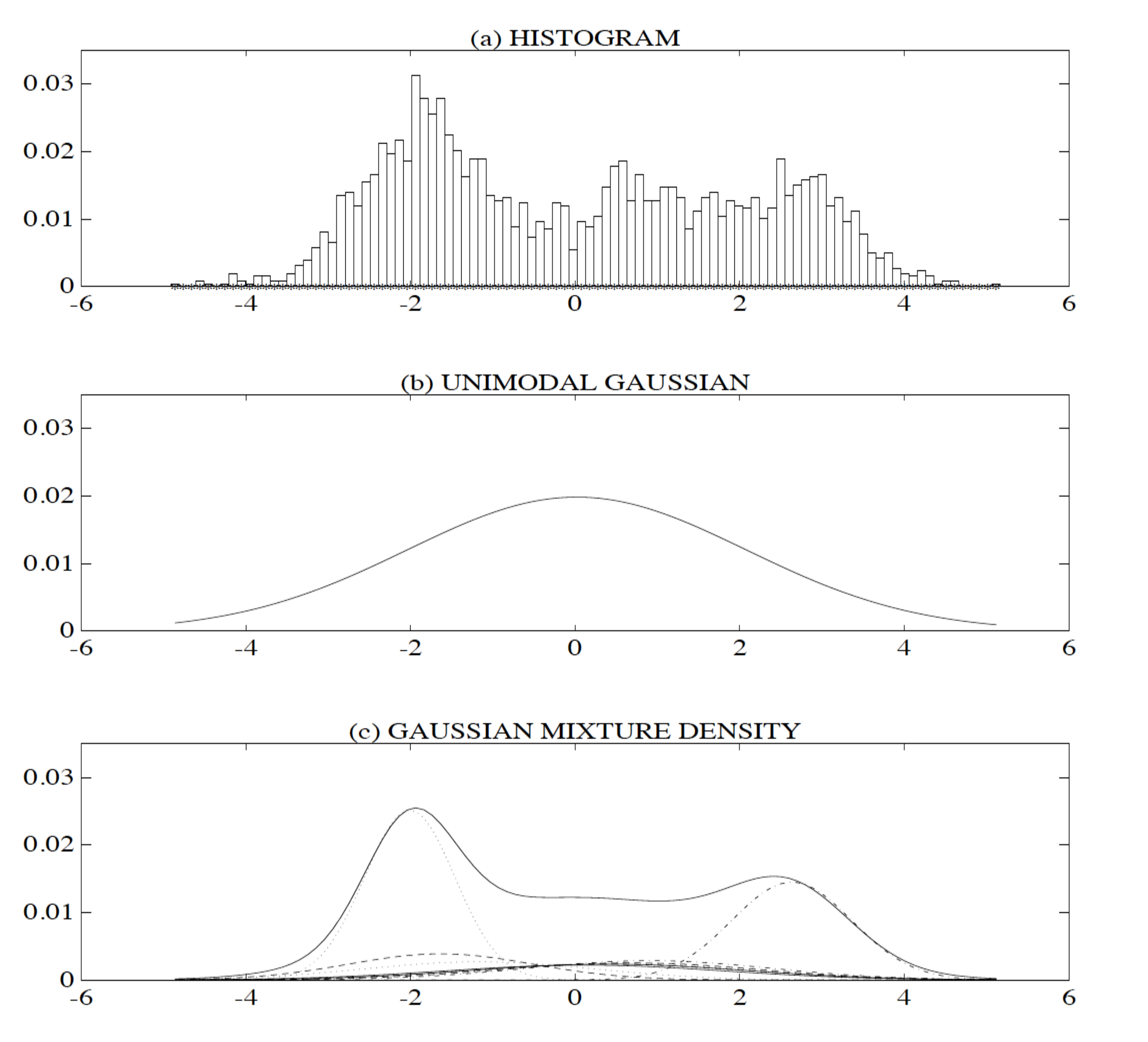


Need mixture models

Can fit any arbitrary distribution.

Gaussian Distribution

1-D example



Often the data lies in clusters

Gaussian Mixture Models

A Gaussian Mixture Model (GMM) is defined as

$$p(\mathbf{x}|\Theta) = \sum_{k=1}^{K} \alpha_k p(\mathbf{x}|\boldsymbol{\theta}_k)$$

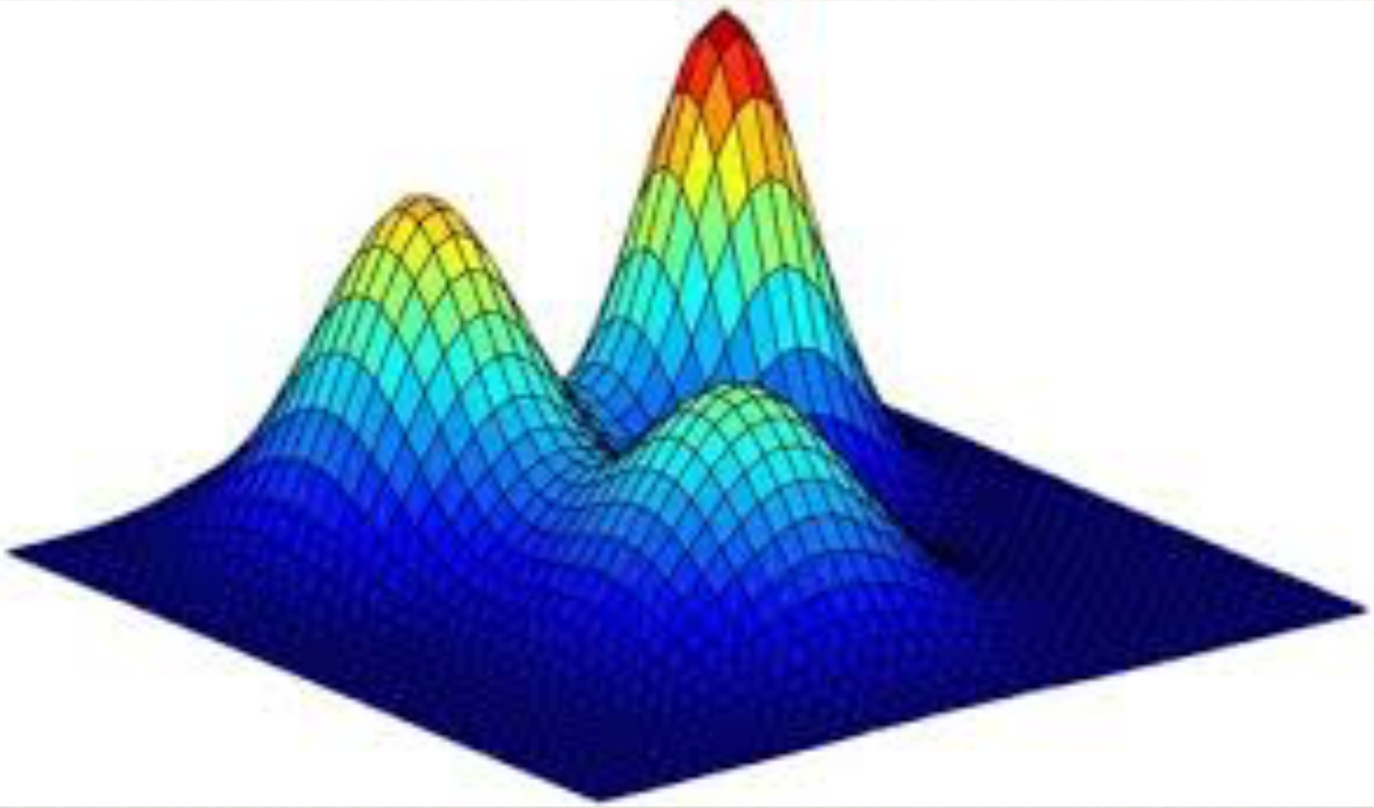
$$p(\mathbf{x}|\boldsymbol{\theta}_k) = \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}_k|}} exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^*\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)\right\}$$

