Hard EM for Mixture of Gaussians

1. Guess initial parameters p_k, μ_k, Σ_k for each class k
2. Repeat until convergence:
   (a) *E-step:* For each instance i, compute the most likely class:

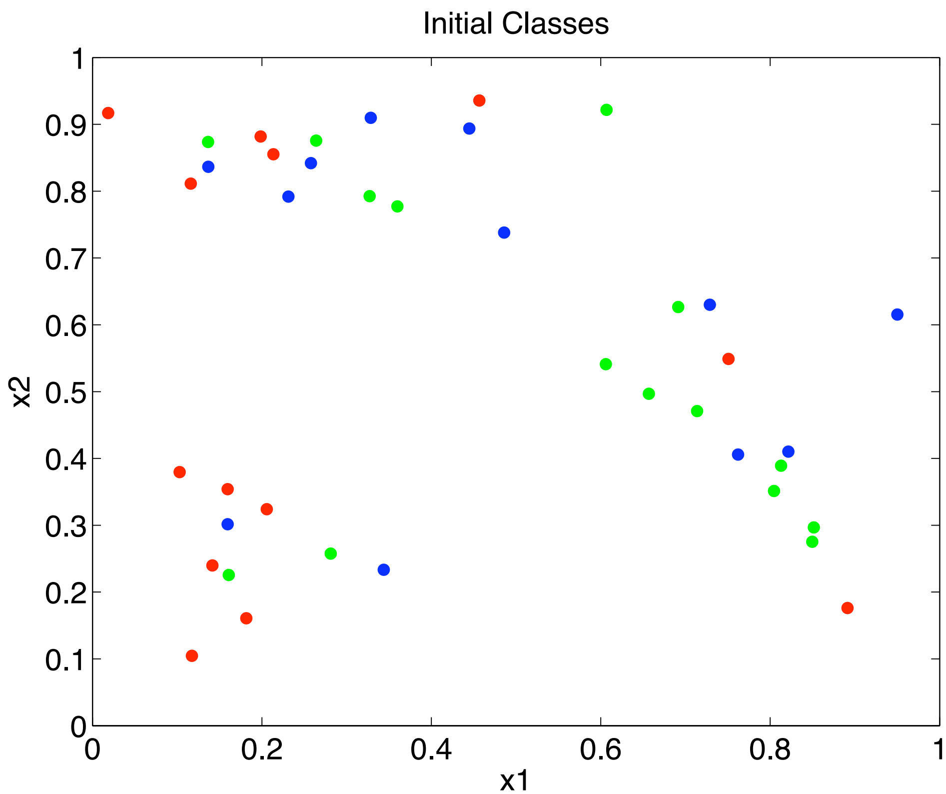
$$y_i = \arg\max_k P(y=k|\mathbf{x}_i) = \arg\max_k P(\mathbf{x}_i|y=k)P(y=k)$$

   (b) *M-step:* Update the parameters of the model $(p_k, \mu_k, \Sigma_k, \forall k = 1 \ldots K)$ to maximize the likelihood of the data:

$$p_k = \frac{1}{m}\sum_{i=1}^{m}\delta_{ik} \qquad \mu_k = \frac{\sum_{i=1}^{m}\delta_{ik}\mathbf{x}_i}{\sum_{i=1}^{m}\delta_{ik}}$$

