E9 205 Machine Learning for Signal Processing

EM Algorithm

For Mixture Gaussian Models

09-09-2019

Instructor - Sriram Ganapathy (sriramg@iisc.ac.in)

Teaching Assistant - Prachi Singh (prachisingh@iisc.ac.in).

Basics of Information Theory

- Entropy of distribution
- KL divergence
- Jensen’s inequality
- Expectation Maximization Algorithm for MLE

Gaussian Mixture Models

A Gaussian Mixture Model (GMM) is defined as

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

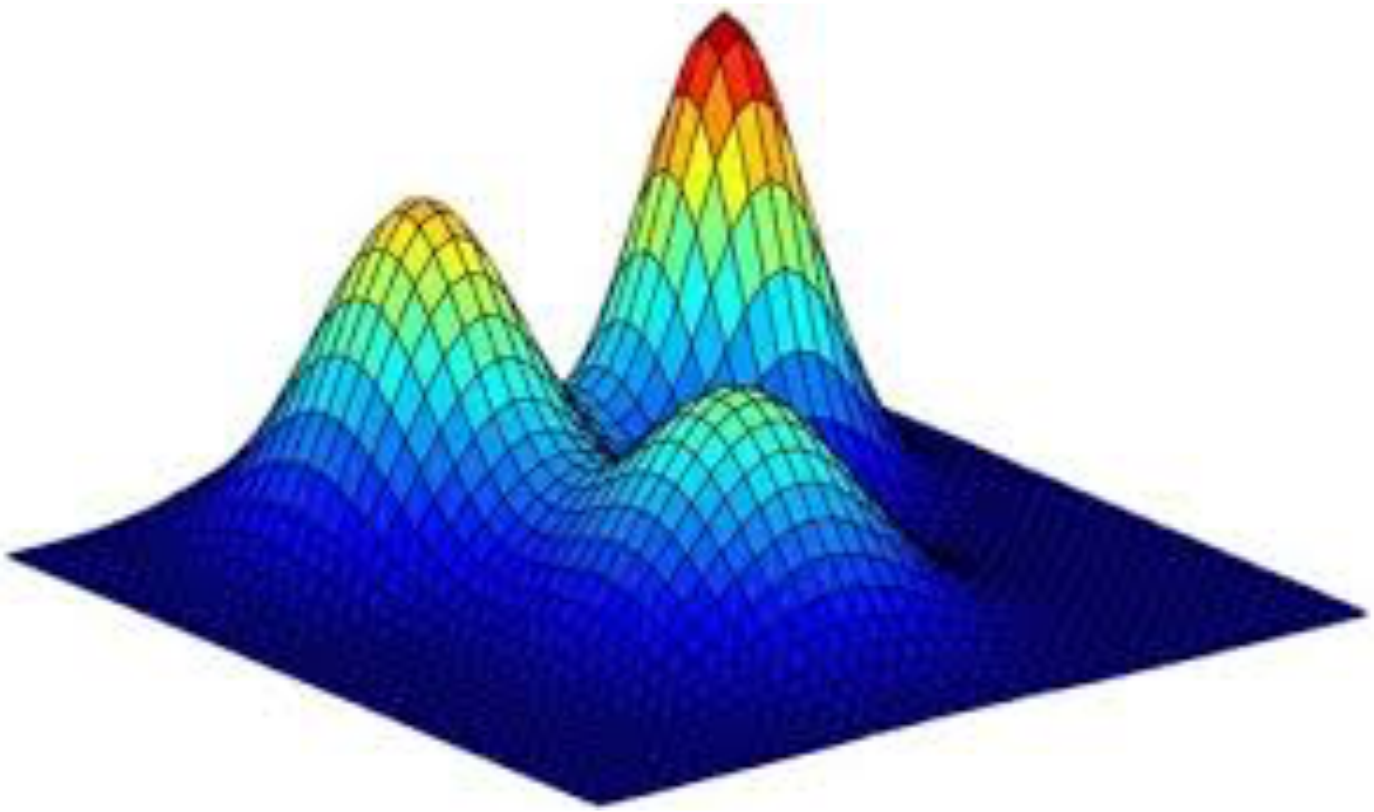
$$p(\mathbf{x}|\boldsymbol{\theta}_k) = \frac{1}{\sqrt{(2\pi)^D |\mathbf{\Sigma}_k|}} exp\left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^* \mathbf{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \right\}$$

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

$$\Theta_k = \{\alpha_k, \theta_k\}_{k=1}^{K} \qquad \theta_k = \{\mu_k, \Sigma_k\}$$

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

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

- Solving for the optimal parameters using MLE for GMM is not straight forward.

- Resort to the Expectation Maximization (EM) algorithm

Basics of Information Theory

- Entropy of distribution
- KL divergence
- Jensen’s inequality
- Expectation Maximization Algorithm for MLE

Expectation Maximization Algorithm

- Iterative procedure.
- Assume the existence of hidden variable $\mathbf{z}_i$ associated with each data sample $\mathbf{x}_i$
- Let the current estimate (at iteration n) be $\boldsymbol{\Theta}^n$ Define the Q function as

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

$$= \sum_{\mathbf{z}} \log(P(\mathbf{X}, \mathbf{z}|\boldsymbol{\Theta}))P(\mathbf{z}|\mathbf{X}, \boldsymbol{\Theta}^n)$$

Established 1911

Expectation Maximization Algorithm

- It can be proven that if we choose

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

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

- 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.

Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

EM Algorithm Summary

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

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

- 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.

