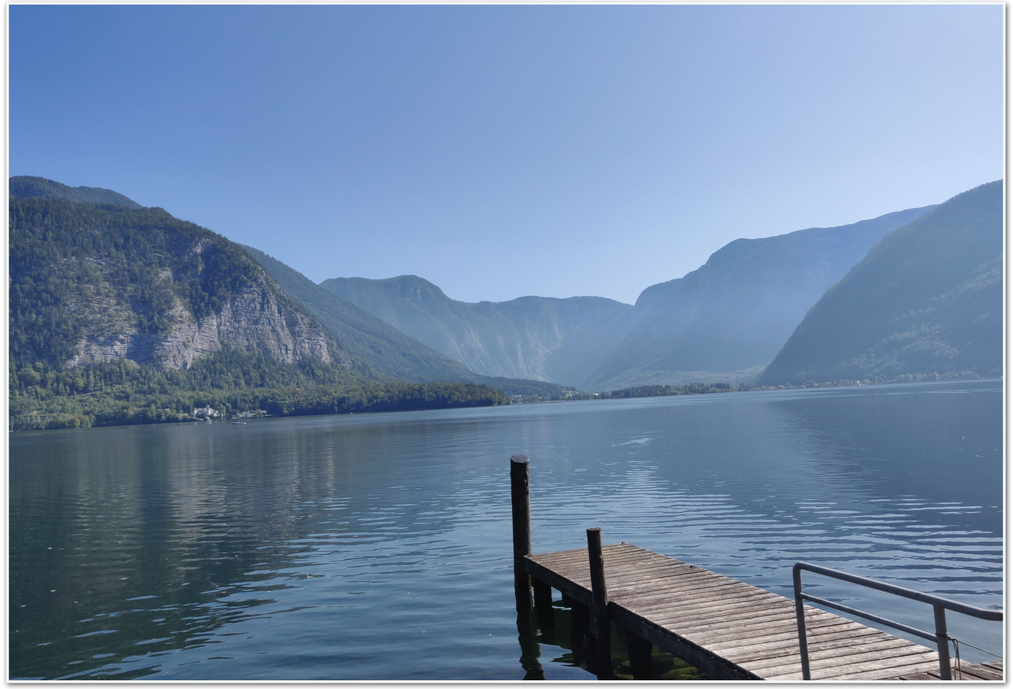
MACHINE LEARNING FOR SIGNAL PROCESSING



27-1-2025

Sriram Ganapathy
LEAP lab, Electrical Engineering, Indian Institute of Science,
sriramg@iisc.ac.in

Viveka Salinamakki, Varada R.
LEAP lab, Electrical Engineering, Indian Institute of Science

http://leap.ee.iisc.ac.in/sriram/teaching/MLSP25/

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

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

MLE for GMM

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

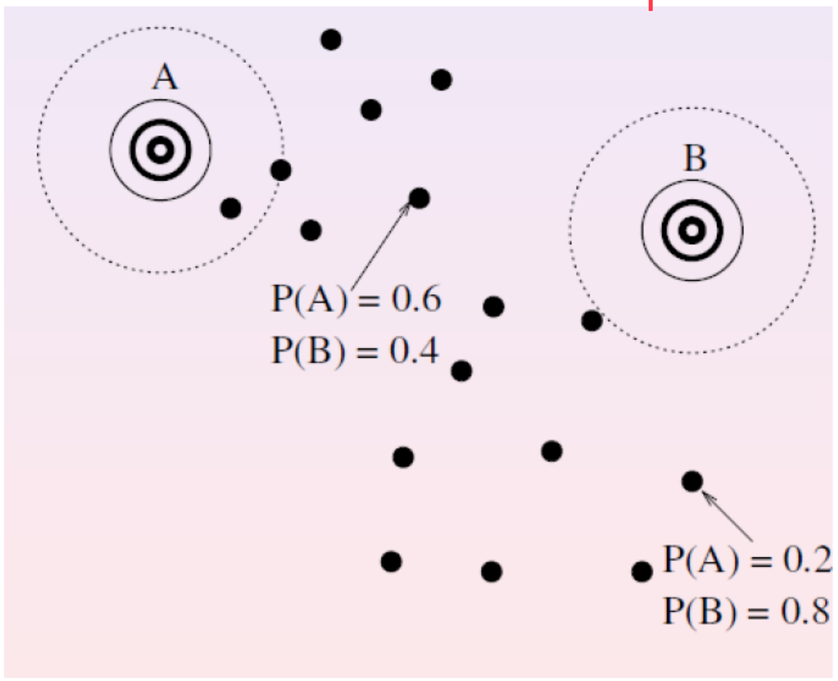
$$\log L(\boldsymbol{\Theta}) = \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} \alpha_k p(\mathbf{x}_i | \boldsymbol{\theta}_k) \right)$$

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

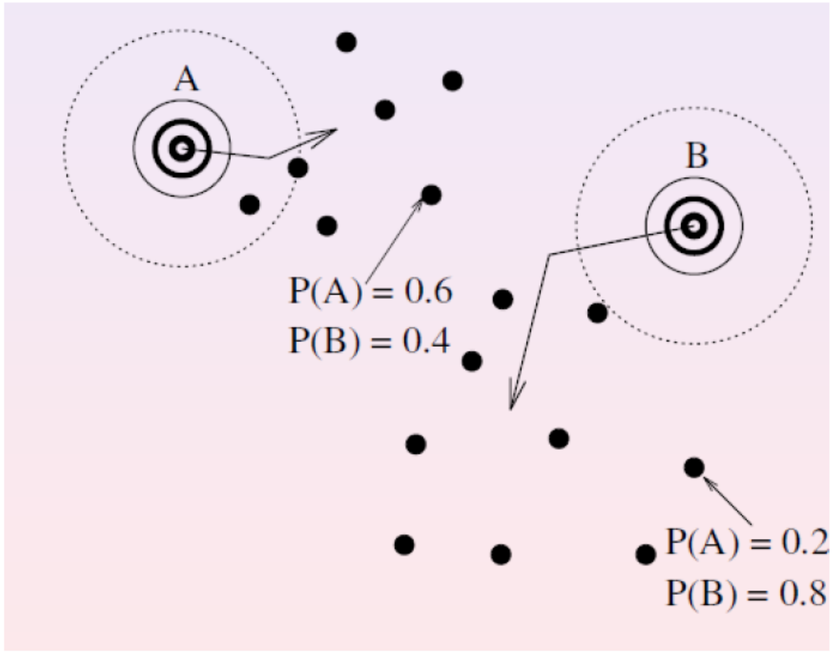
EM Algorithm for GMM

- The hidden variables will be the index of the mixture component which generated
- Re-estimation formulae