The weighting coefficients have the property

$$\sum_{k=1}^{K} \alpha_k = 1$$

Gaussian Mixture Models

- Properties of GMM
  - Can model multi-modal data.
  - Identify data clusters.
  - Can model arbitrarily complex data distributions



The set of parameters for the model are

$$\boldsymbol{\Theta}_k = \{\alpha_k, \boldsymbol{\theta}_k\}_{k=1}^{K} \qquad \boldsymbol{\theta}_k = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$$

The number of parameters is $KD^2 + KD + K$

MLE for GMM

- The log-likelihood function over the entire data in this case will have a logarithm of a summation

$$\log L(\boldsymbol{\Theta}) = \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} \alpha_k p(\mathbf{x}_i | \boldsymbol{\theta}_k) \right)$$

- Solving for the optimal parameters using MLE for GMM is not straight forward.
- Resort to the Expectation Maximization (EM) algorithm

Expectation Maximization Algorithm

- Iterative procedure.
- Assume the existence of hidden variable associated with each data sample
- Let the current estimate (at iteration n) be Define the Q function as

$$Q(\mathbf{\Theta}, \mathbf{\Theta}^n) = E_{\mathbf{z}|\mathbf{X}, \mathbf{\Theta}^n}\left[\log(P(\mathbf{X}, \mathbf{z}|\mathbf{\Theta})\right]$$
$$= \sum_{\mathbf{z}} \log(P(\mathbf{X}, \mathbf{z}|\mathbf{\Theta}))P(\mathbf{z}|\mathbf{X}, \mathbf{\Theta}^n)$$

Expectation Maximization Algorithm

- It can be proven that if we choose

$$\mathbf{\Theta}^{n+1} = arg \max_{\mathbf{\Theta}} Q(\mathbf{\Theta}, \mathbf{\Theta}^n)$$

then $$L(\mathbf{\Theta}^{n+1}) \geq L(\mathbf{\Theta}^n)$$

- In many cases, finding the maximum for the Q function may be easier than likelihood function w.r.t. the parameters.
- Solution is dependent on finding a good choice of the hidden variables which eases the computation
- Iteratively improve the log-likelihood function.

EM Algorithm Summary

- Initialize with a set of model parameters (n=1)
- Compute the conditional expectation (E-step)

$$E_{\mathbf{z}|\mathbf{X},\boldsymbol{\Theta}^n}\left[\log(P(\mathbf{X},\mathbf{z}|\boldsymbol{\Theta})\right]$$

- Maximize the conditional expectation w.r.t. parameter. (M-step) (n = n+1)
- Check for convergence
- Go back to E-step if model has not converged.

EM Algorithm for GMM

- The hidden variables $\mathbf{z}_i = l$ will be the index of the mixture component which generated $\mathbf{x}_i$
- Re-estimation formulae

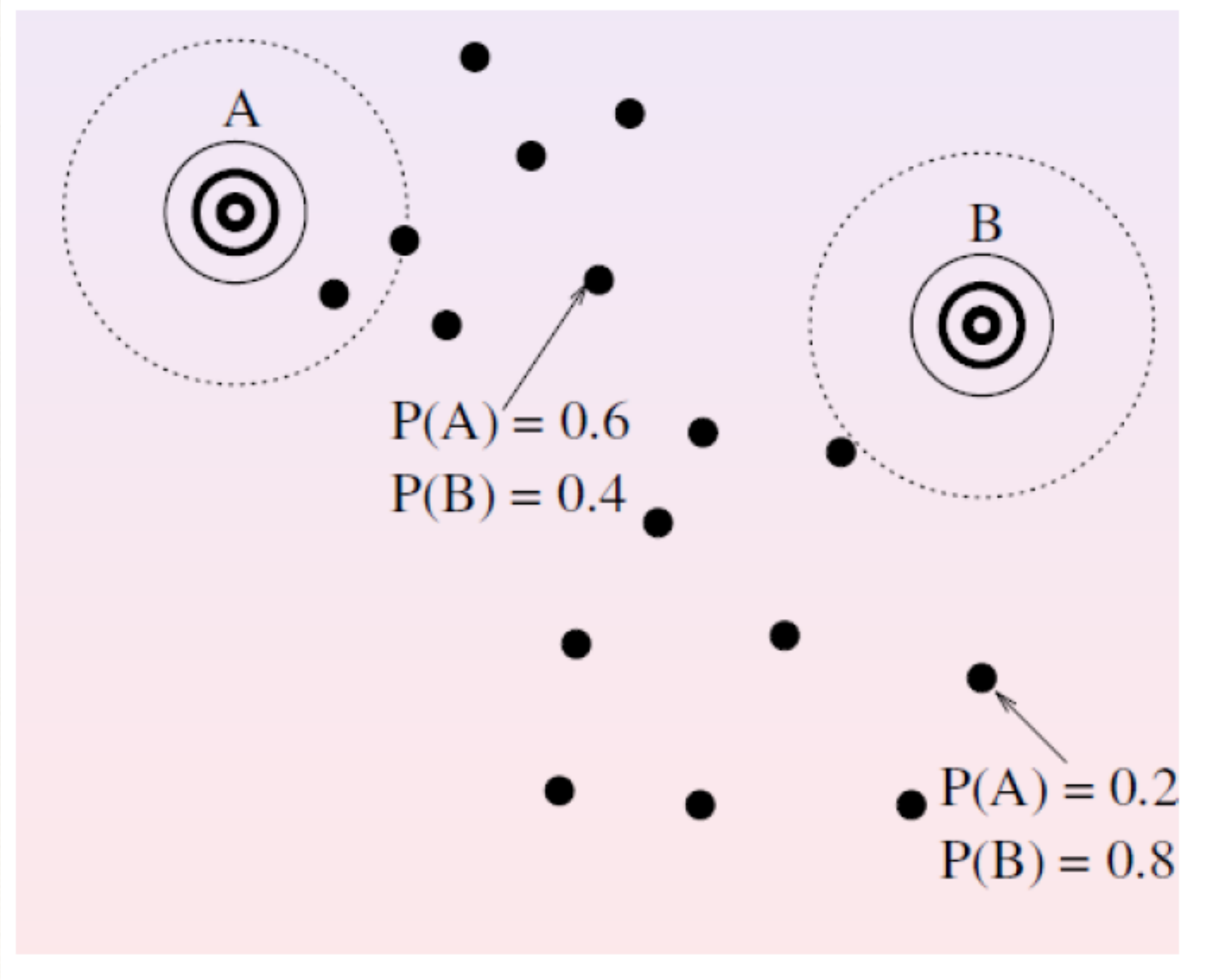
$$\alpha_\ell^{new} = \frac{1}{N} \sum_{i=1}^{N} p(\ell | x_i, \Theta^g)$$

$$\mu_\ell^{new} = \frac{\sum_{i=1}^{N} x_i p(\ell | x_i, \Theta^g)}{\sum_{i=1}^{N} p(\ell | x_i, \Theta^g)}$$