Established 1911

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

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

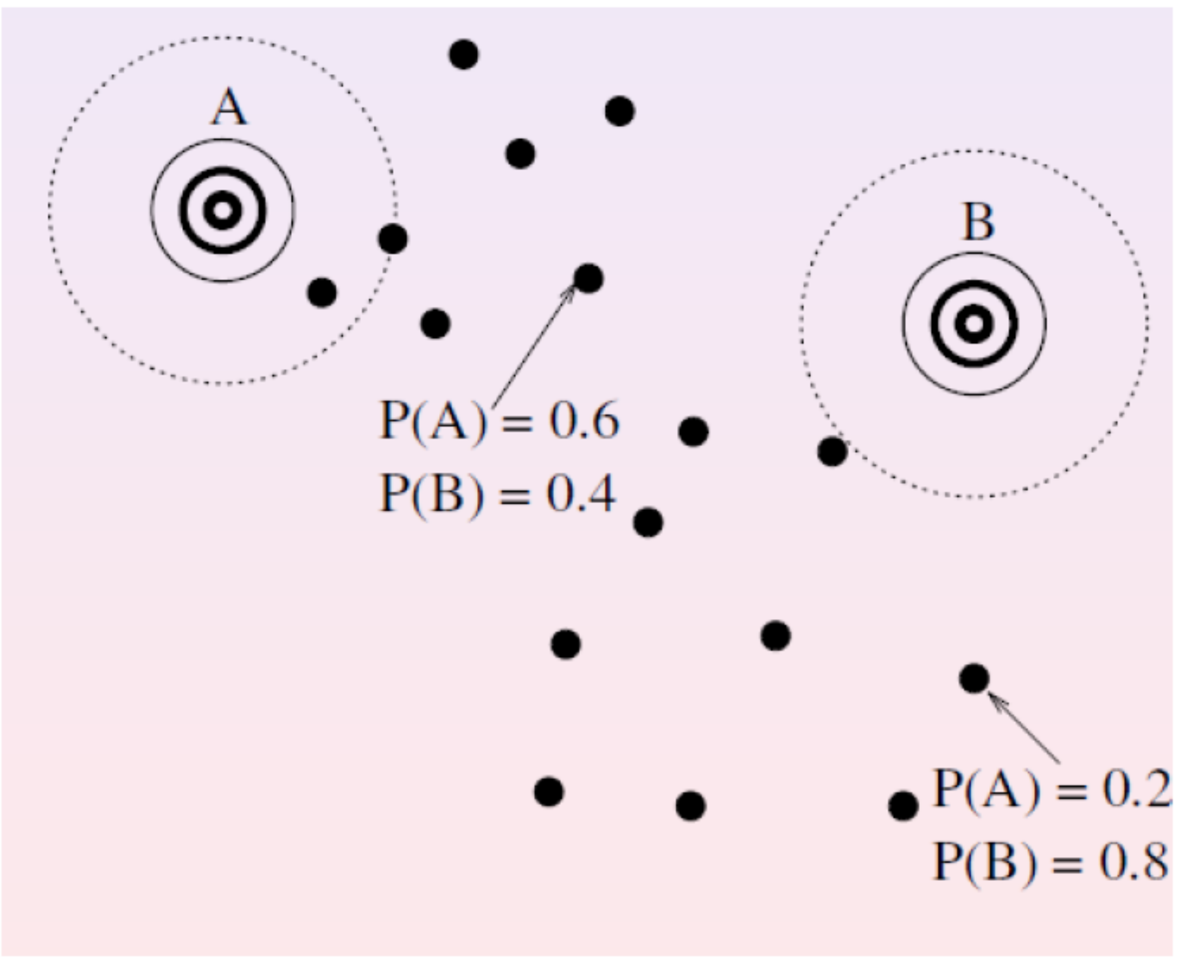
$$\mu_\ell^{new} = \frac{\sum_{i=1}^{N} x_i p(\ell | x_i, \Theta^g)}{\sum_{i=1}^{N} p(\ell | x_i, \Theta^g)}$$