E-step



M-step



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

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

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

Expectation Maximization Algorithm

- It can be proven that if we choose

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

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

- 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

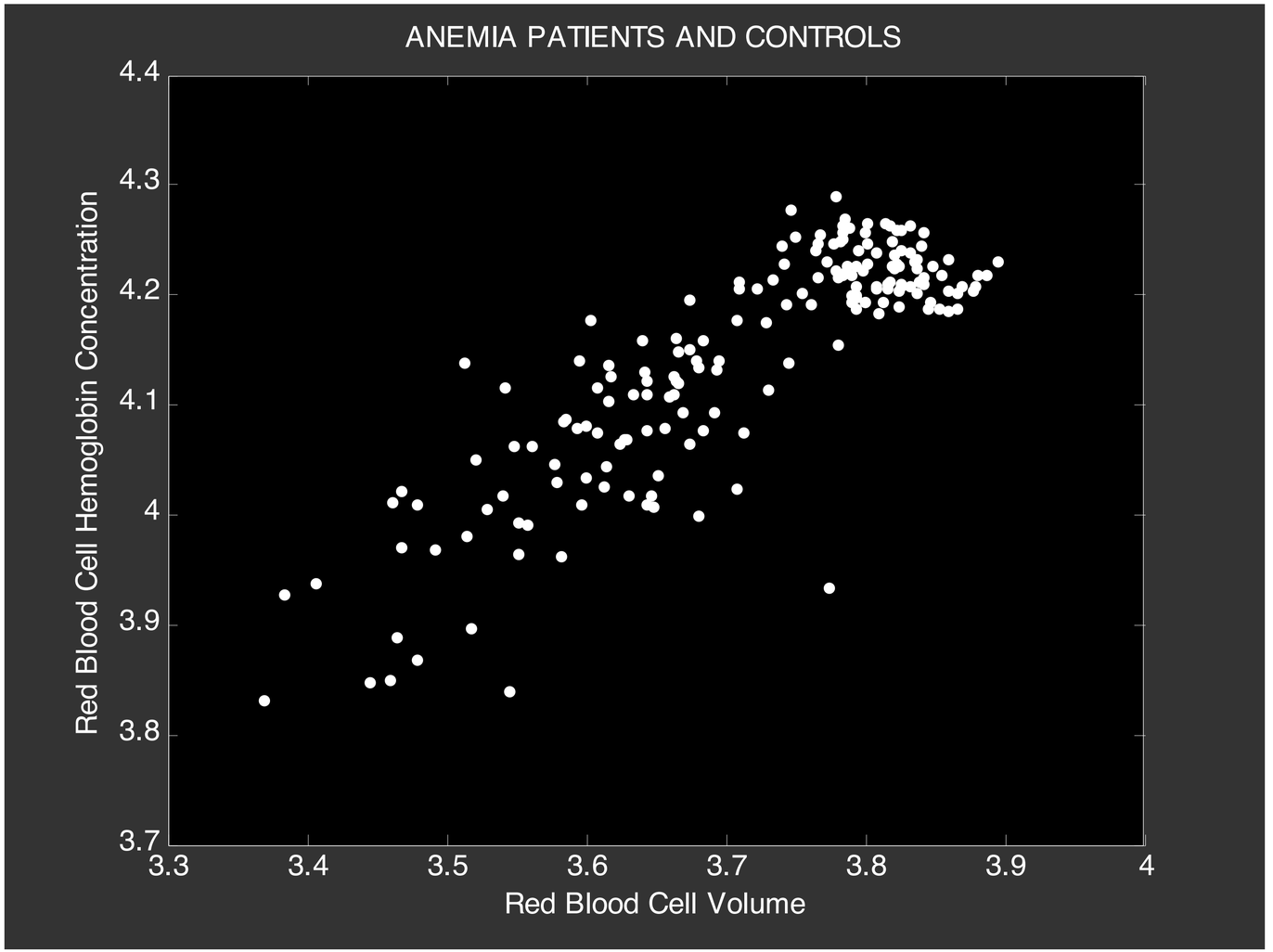
- The hidden variables $\mathbf{z}_i = l$ will be the index of the mixture component which generated $\mathbf{x}_i$
- Re-estimation formulae

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

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

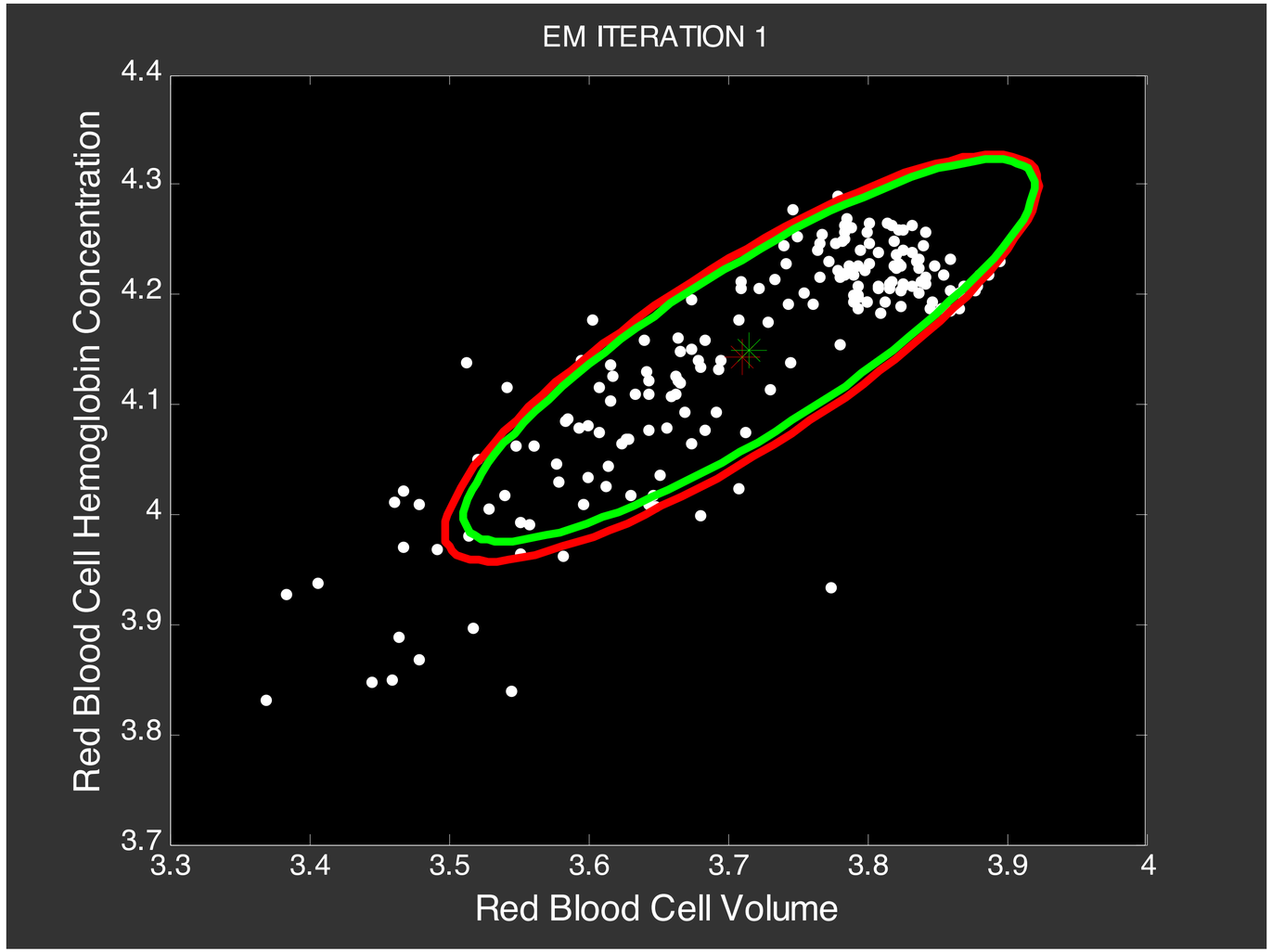
$$\Sigma_\ell^{new} = \frac{\sum_{i=1}^{N} p(\ell|x_i, \Theta^g)(x_i - \mu_\ell^{new})(x_i - \mu_\ell^{new})^T}{\sum_{i=1}^{N} p(\ell|x_i, \Theta^g)}$$