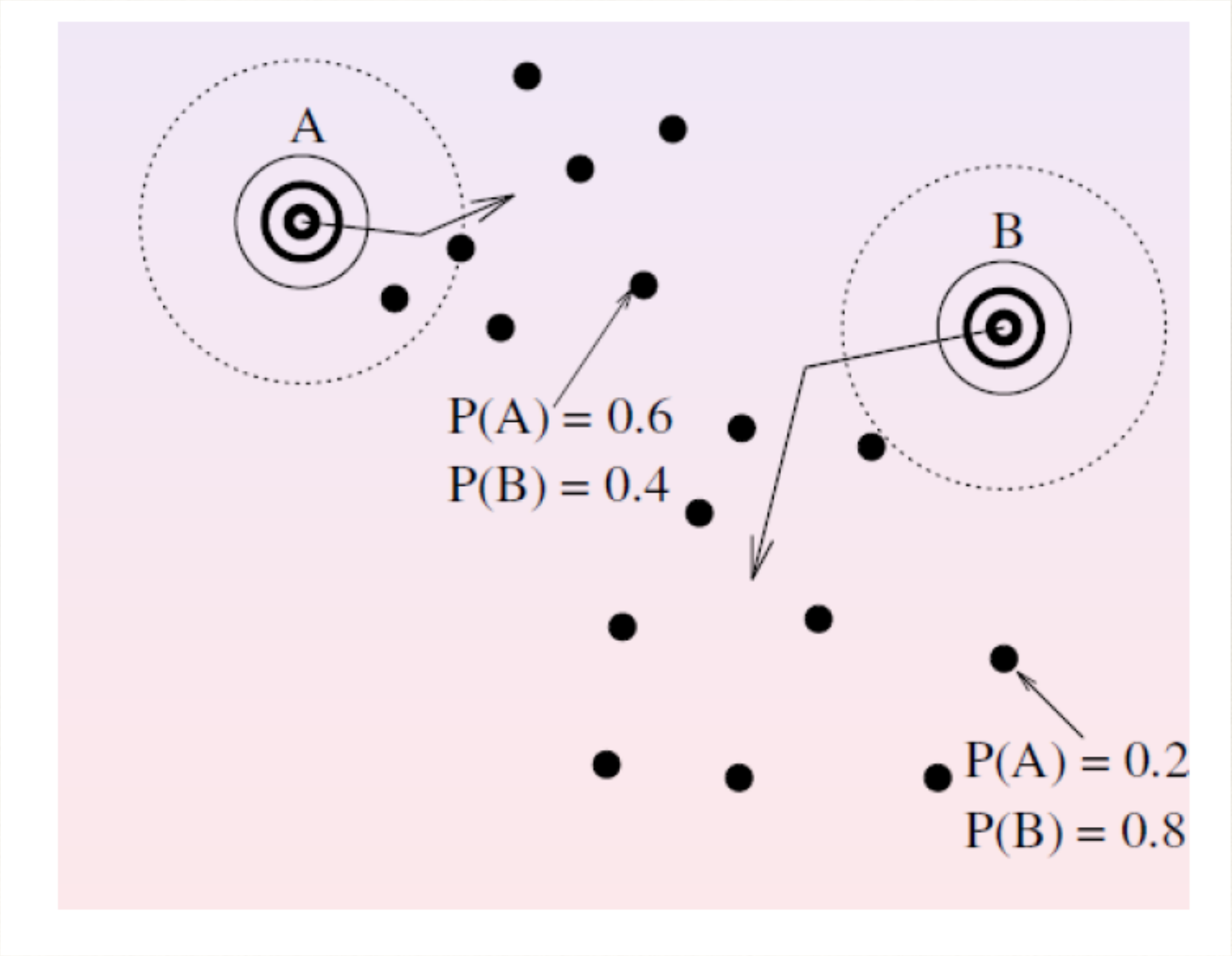
$$\Sigma_\ell^{new} = \frac{\sum_{i=1}^{N} p(\ell | x_i, \Theta^g)(x_i - \mu_\ell^{new})(x_i - \mu_\ell^{new})^T}{\sum_{i=1}^{N} p(\ell | x_i, \Theta^g)}$$