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

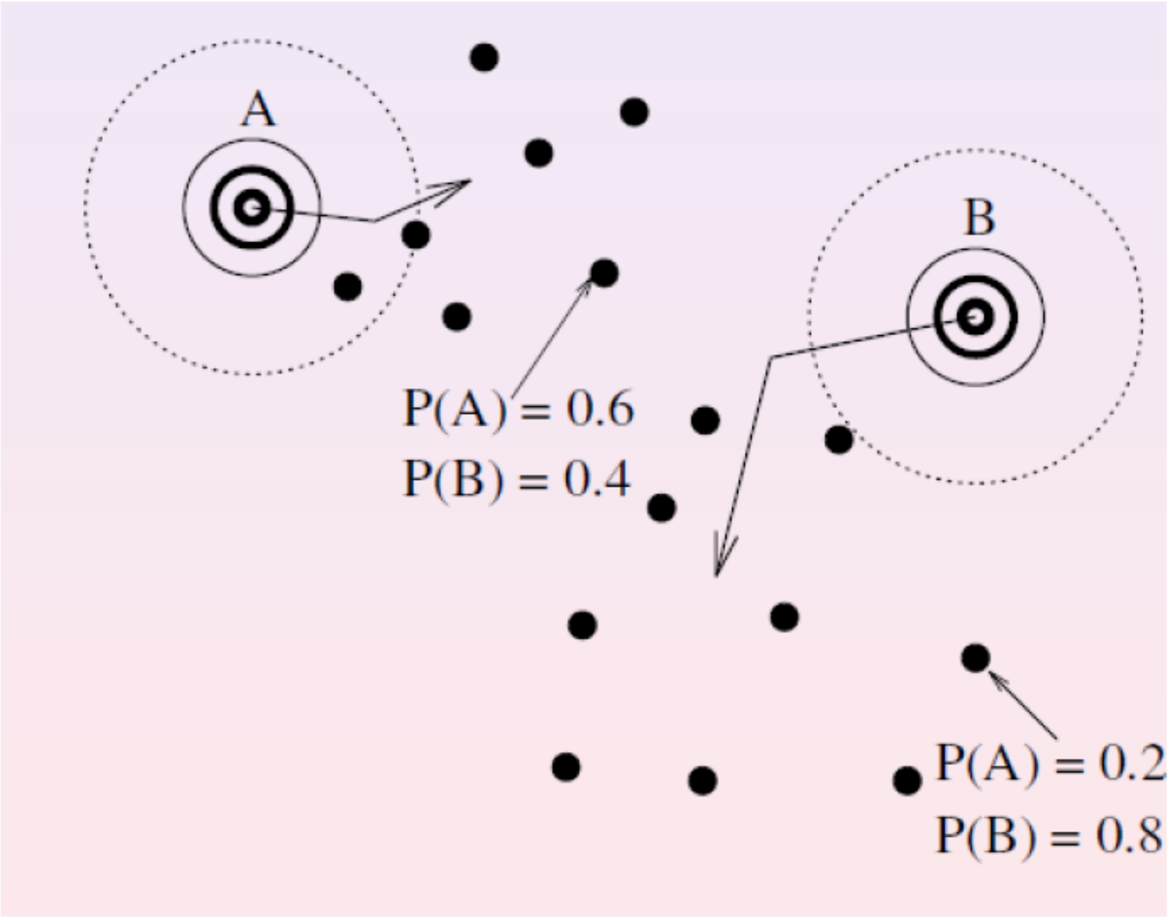
Established 1911

EM Algorithm for GMM

E-step



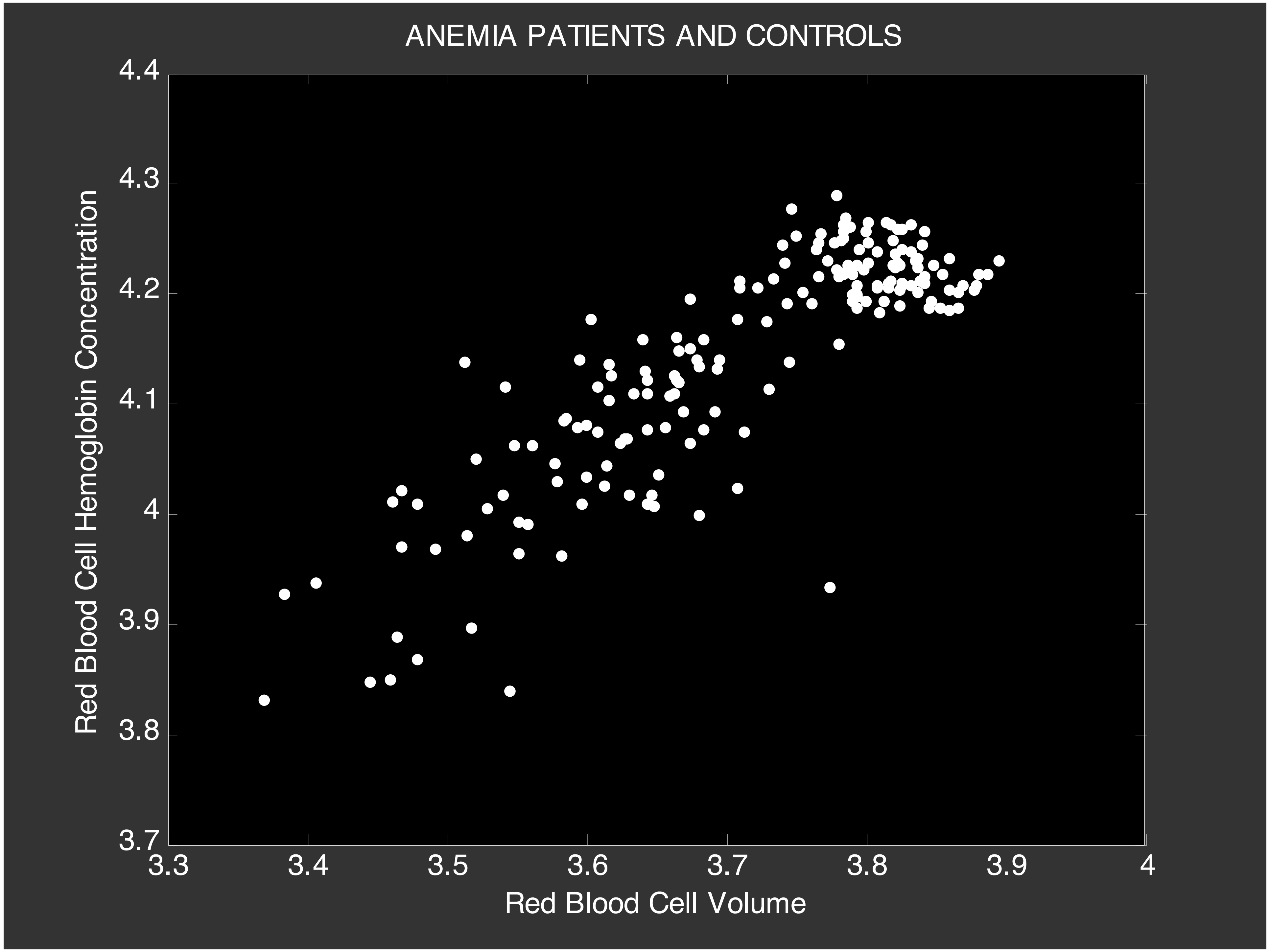
M-step



Established 1911

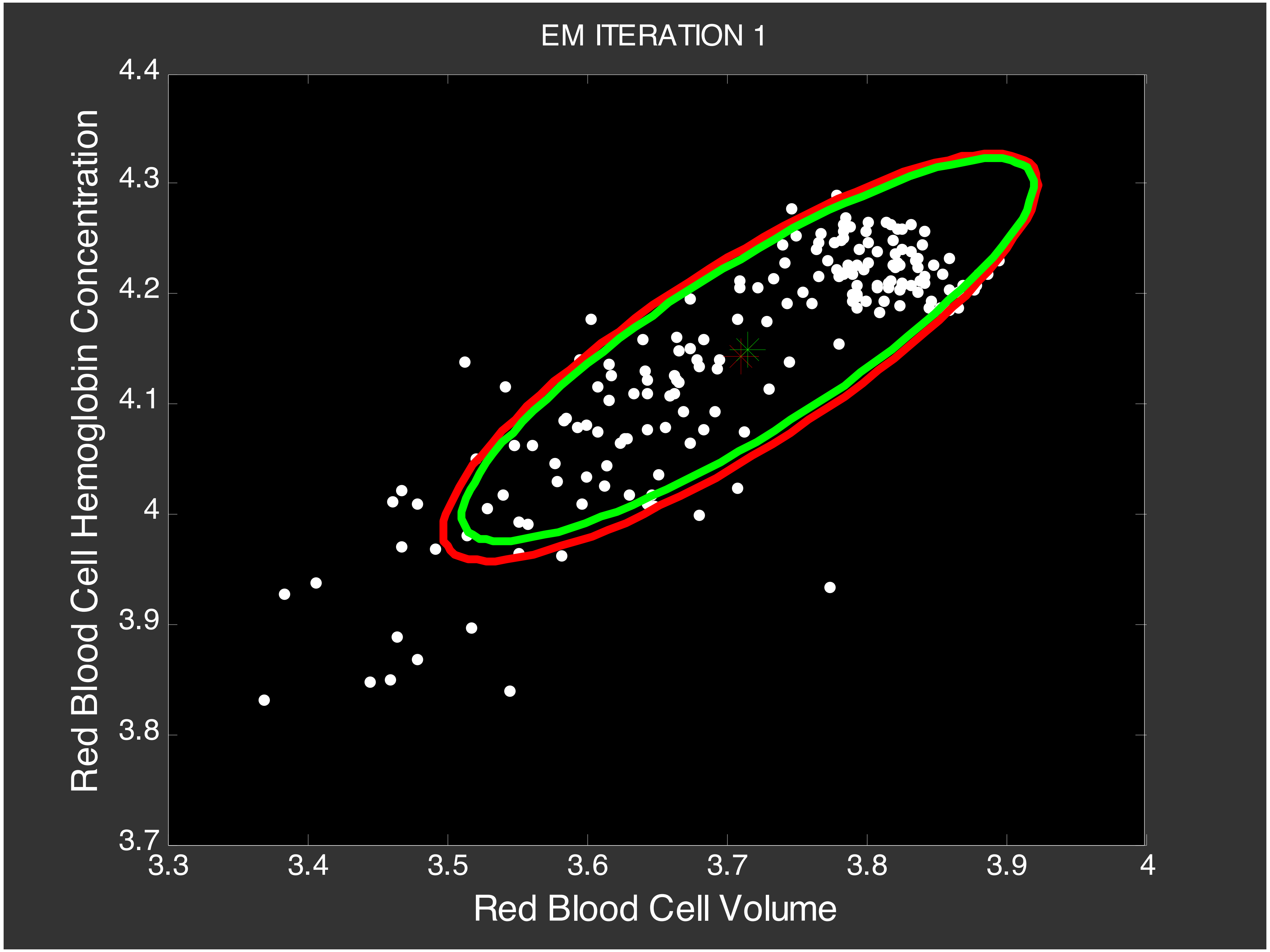
INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