EM Algorithm for GMM



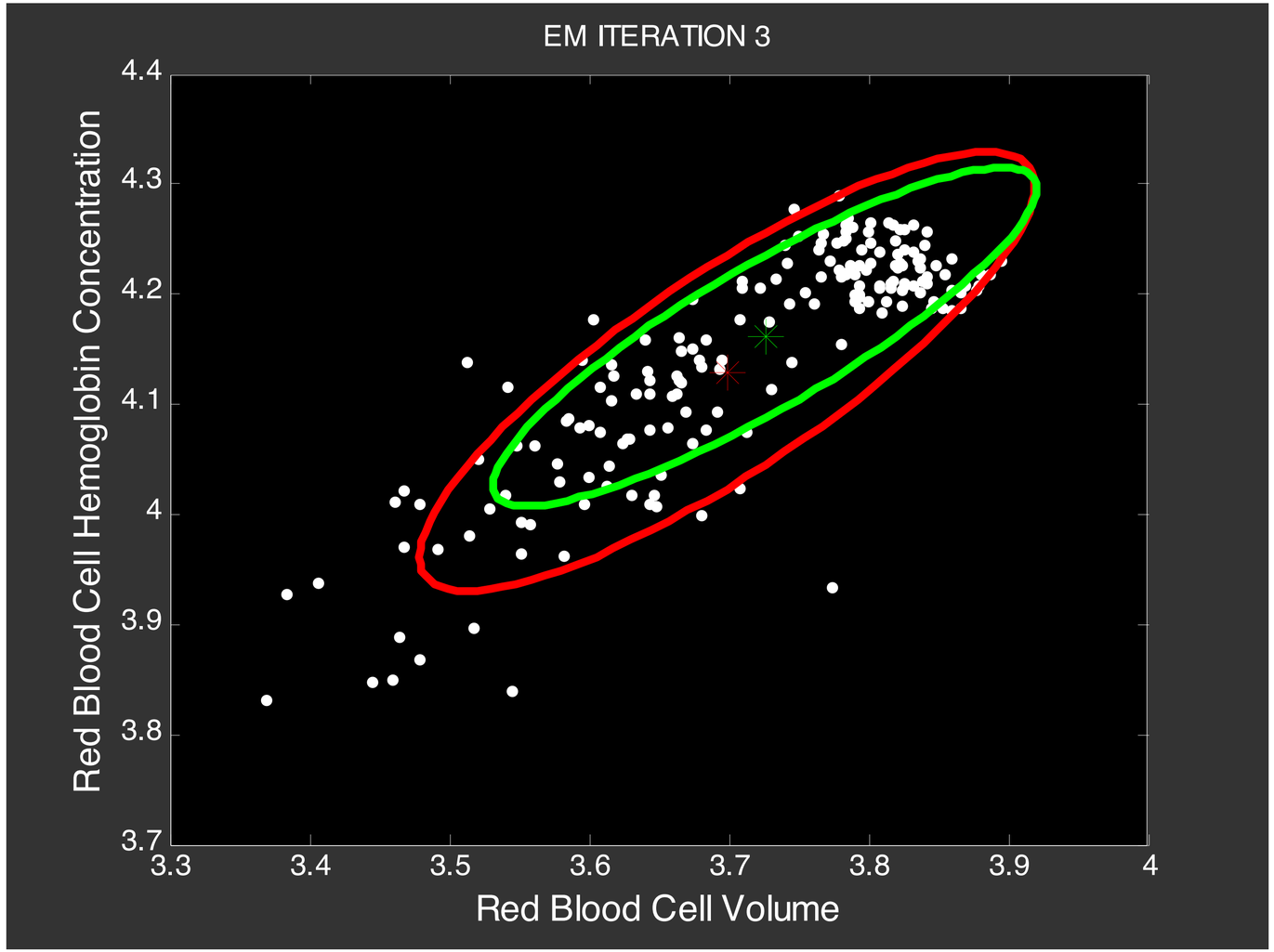
Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

EM Algorithm for GMM



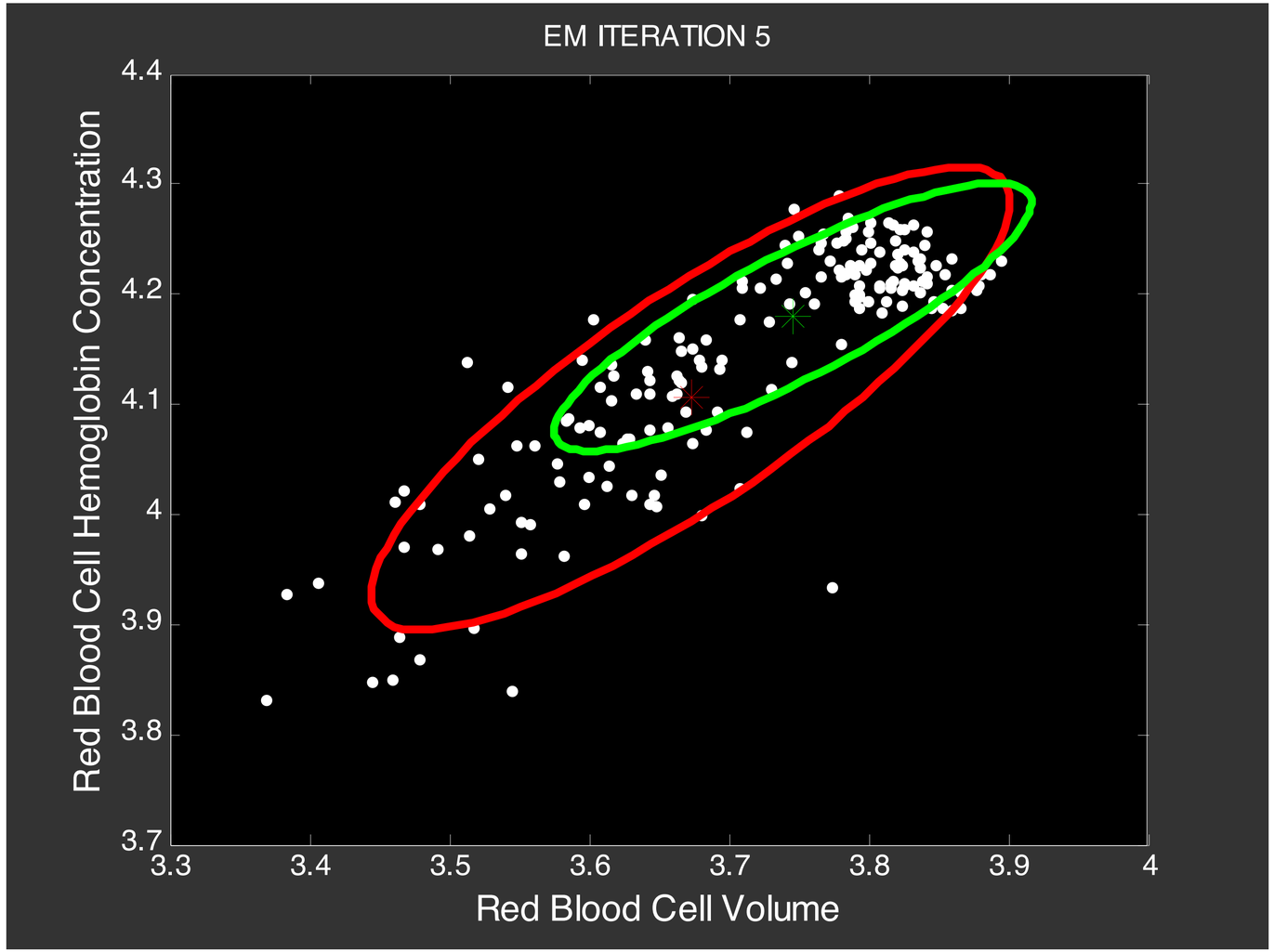
Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

EM Algorithm for GMM



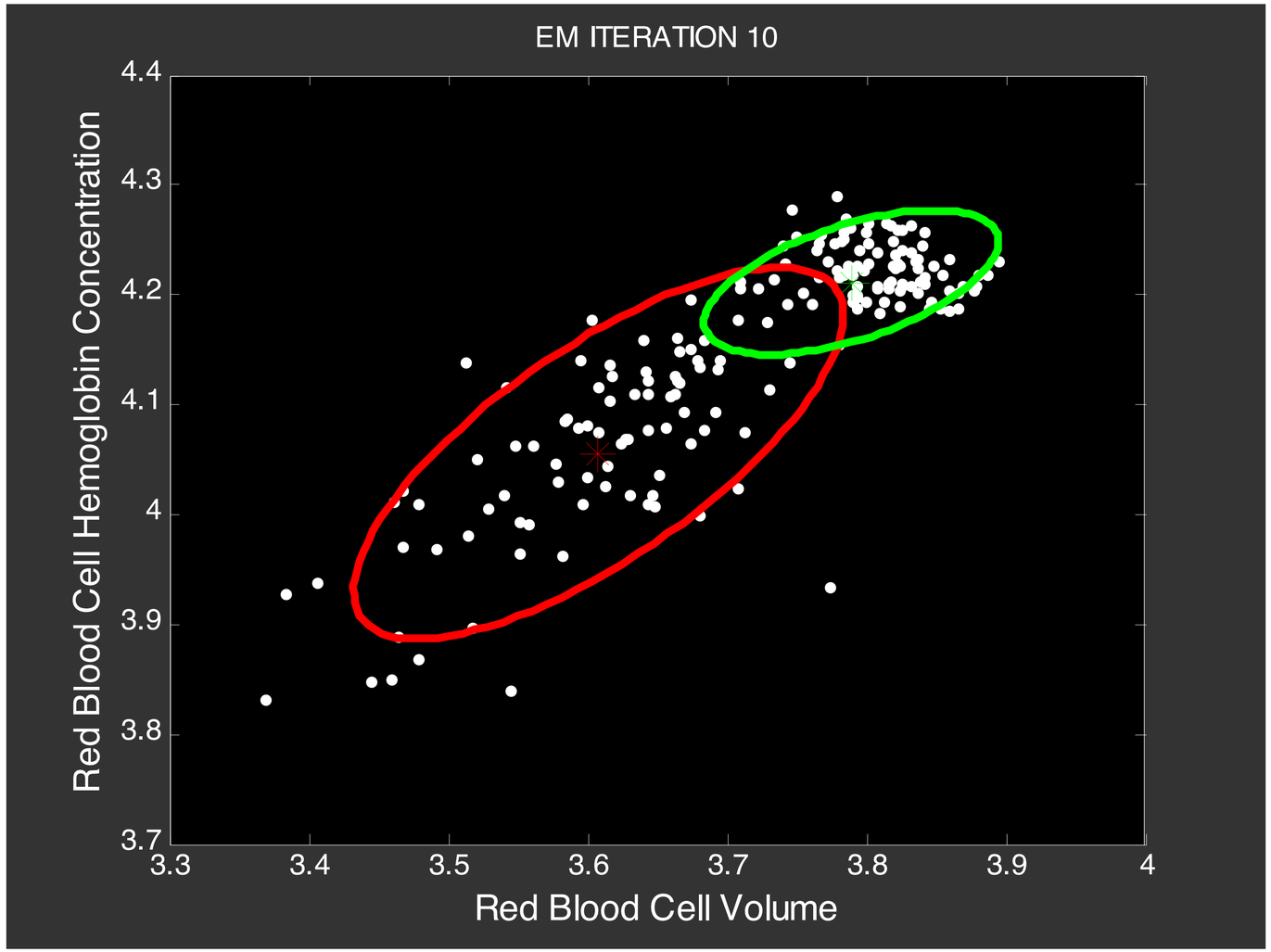
Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