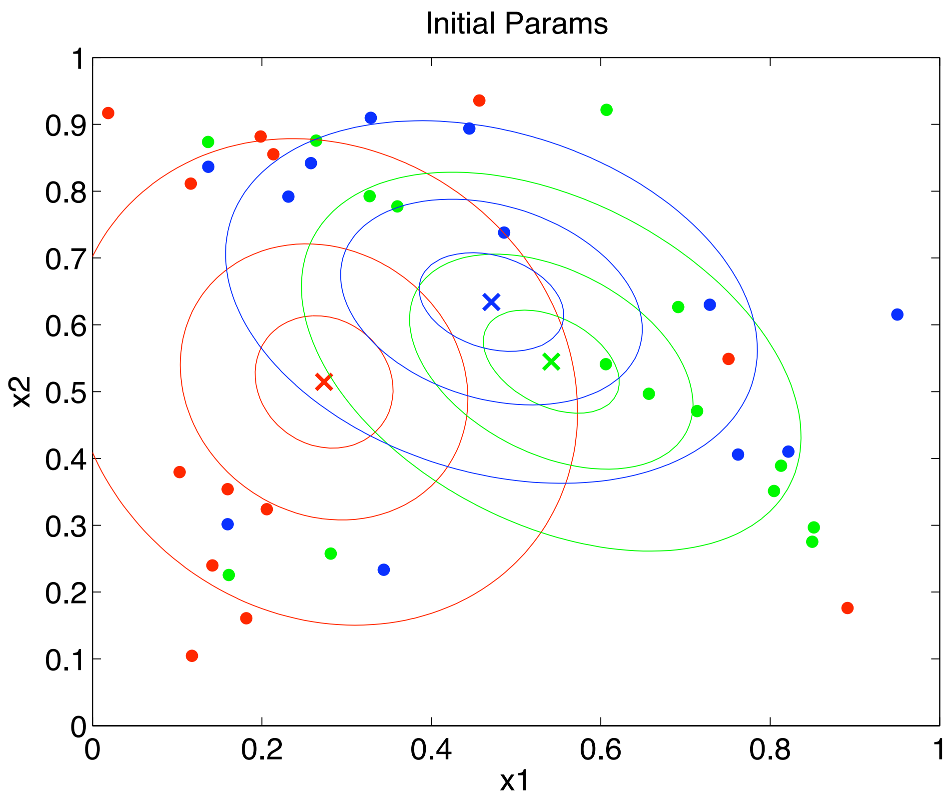
$$\Sigma_k = \frac{\sum_{i=1}^{m}\delta_{ik}\left(\mathbf{x}_i-\mu_k\right)\left(\mathbf{x}_i-\mu_k\right)^T}{\sum_{i=1}^{m}\delta_{ik}}$$

Hard EM for Mixture of Gaussians: Example



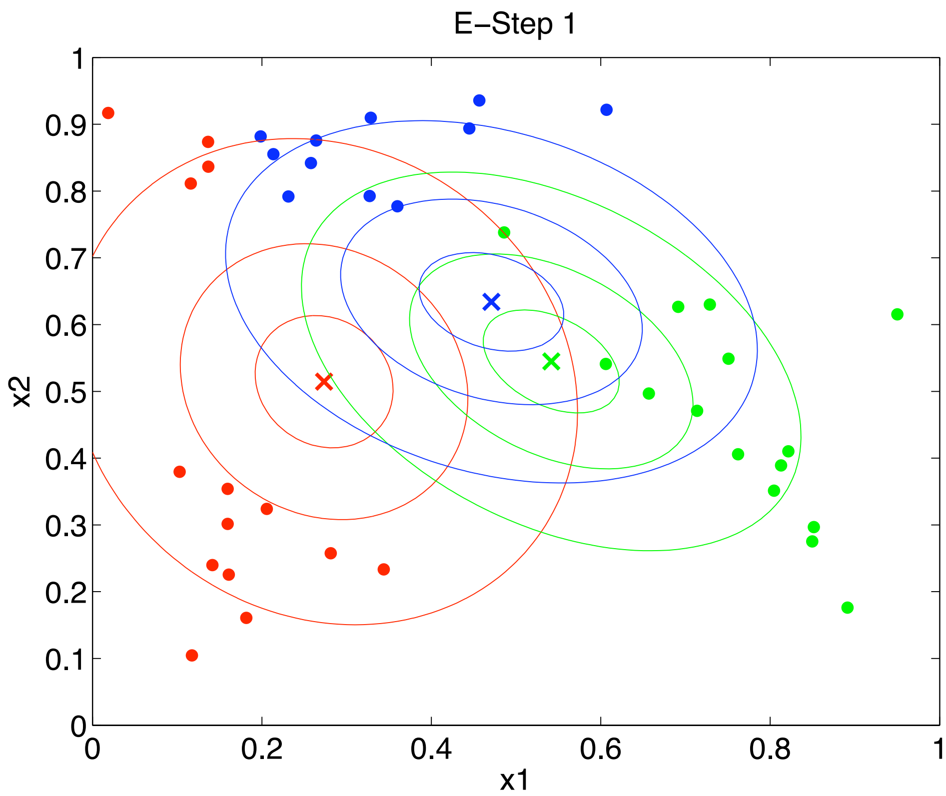
$K = 3$, initial assignment of points to components is random