EM Algorithm for GMM



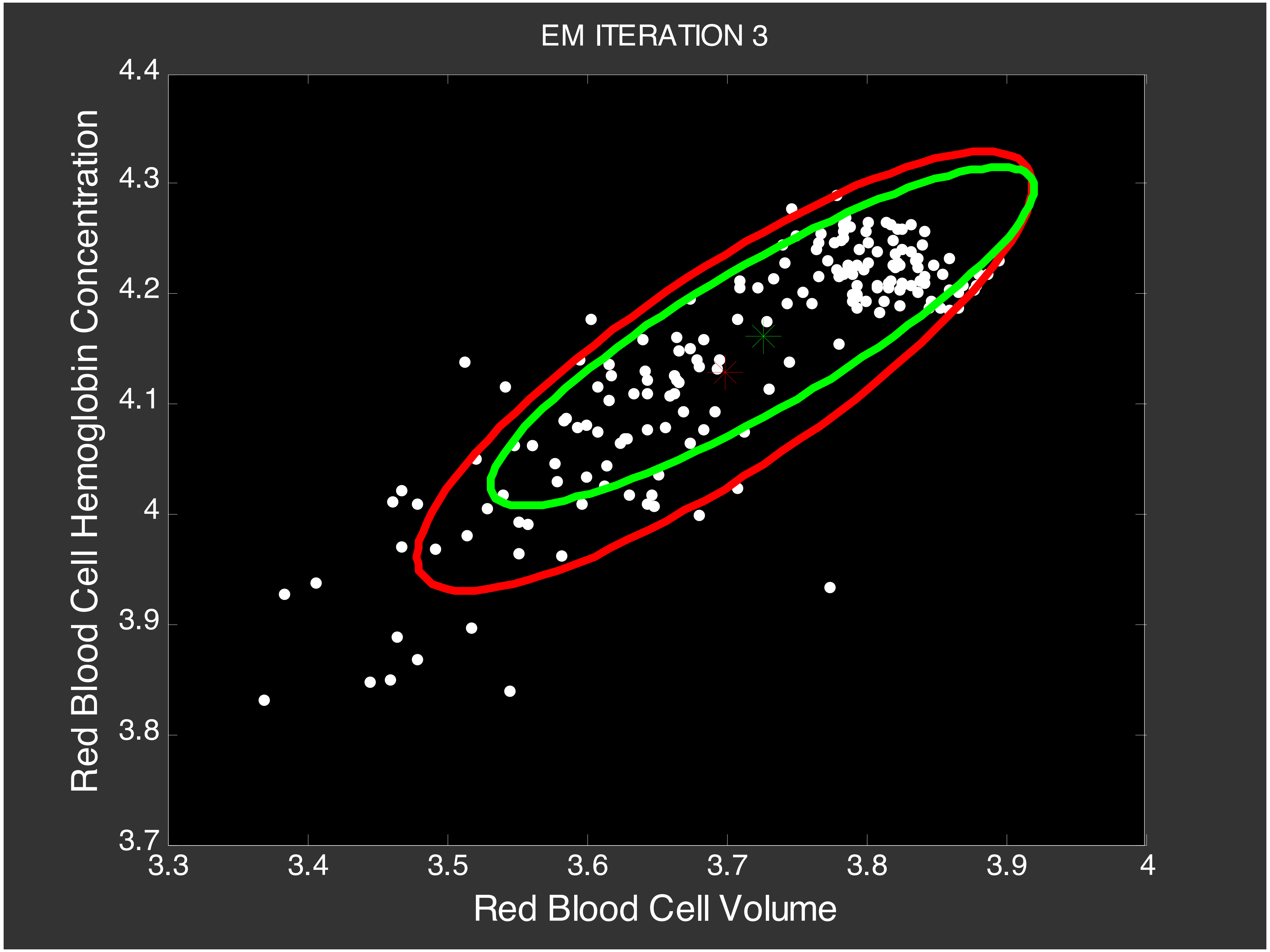
Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