EM Algorithm for GMM



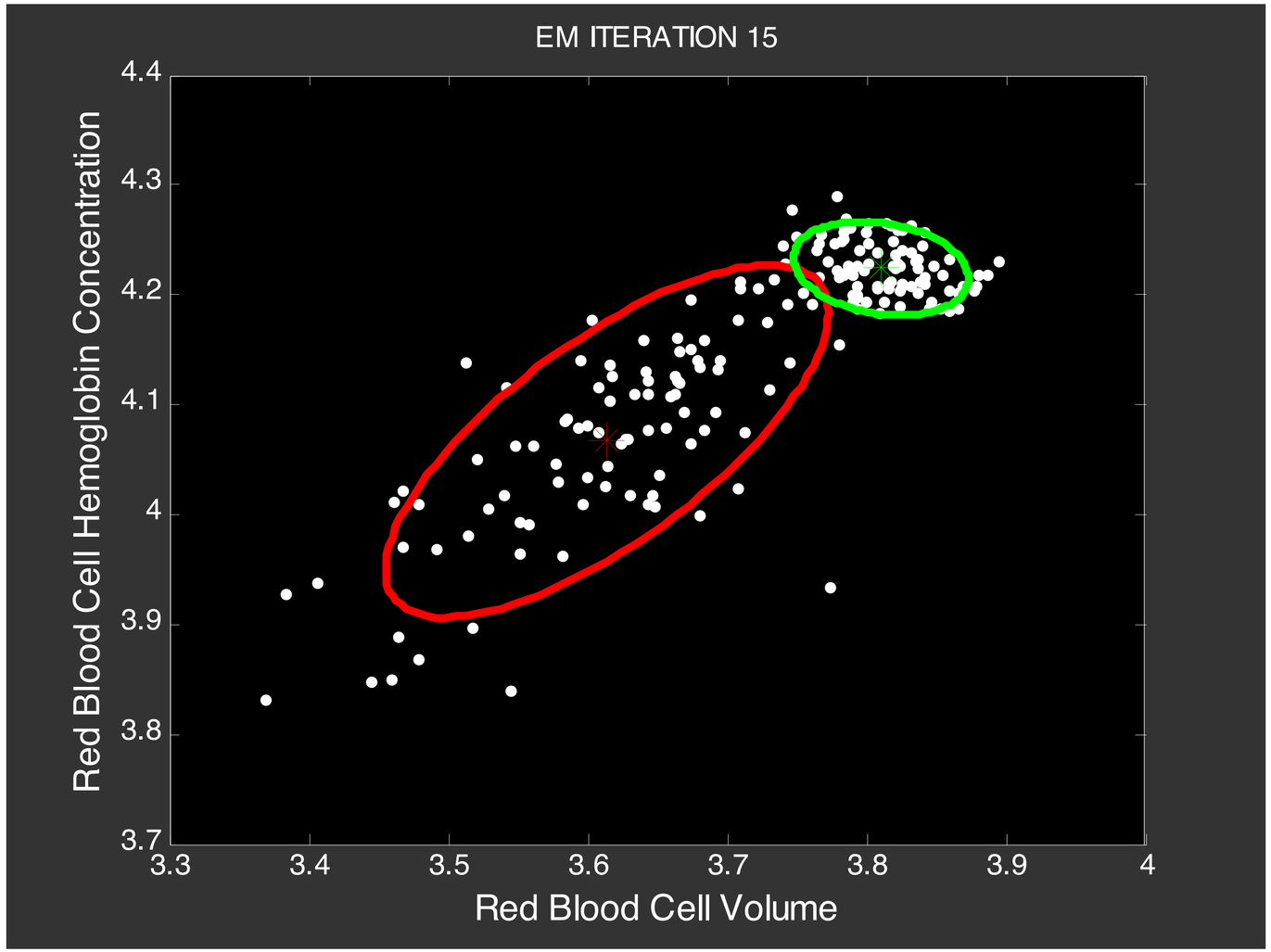
Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

EM Algorithm for GMM



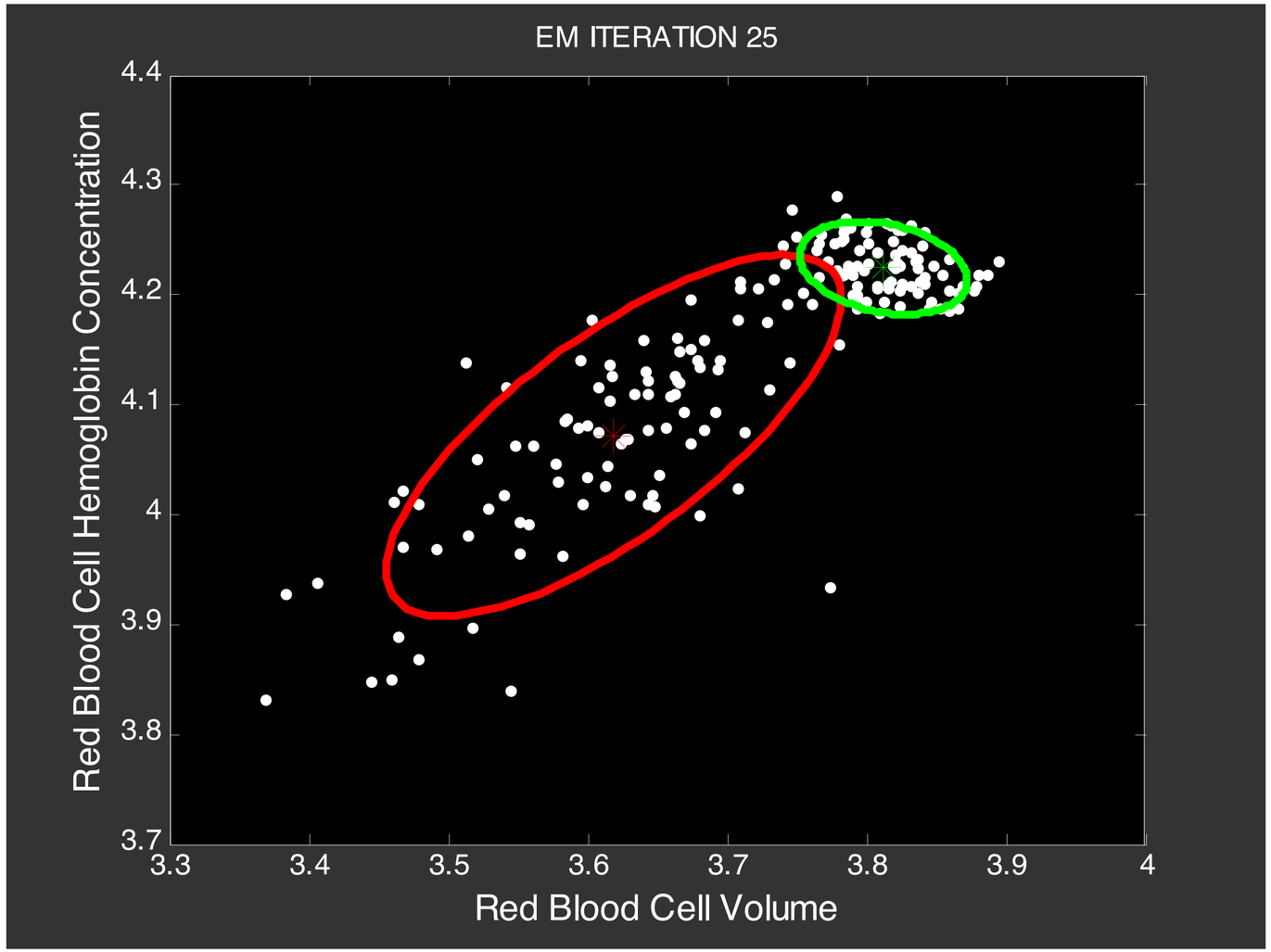
Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