EM Algorithm for GMM

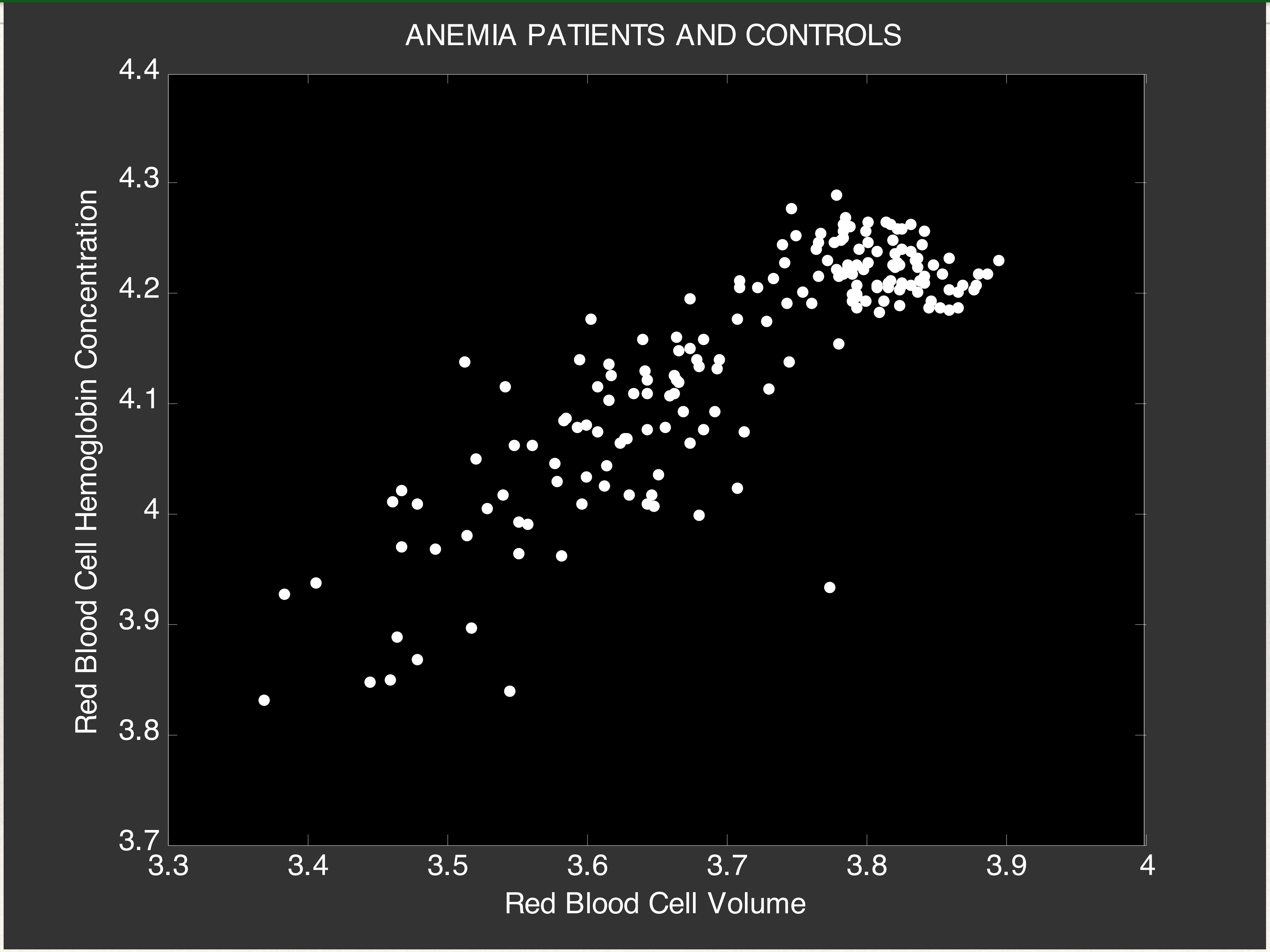
E-step



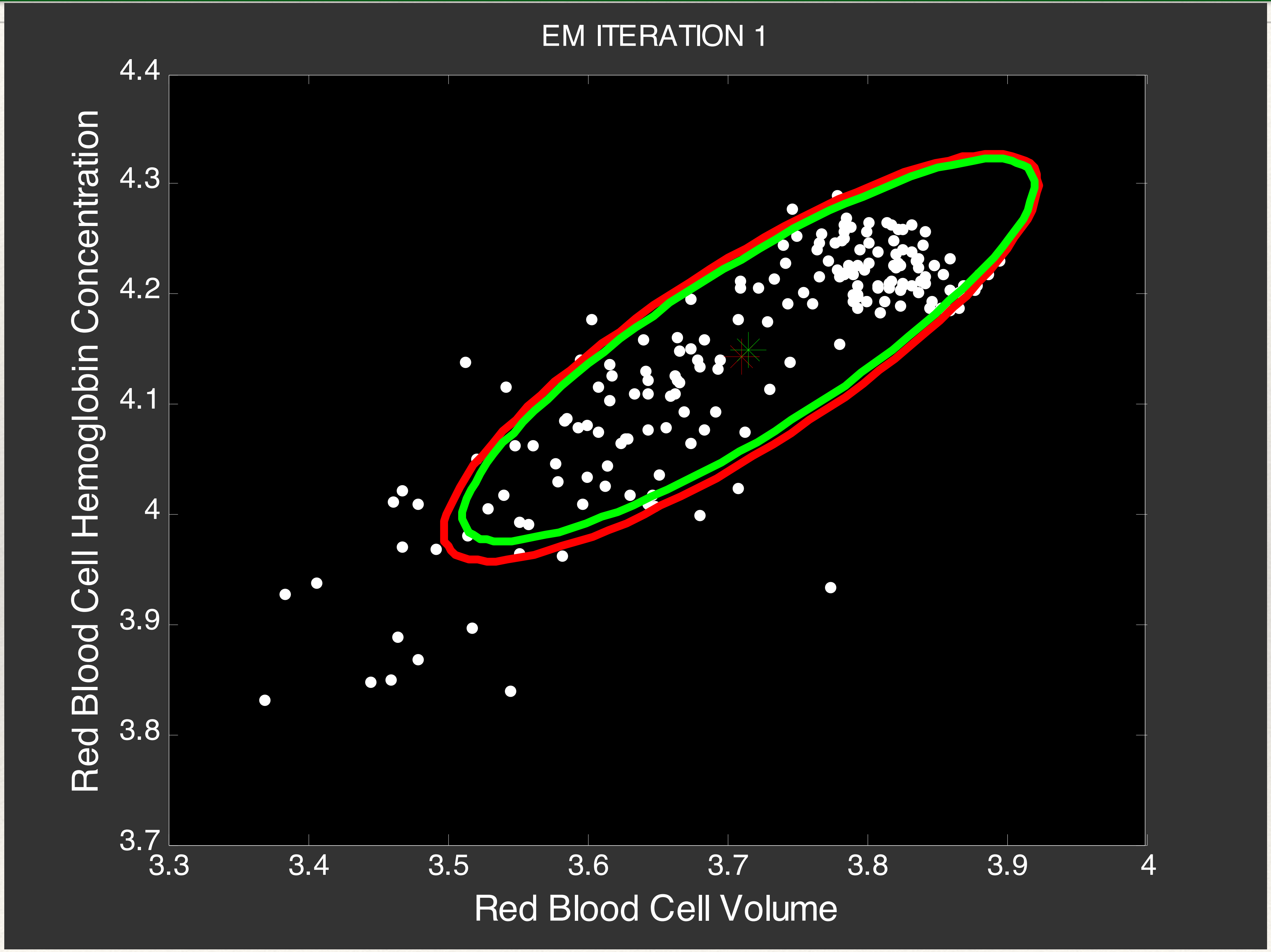
M-step



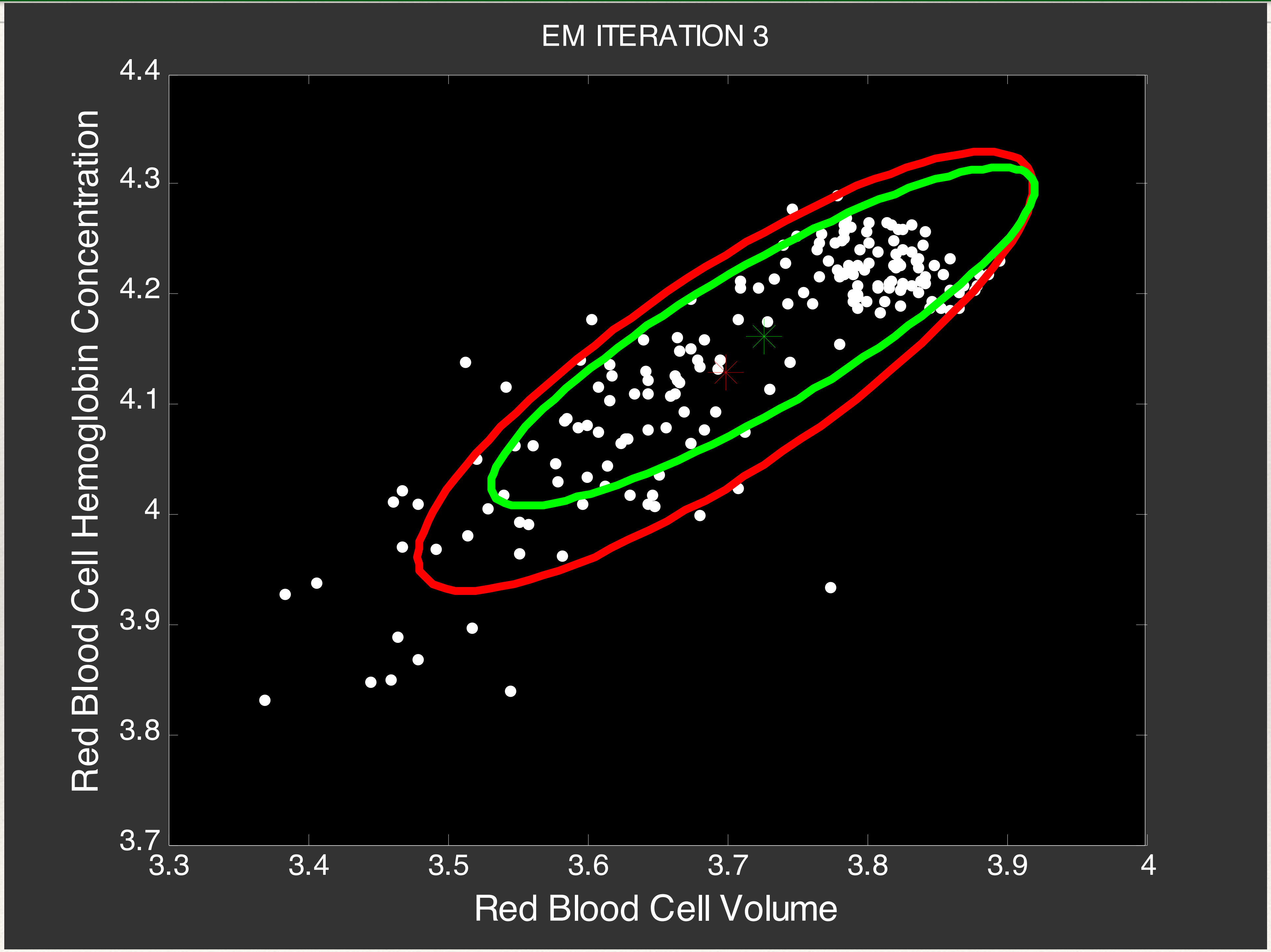
EM Algorithm for GMM



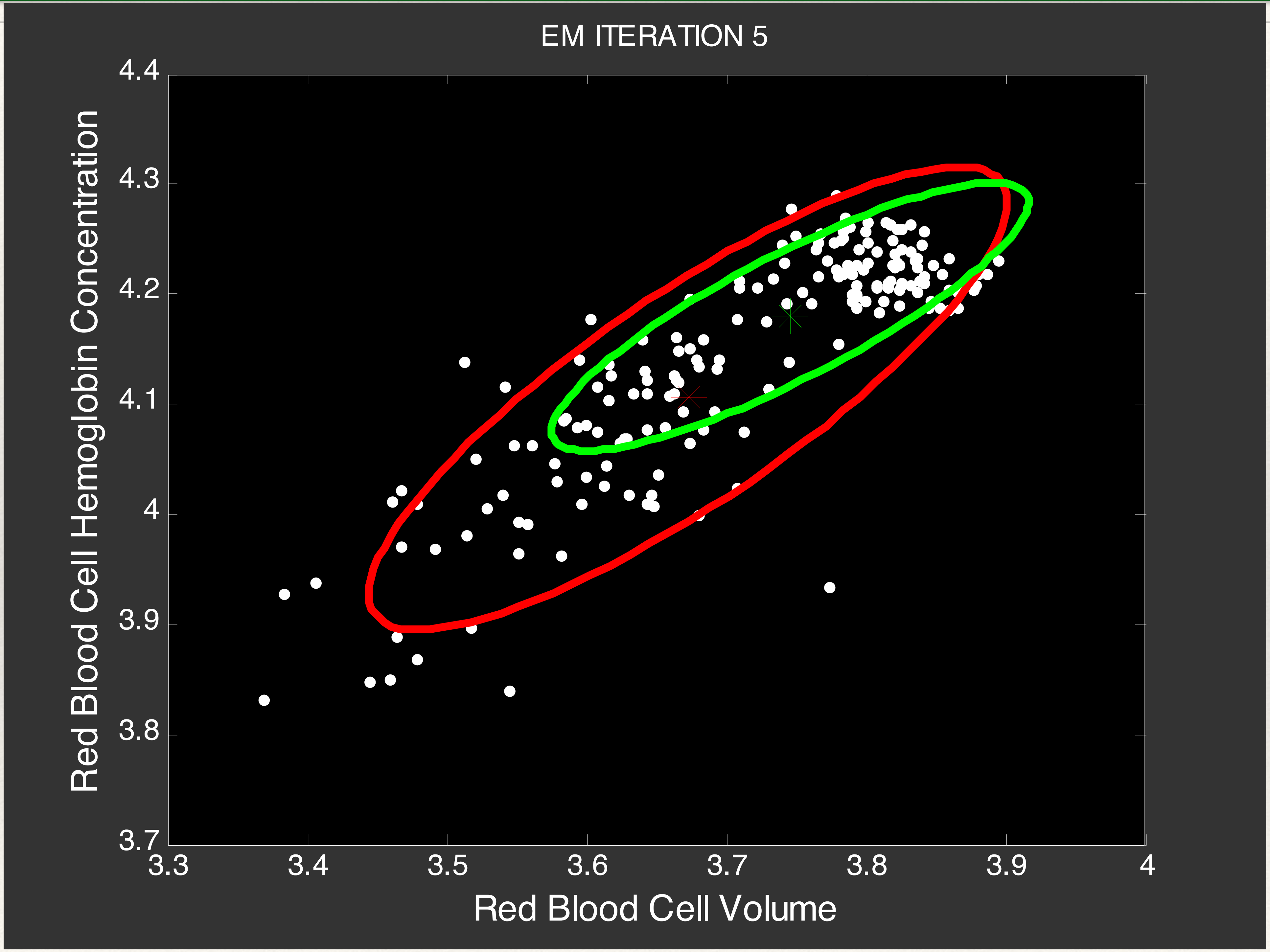
EM Algorithm for GMM