EM Algorithm for GMM



Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

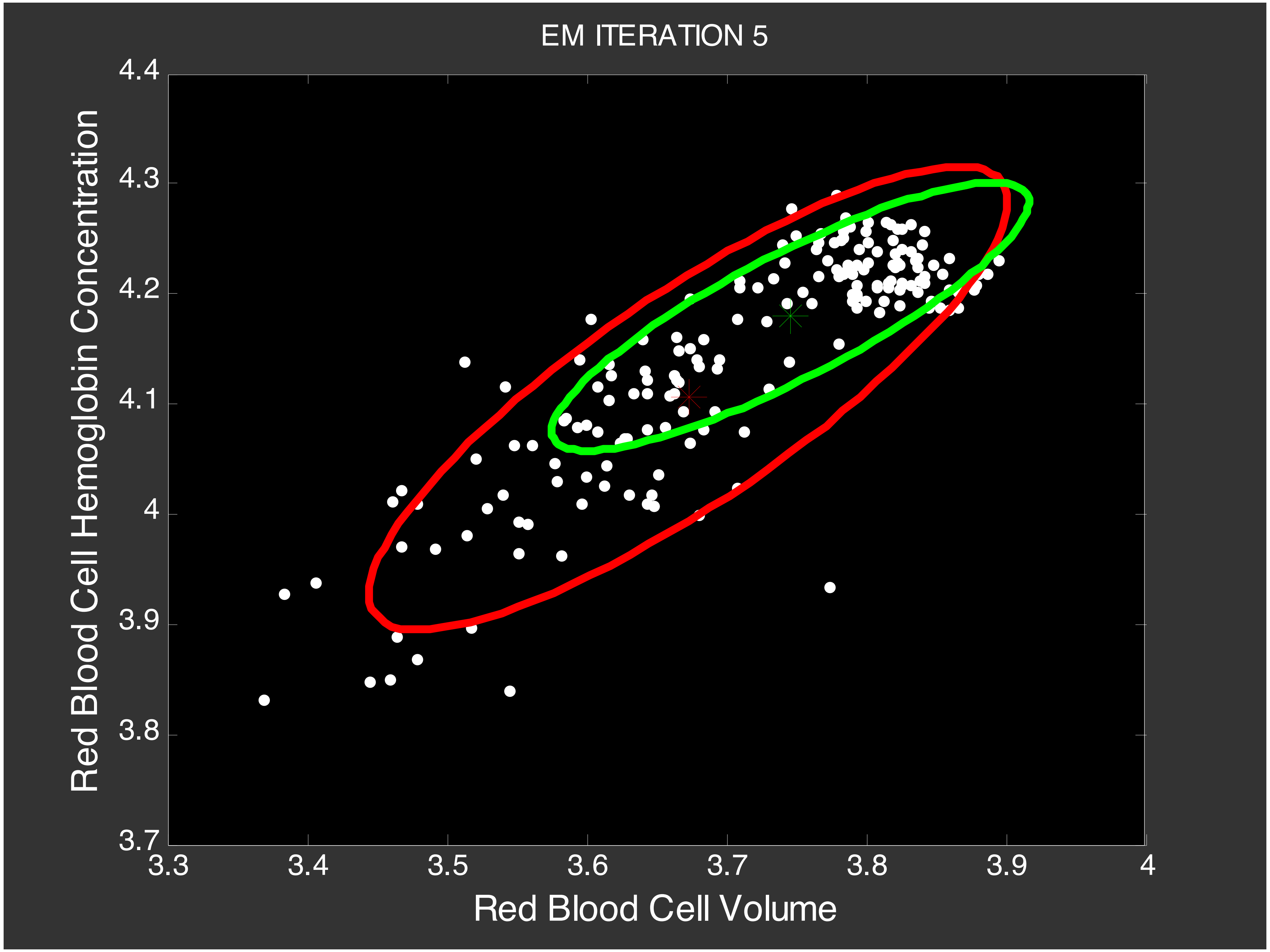
EM Algorithm for GMM



Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

Established 1911

EM Algorithm for GMM

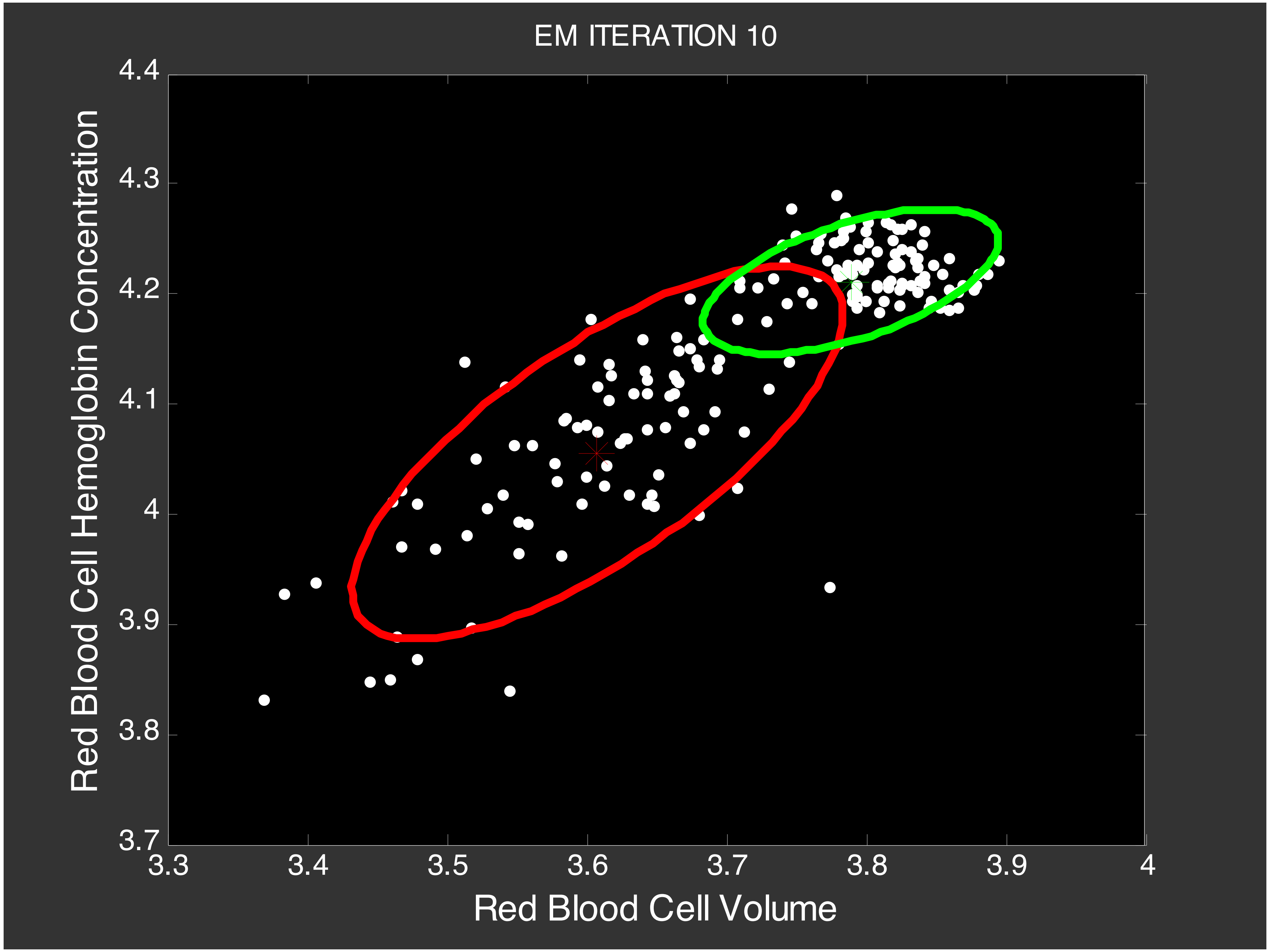


Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

Established 1911

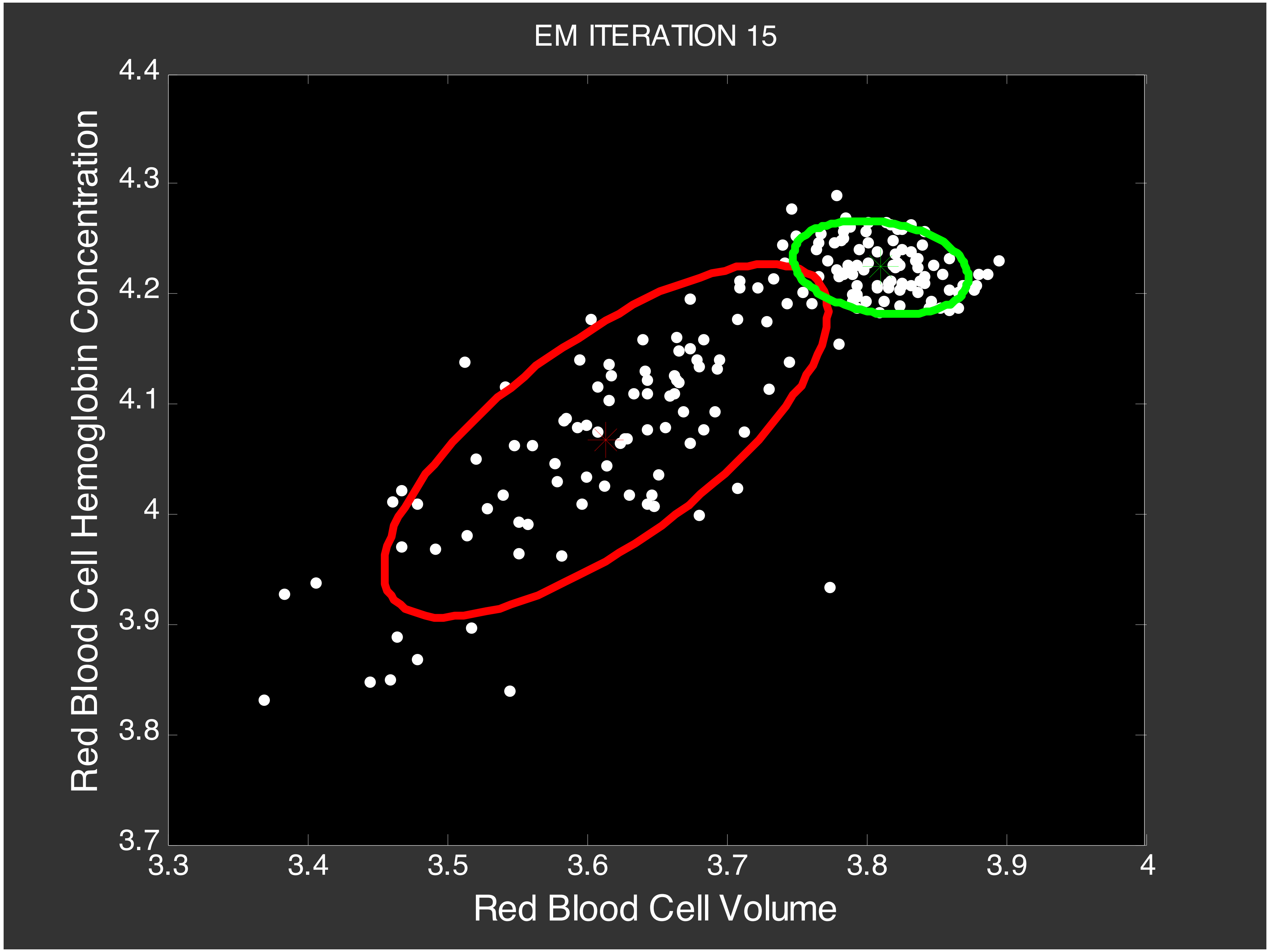
INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

EM Algorithm for GMM



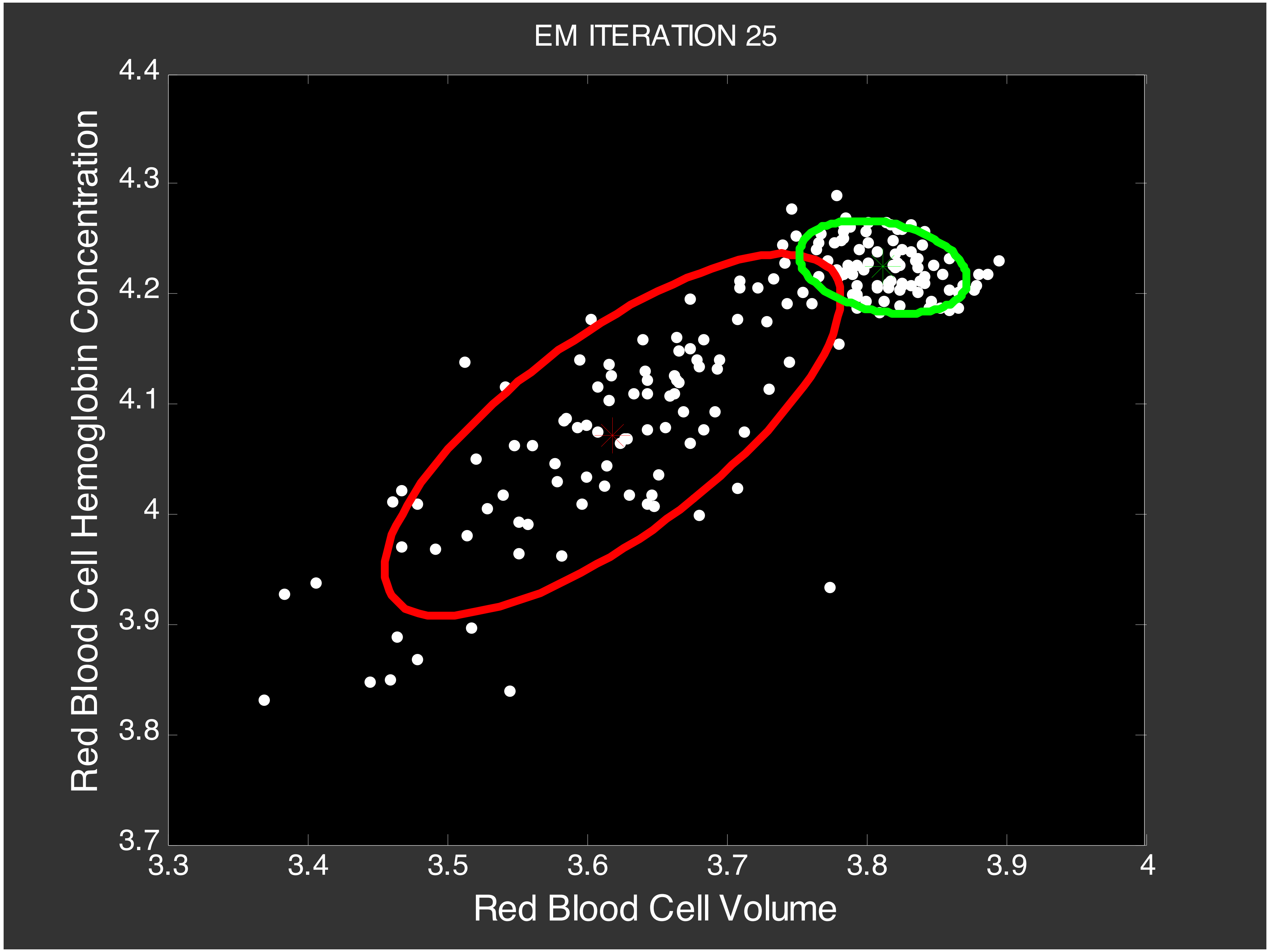
Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