EM Algorithm for GMM



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

EM Algorithm for GMM

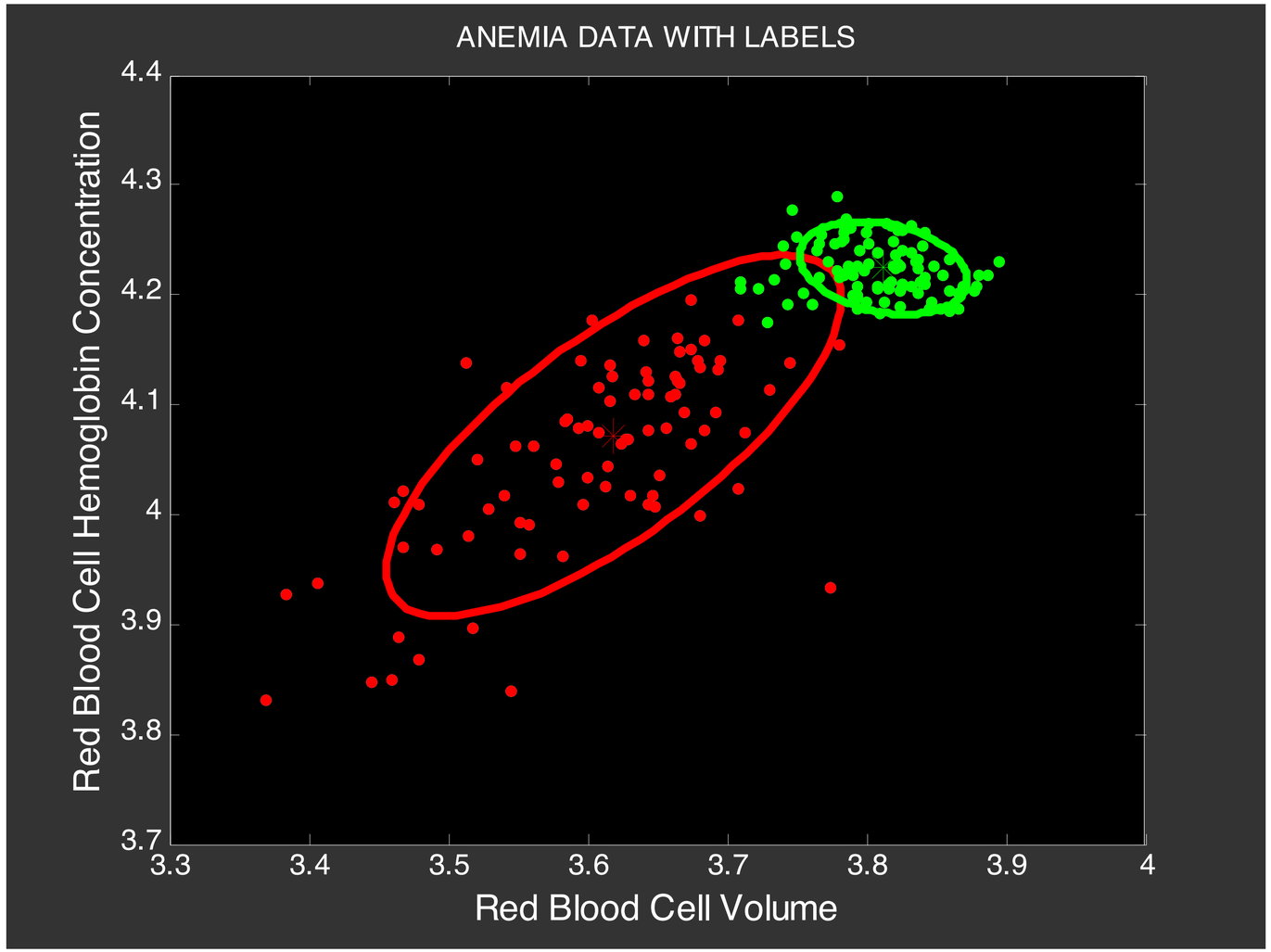


Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

Established 1911

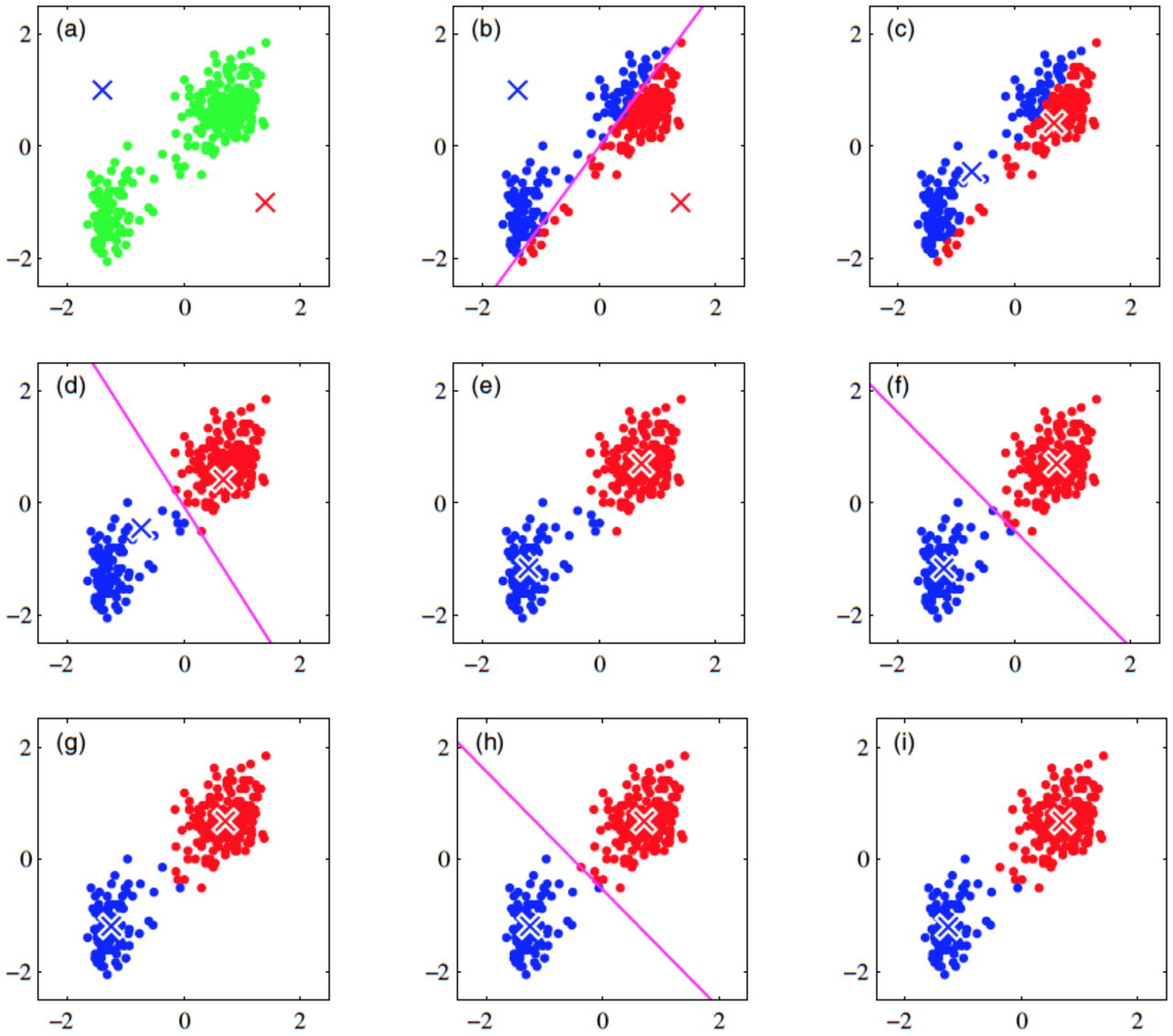
INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