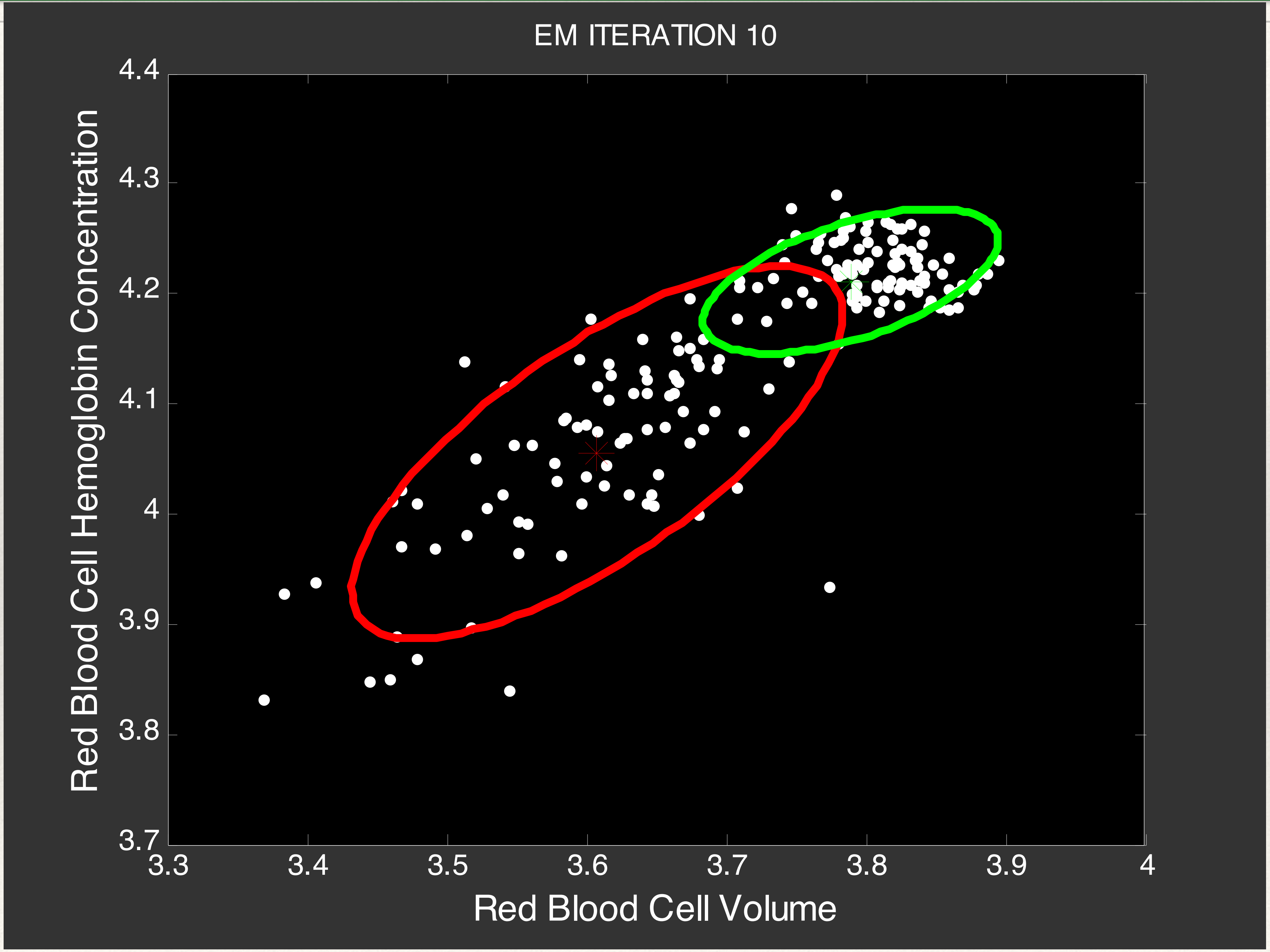
EM Algorithm for GMM



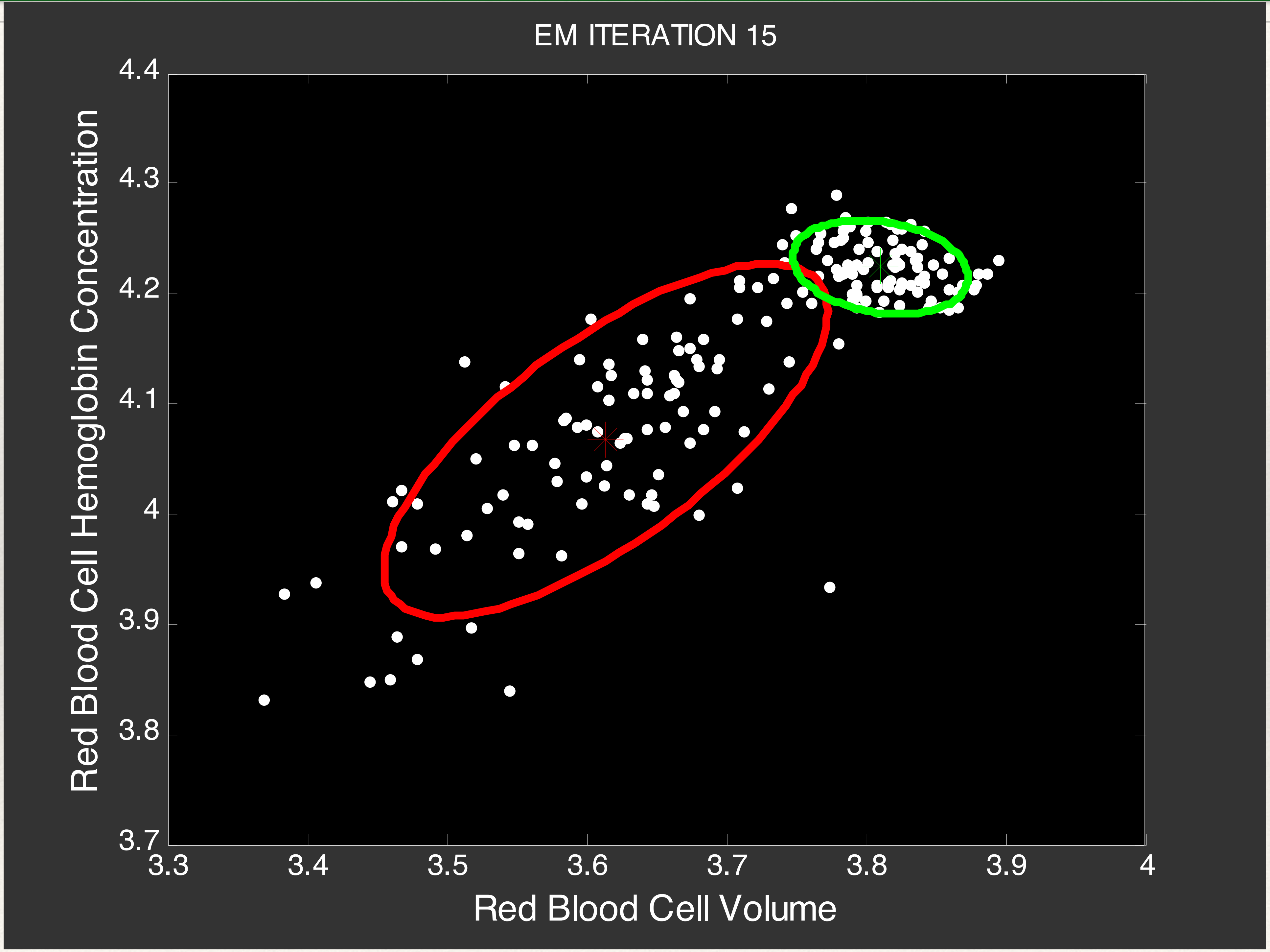
EM Algorithm for GMM



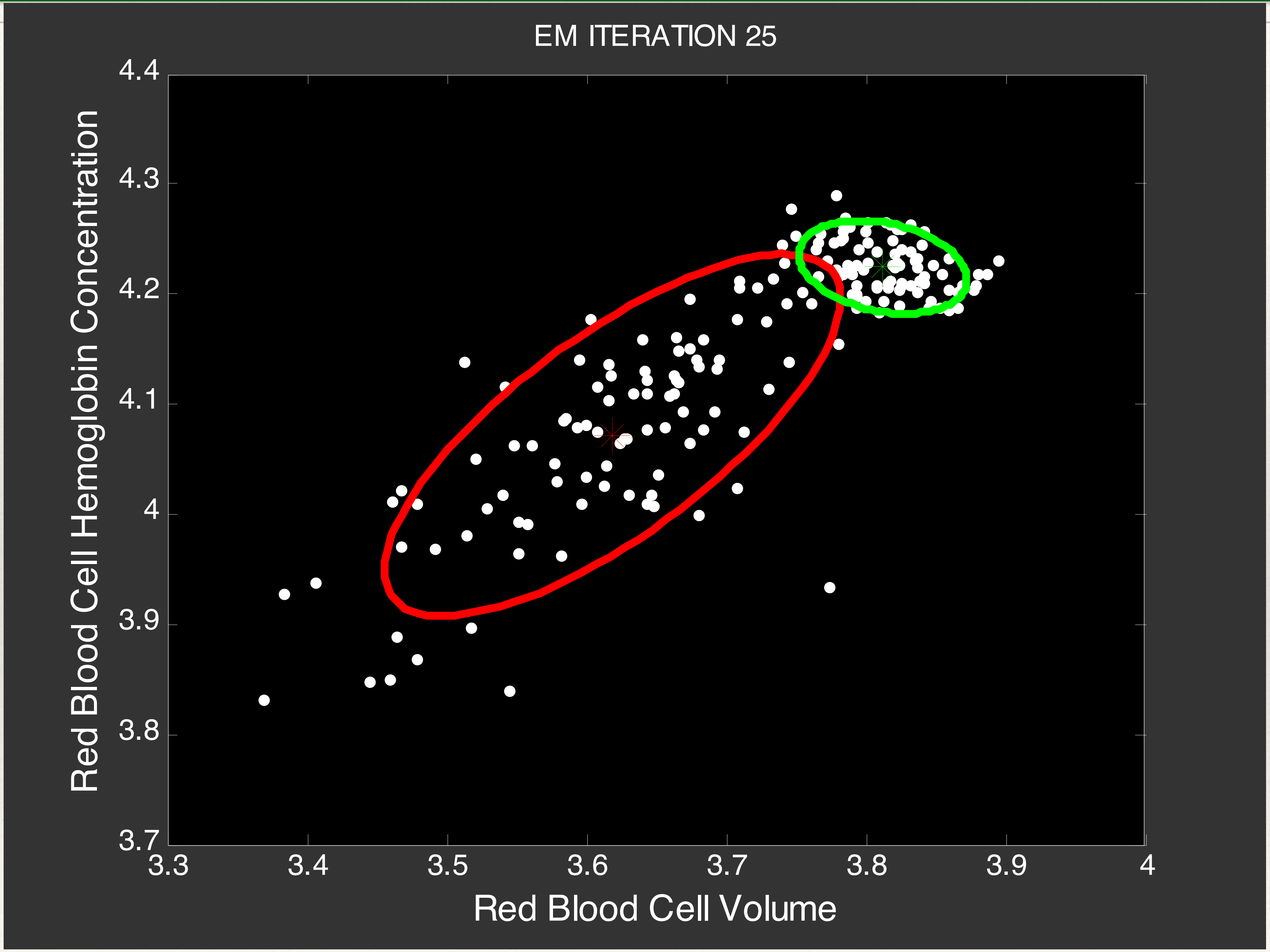
EM Algorithm for GMM



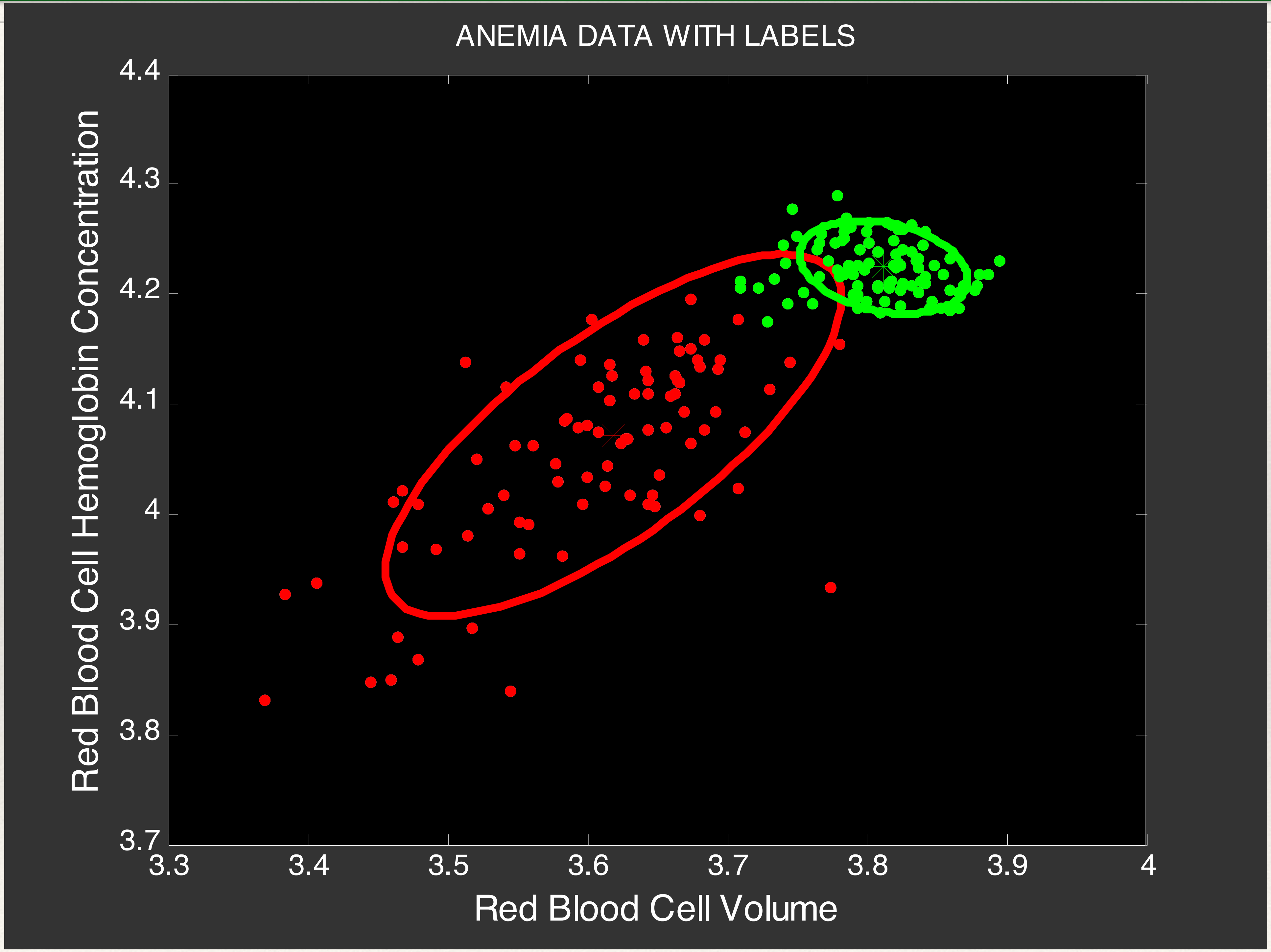
EM Algorithm for GMM