EM Algorithm for GMM



Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

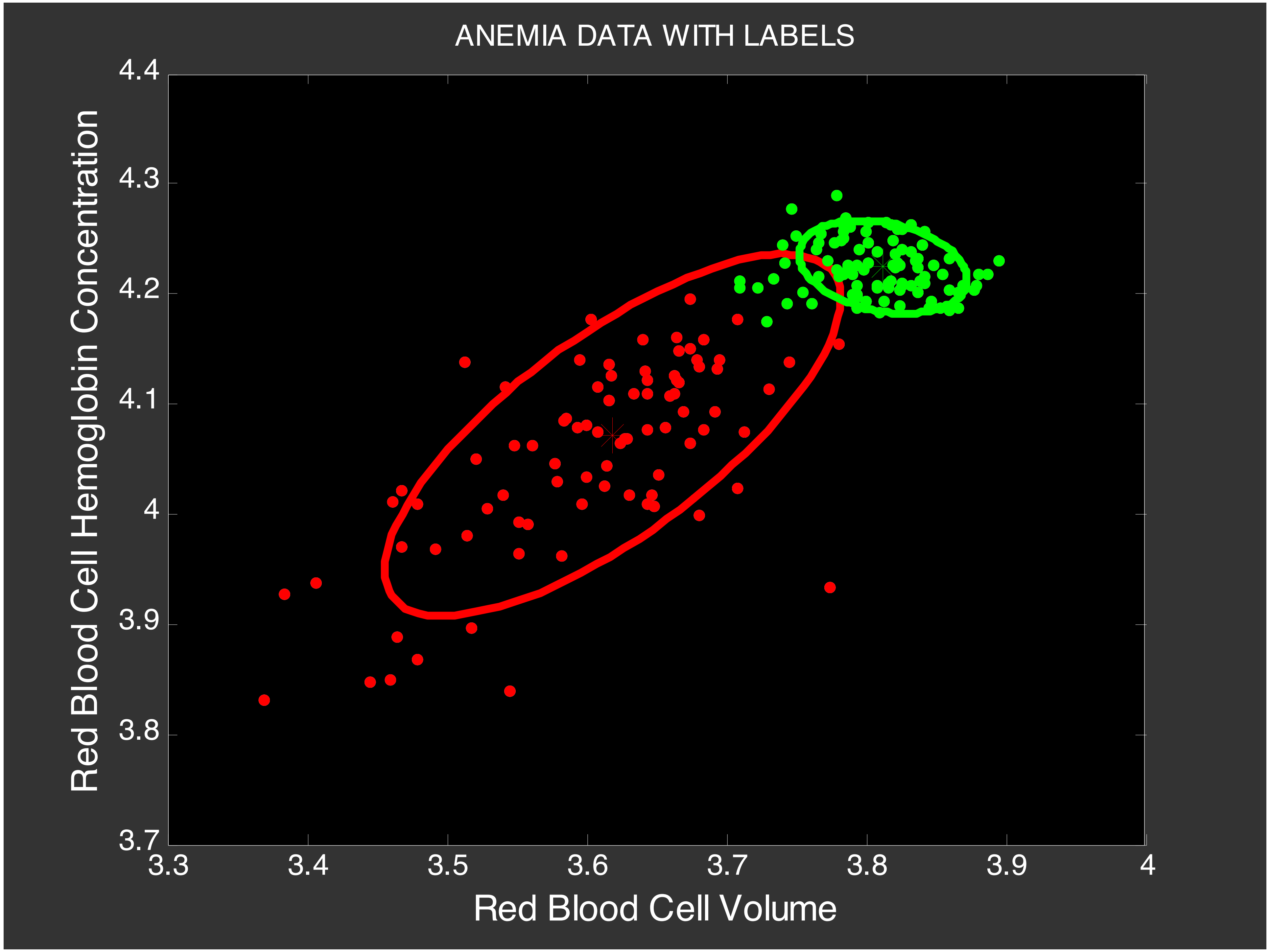
EM Algorithm for GMM



Established 1911

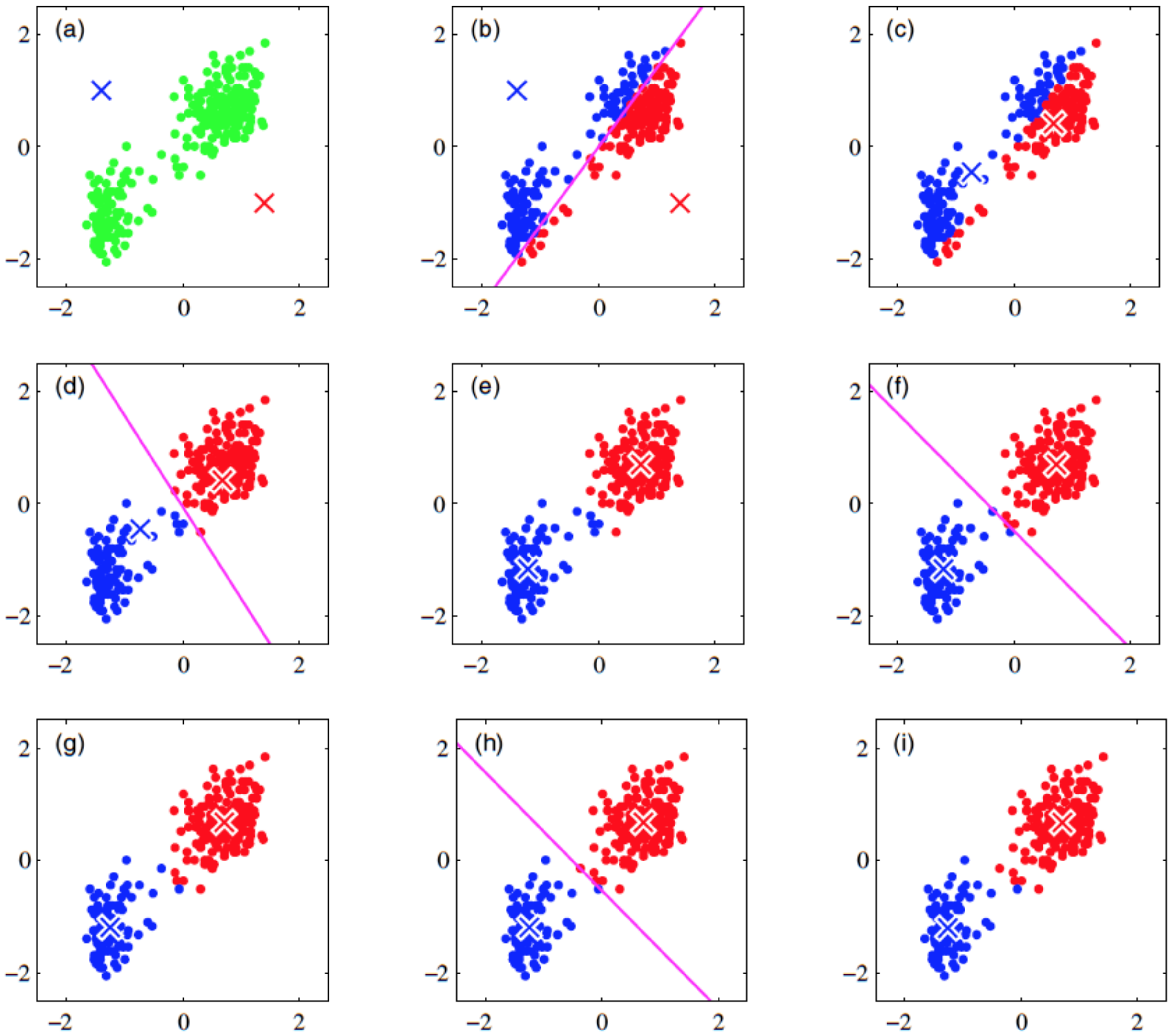
Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