EM Algorithm for GMM



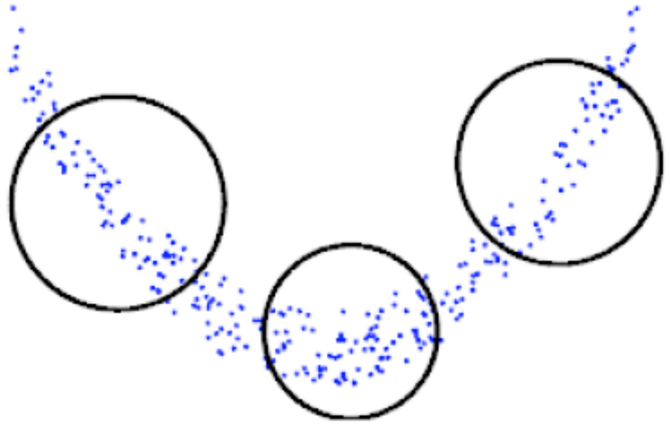
Cadez, Igor V., et al. "Hierarchical models for screening of iron deficiency anemia." *ICML*. 1999.

K-means Algorithm for Initialization



Other Considerations

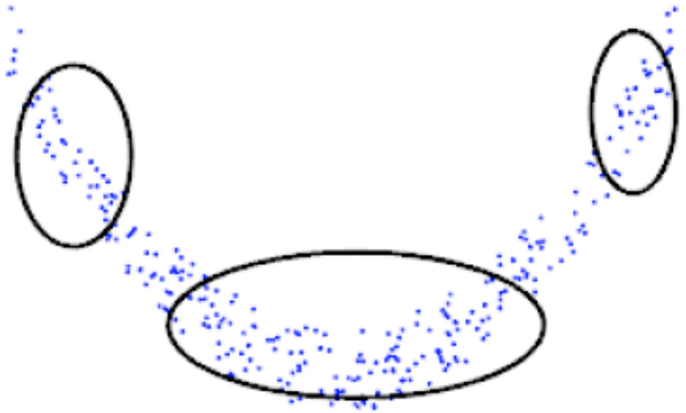
- Initialization - random or k-means
- Number of Gaussians
- Type of Covariance matrix
  - Spherical covariance



-Less precise.
-Very efficient to compute.

Other Considerations

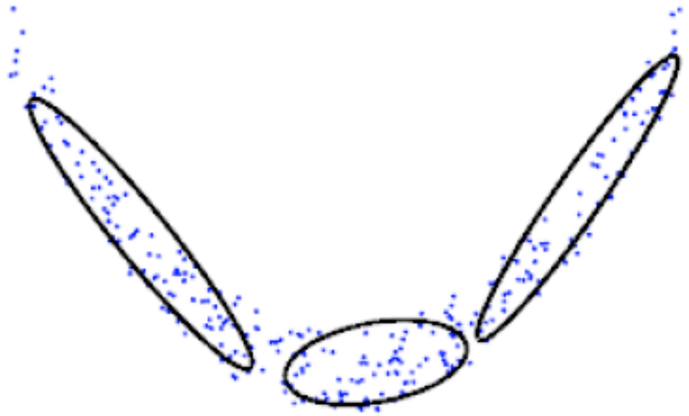
- Initialization - random or k-means
- Number of Gaussians
- Type of Covariance matrix
  - Diagonal covariance



-More precise.
-Efficient to compute.

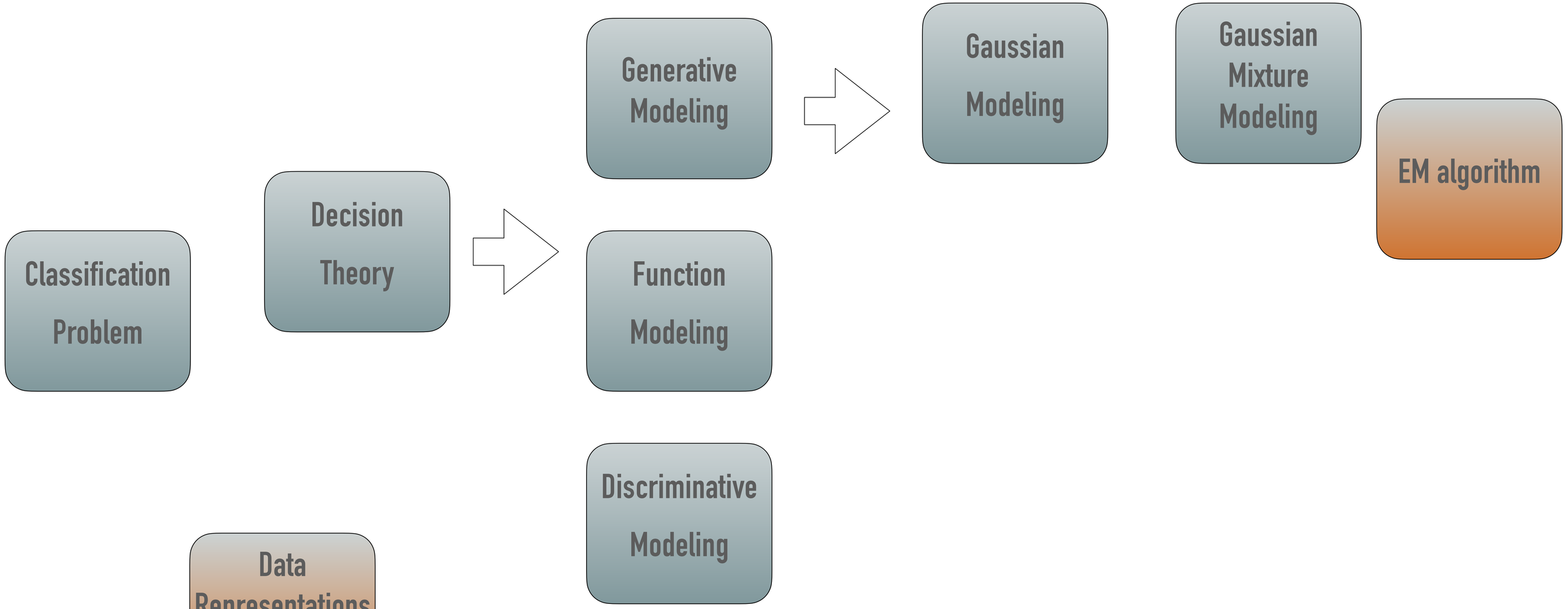
Other Considerations

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



-Very precise.
-Less efficient to compute.