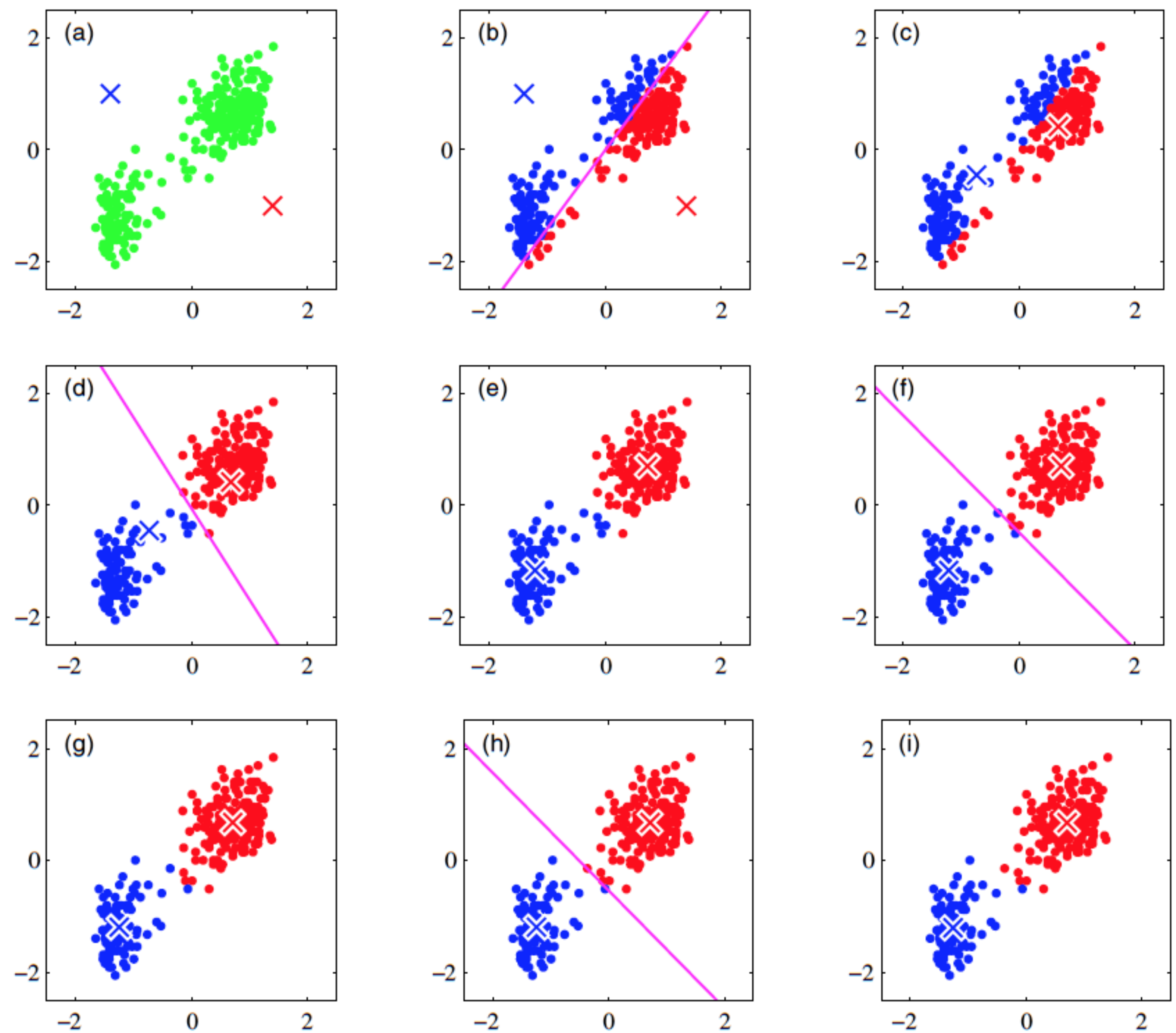
EM Algorithm for GMM



EM Algorithm for GMM

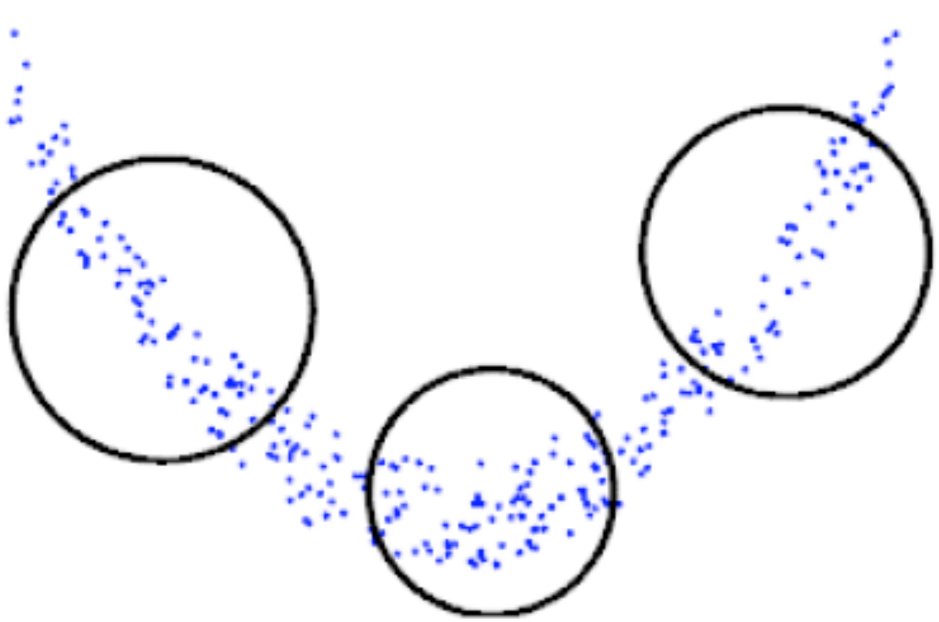


K-means Algorithm for Initialization



Other Considerations

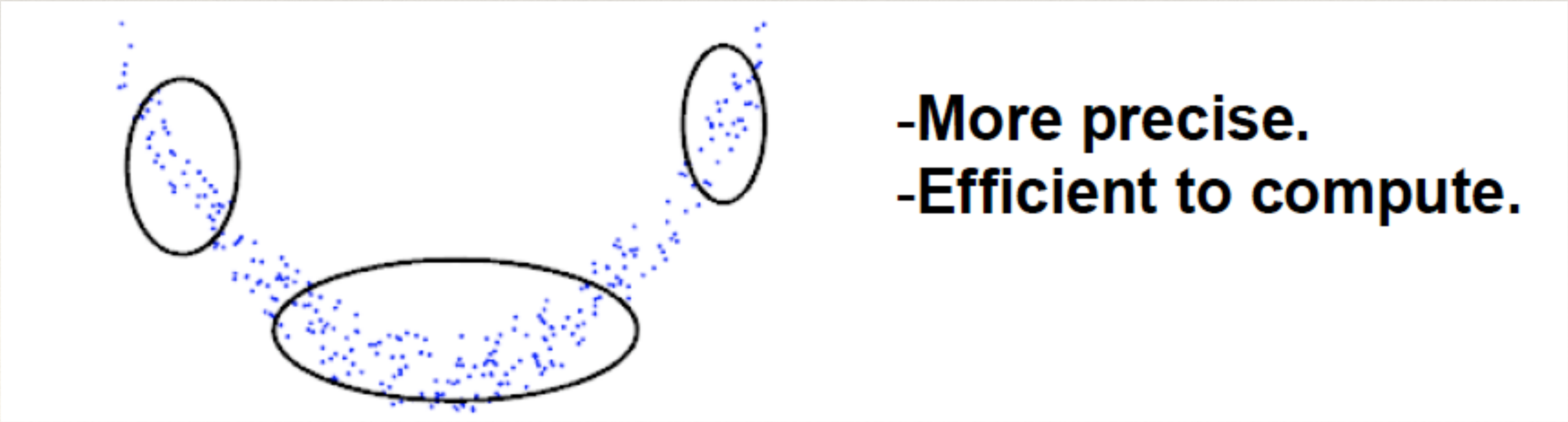