Hard EM for Mixture of Gaussians: Example

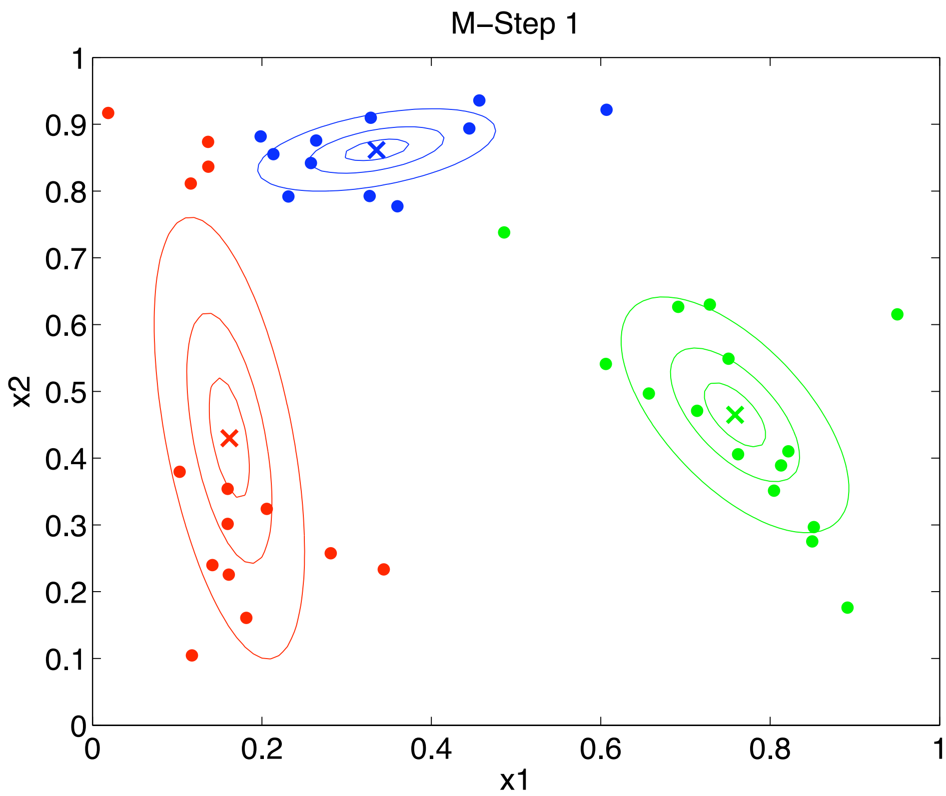


Initial parameters (means and variances) computed from initial assignments

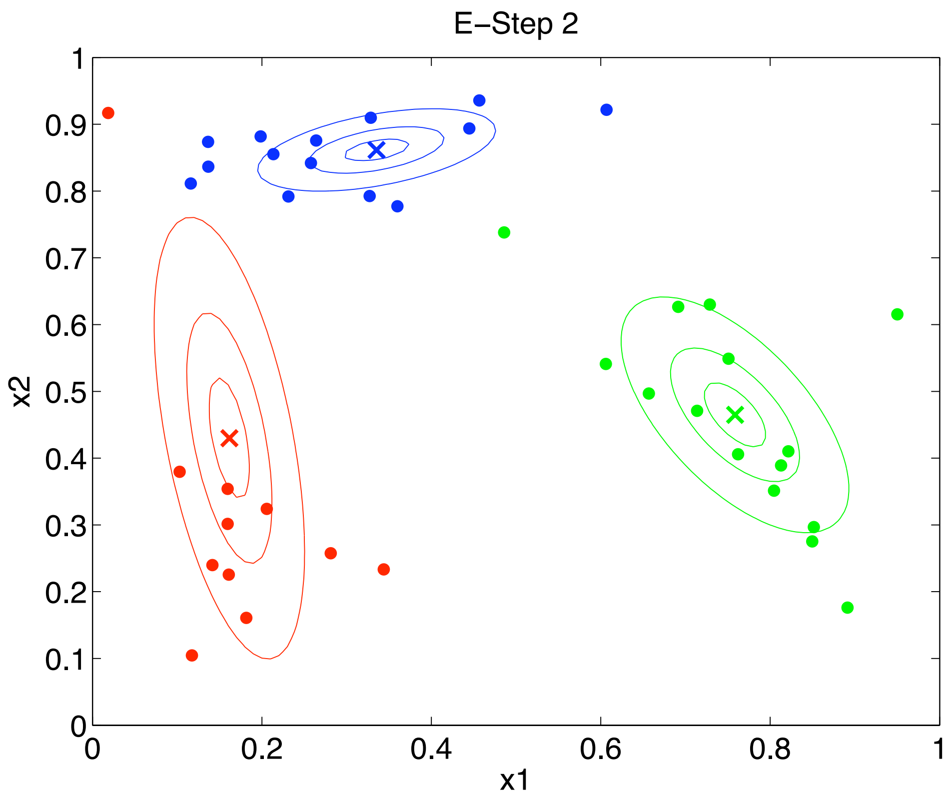
Hard EM for Mixture of Gaussians: Example