STORY SO FAR


Classification Problem
Decision Theory
Generative Modeling
Function Modeling
Discriminative Modeling
Gaussian Modeling
Gaussian Mixture Modeling
EM algorithm
Data Representations PCA, LDA

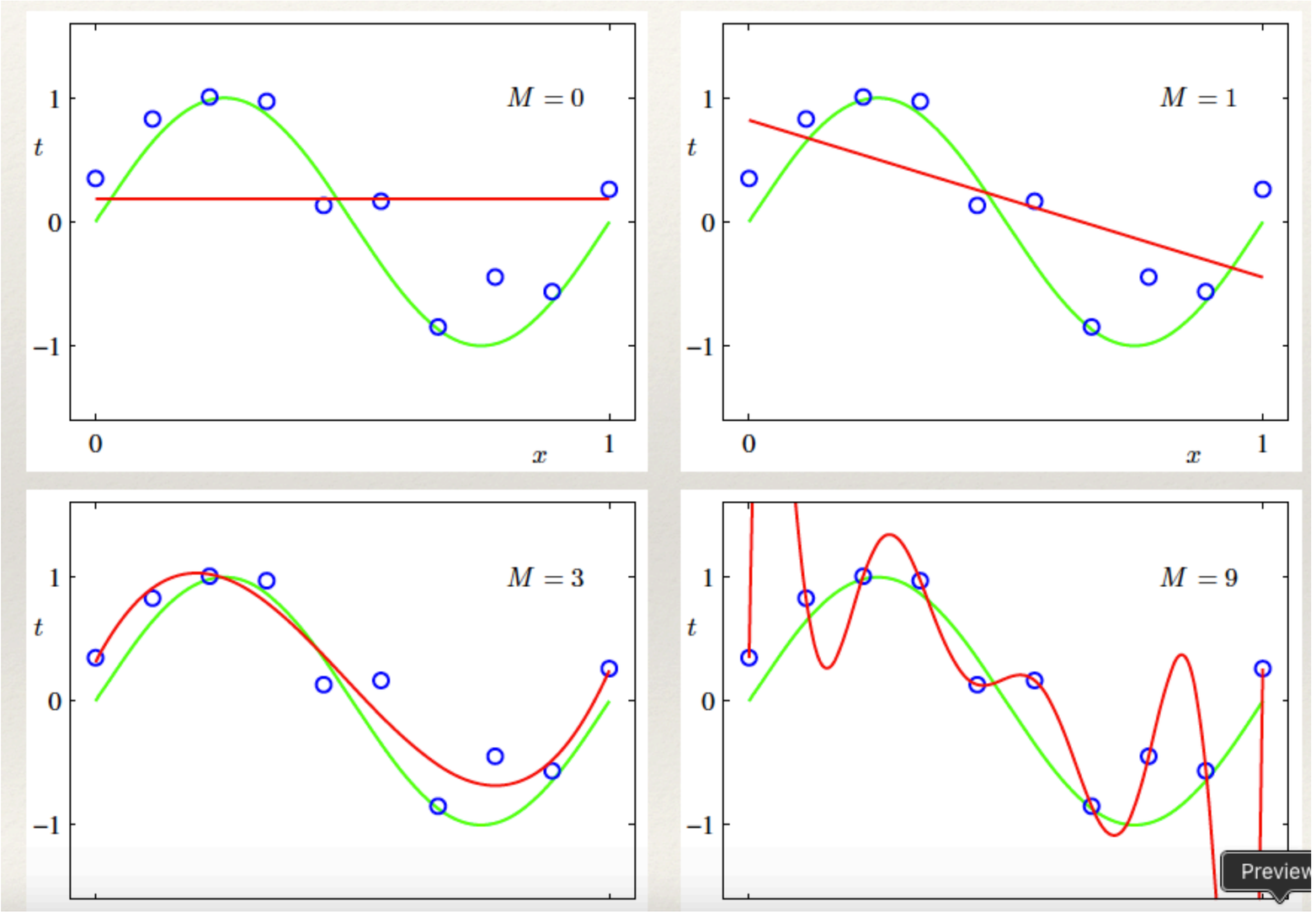
LINEAR REGRESSION

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x})$$

Example

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

LINEAR REGRESSION WITH POLYNOMIAL BASIS



LINEAR REGRESSION

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x})$$

$$t = y(\mathbf{x}, \mathbf{w}) + \epsilon$$

- Solution to Maximum Likelihood problem is the least squares solution

$$\nabla \ln p(\mathbf{t}|\mathbf{w}, \beta) = \sum_{n=1}^{N} \left\{ t_n - \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) \right\} \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}}.$$

LINEAR REGRESSION - CHOICE OF BASIS

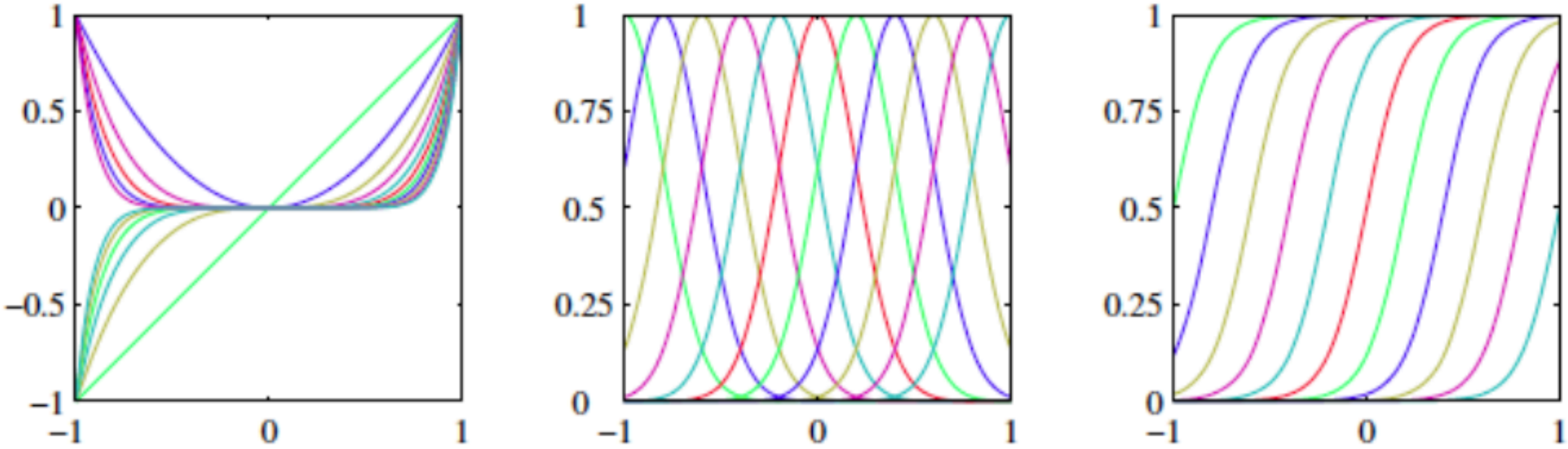
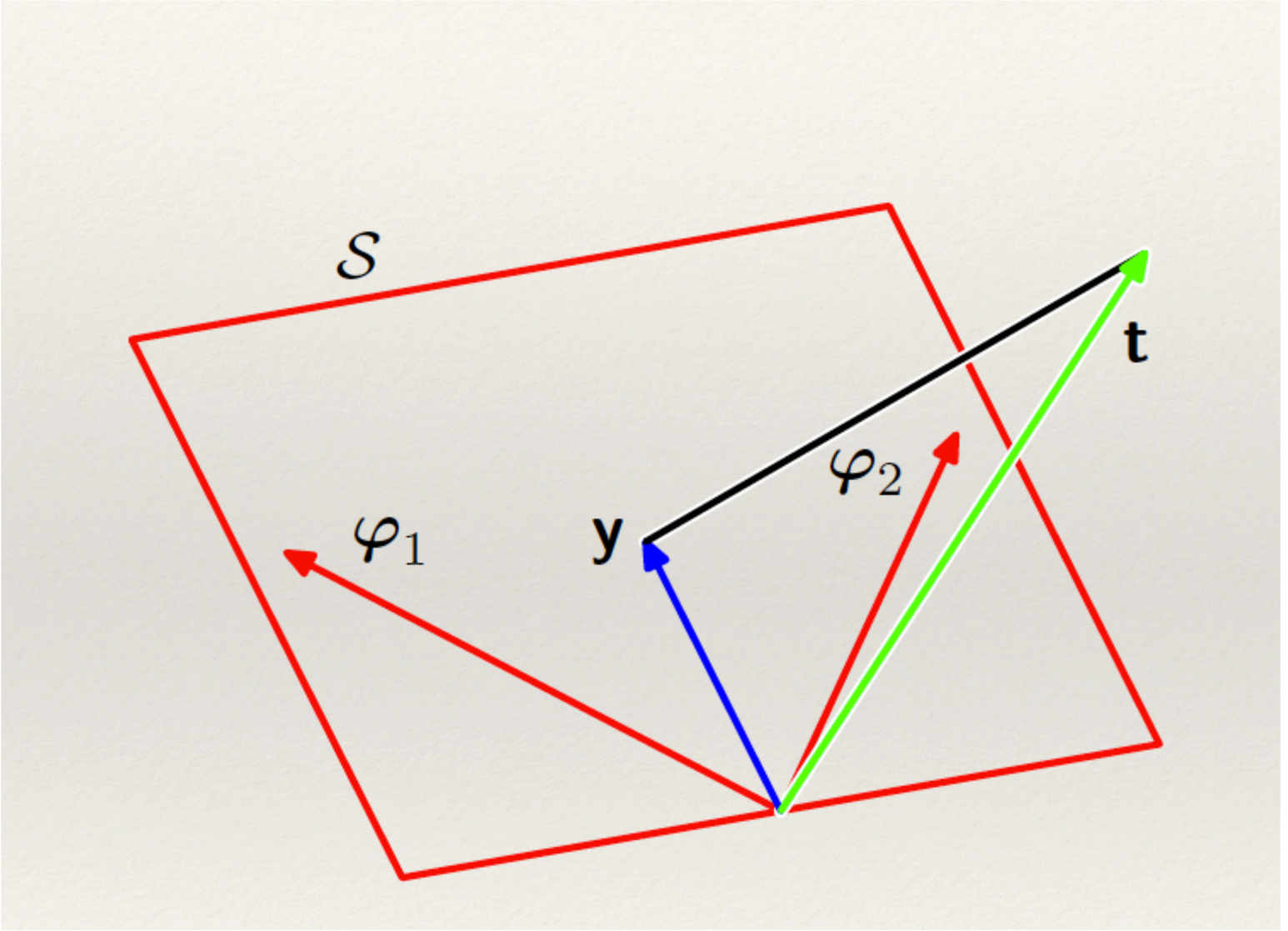