- Initialization - random or k-means
- Number of Gaussians
- Type of Covariance matrix
  - Spherical covariance



-Less precise.
-Very efficient to compute.

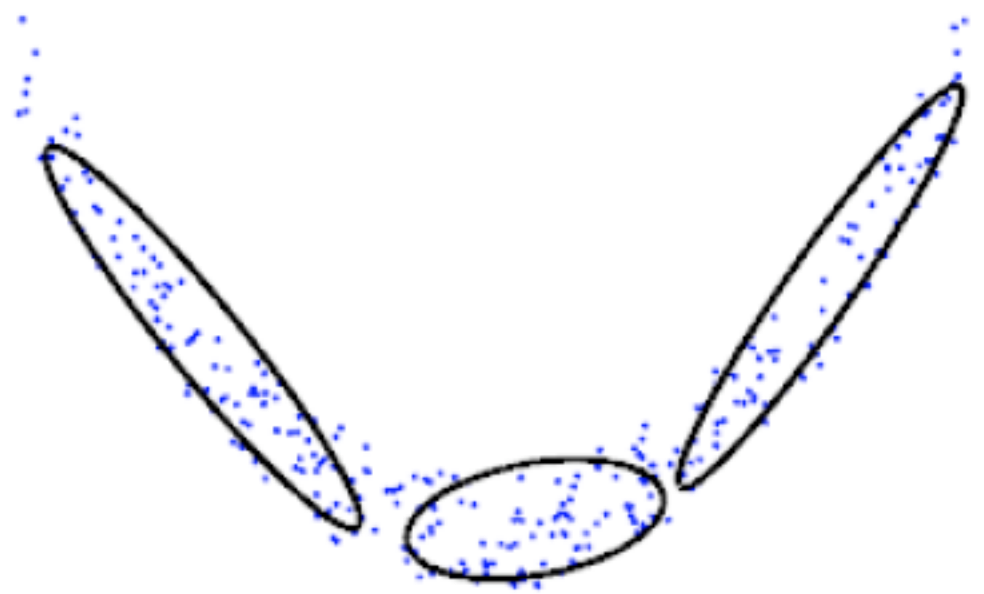
Other Considerations

- Initialization - random or k-means
- Number of Gaussians
- Type of Covariance matrix
  - Diagonal covariance



Other Considerations

- Initialization - random or k-means
- Number of Gaussians
- Type of Covariance matrix
  - Full covariance



-Very precise.
-Less efficient to compute.