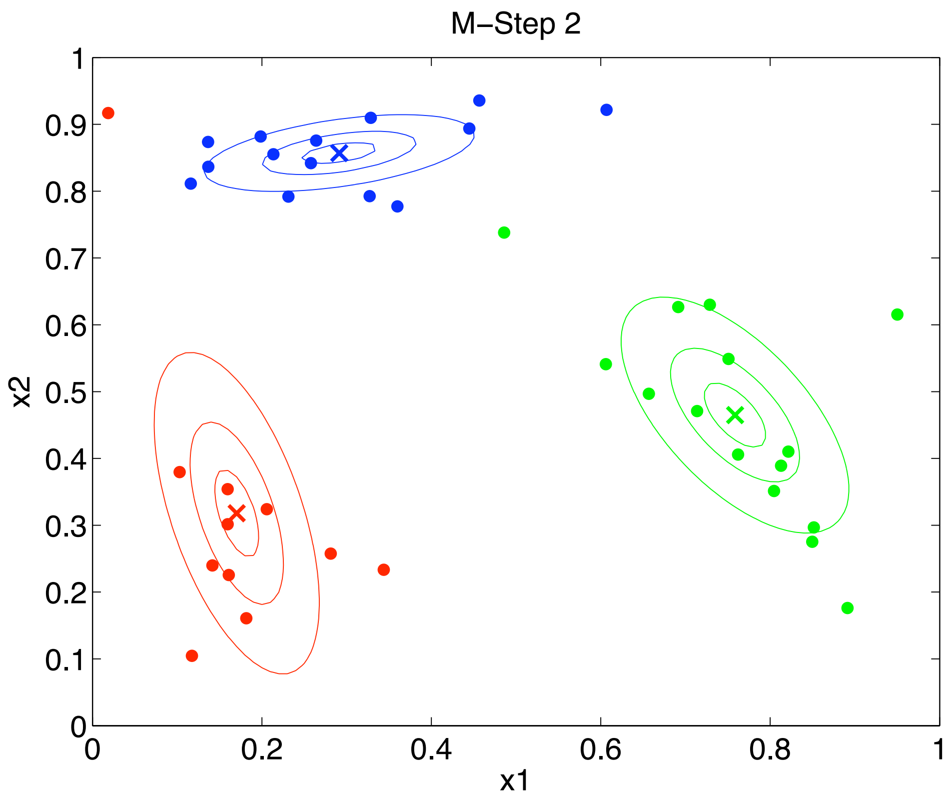
Hard EM for Mixture of Gaussians: Example