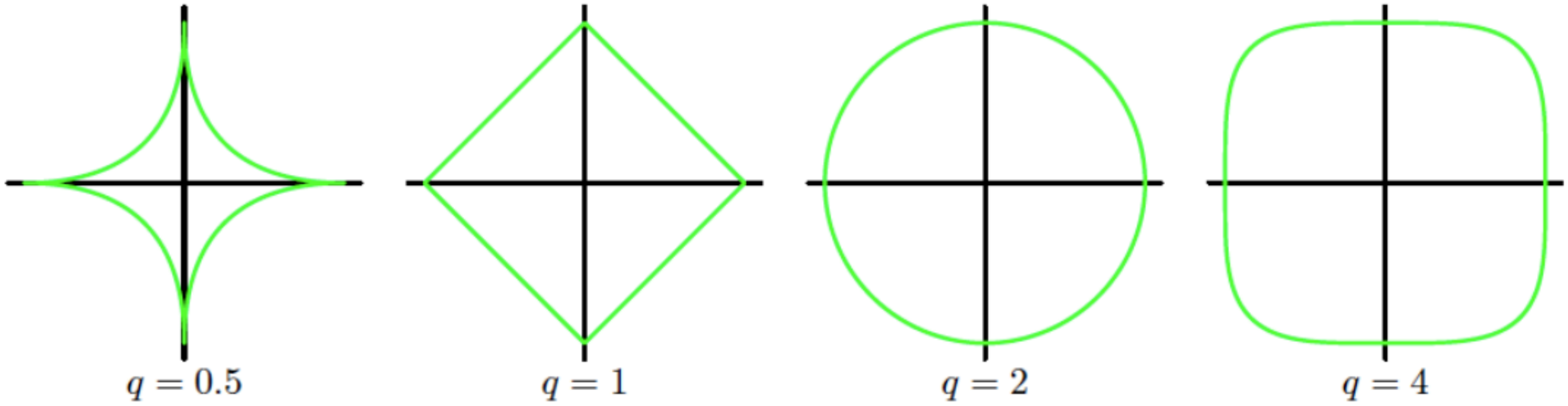
Figure 3.1 Examples of basis functions, showing polynomials on the left, Gaussians of the form (3.4) in the centre, and sigmoidal of the form (3.5) on the right.

LINEAR REGRESSION - GEOMETRY

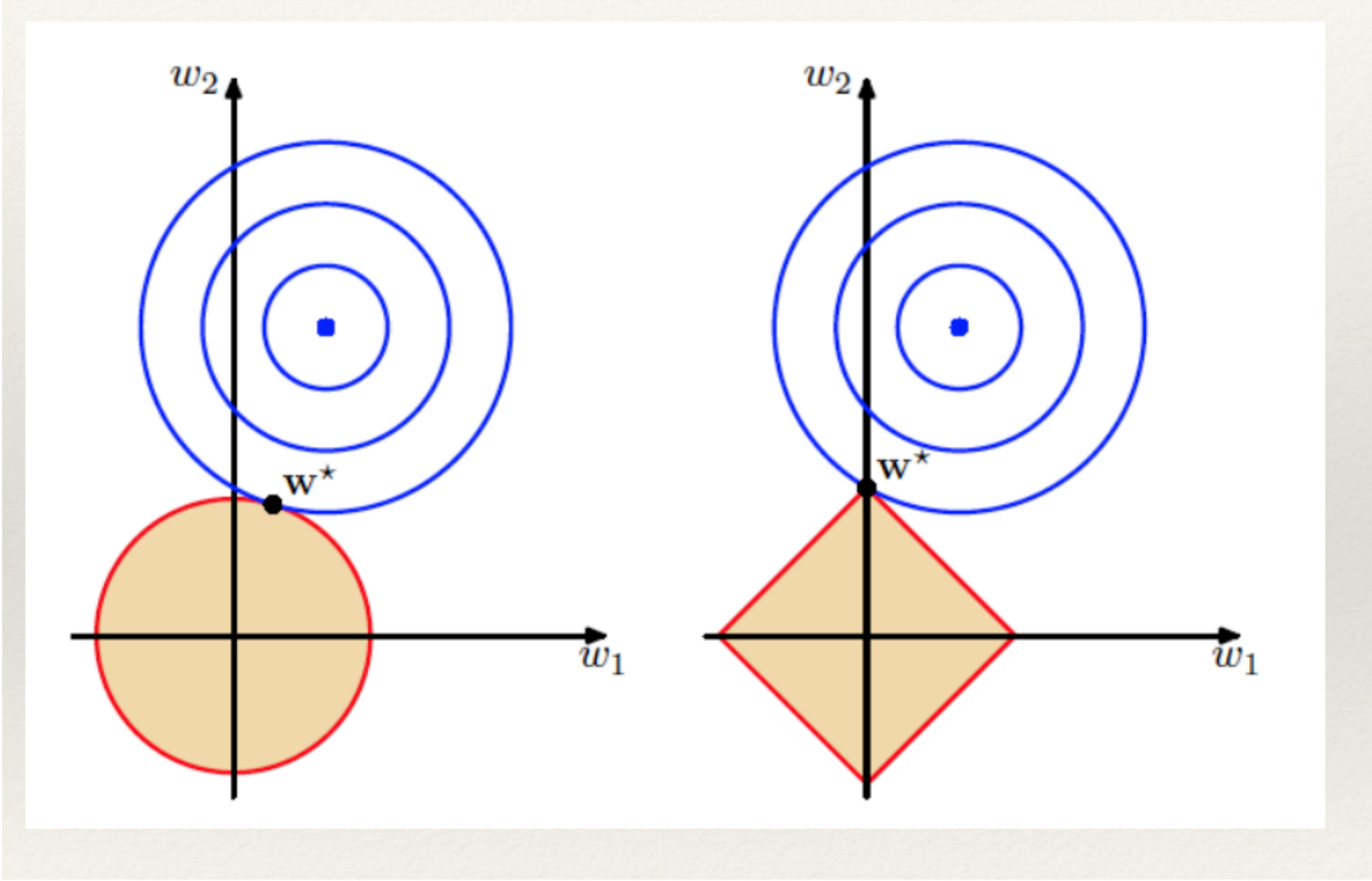


REGULARIZED LINEAR REGRESSION

- Optimize a modified cost function $E_D(\mathbf{w}) + \lambda E_W(\mathbf{w})$



REGULARIZED LINEAR REGRESSION



THANK YOU

Sriram Ganapathy and TA team
LEAP lab, C328, EE, IISc
sriramg@iisc.ac.in