EM Algorithm for GMM



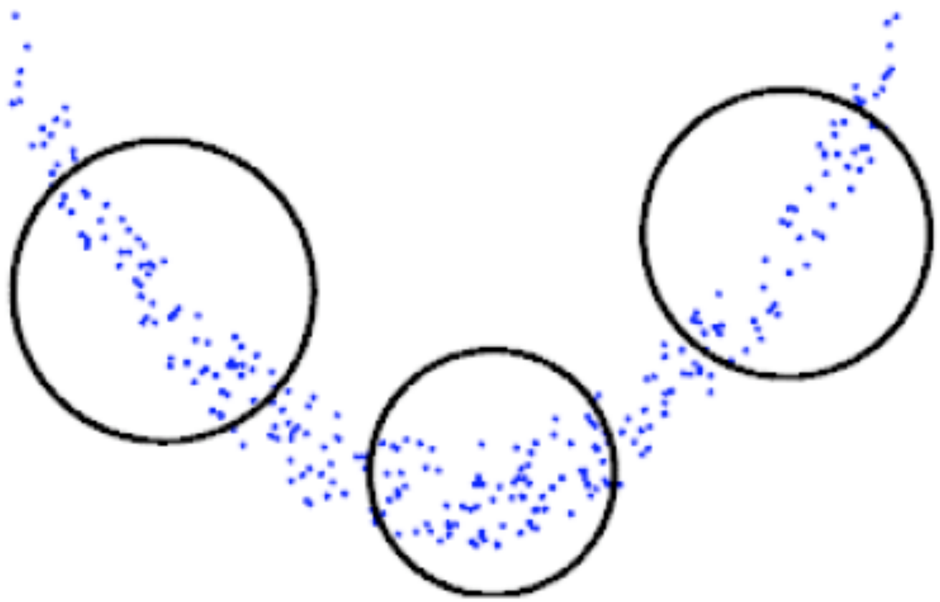
Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

K-means Algorithm for Initialization



Other Considerations

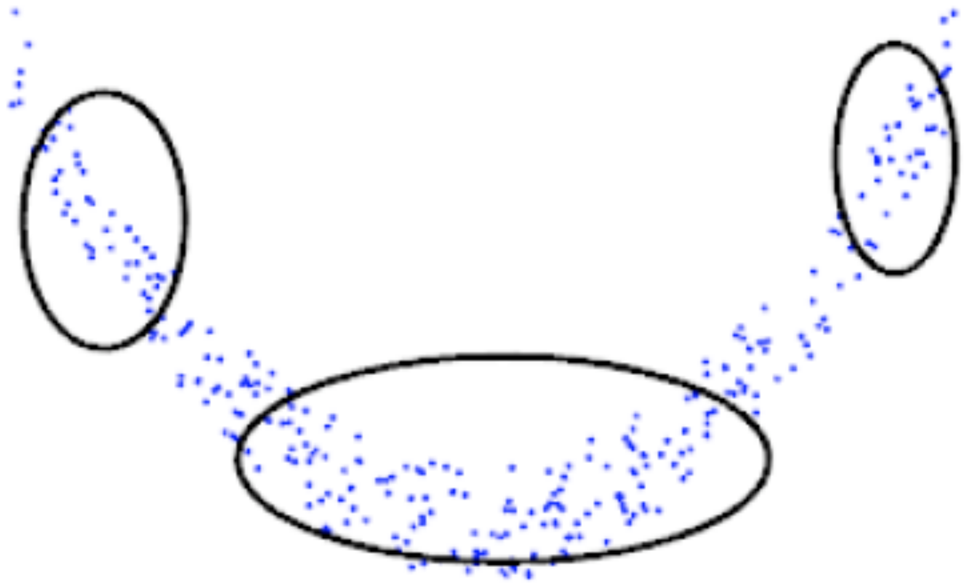
- Initialization - random or k-means
- Number of Gaussians
- Type of Covariance matrix
  - Spherical covariance



-Less precise.
-Very efficient to compute.

Other Considerations

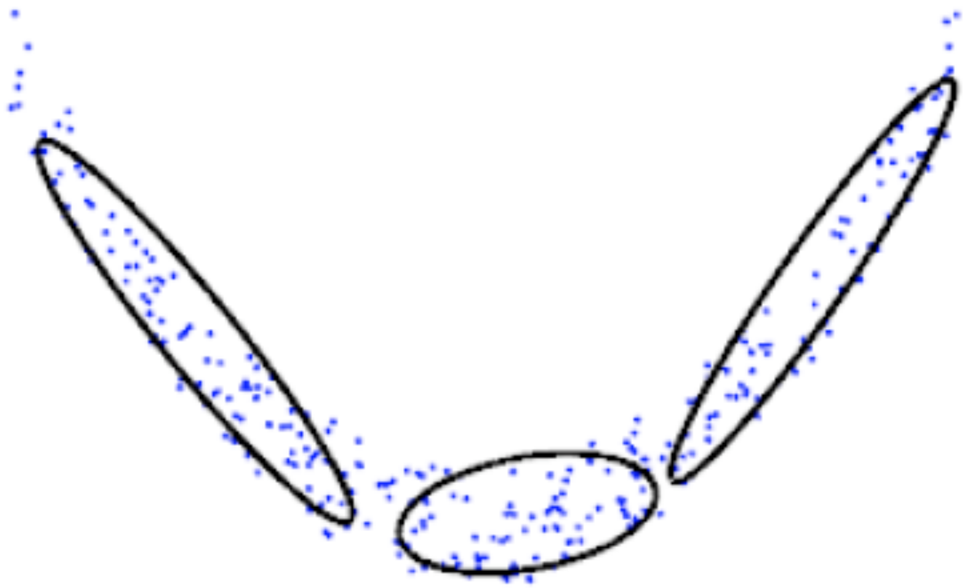
- Initialization - random or k-means
- Number of Gaussians
- Type of Covariance matrix
  - Diagonal covariance



-More precise.
-Efficient to compute.

Other Considerations

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



-Very precise.
-Less efficient to compute.