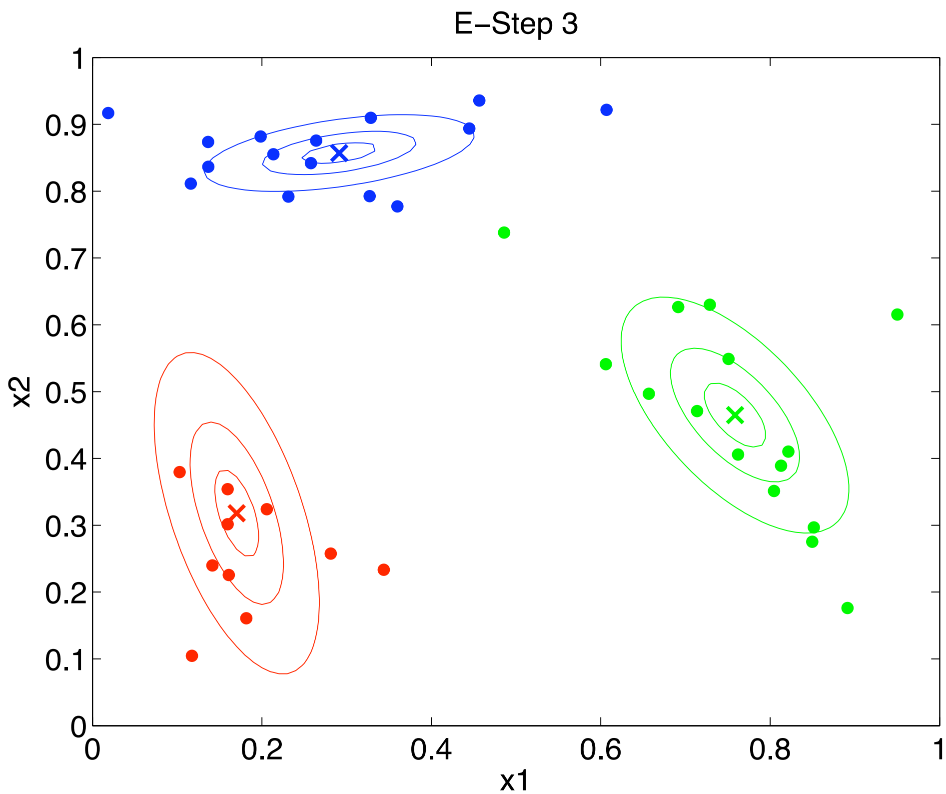
Hard EM for Mixture of Gaussians: Example



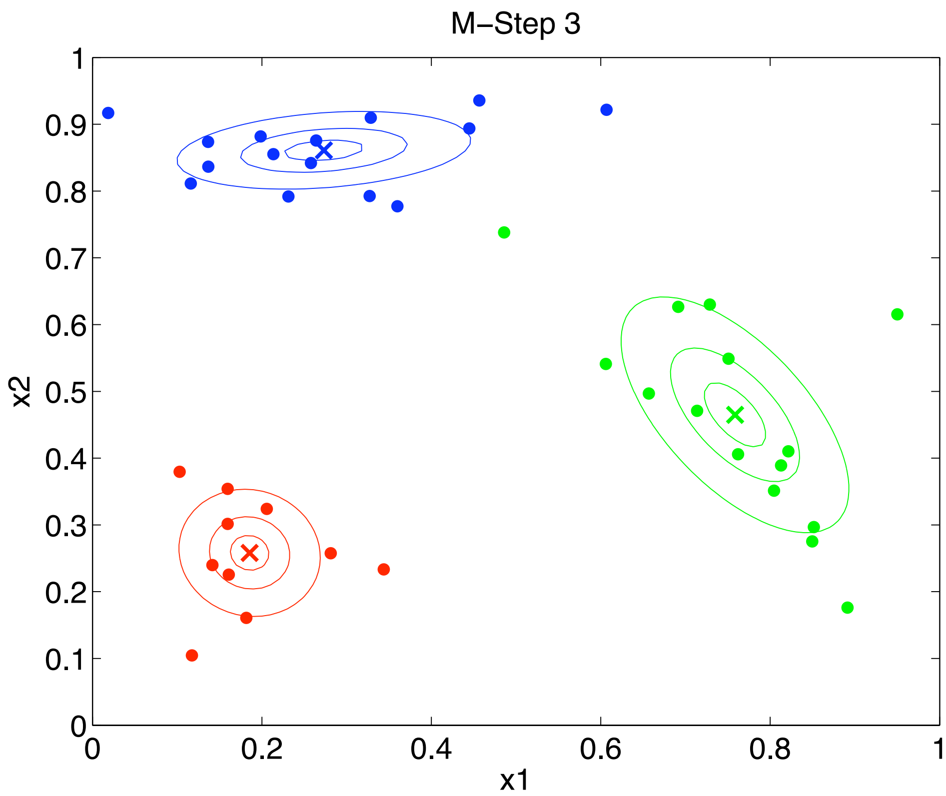
Hard EM for Mixture of Gaussians: Example



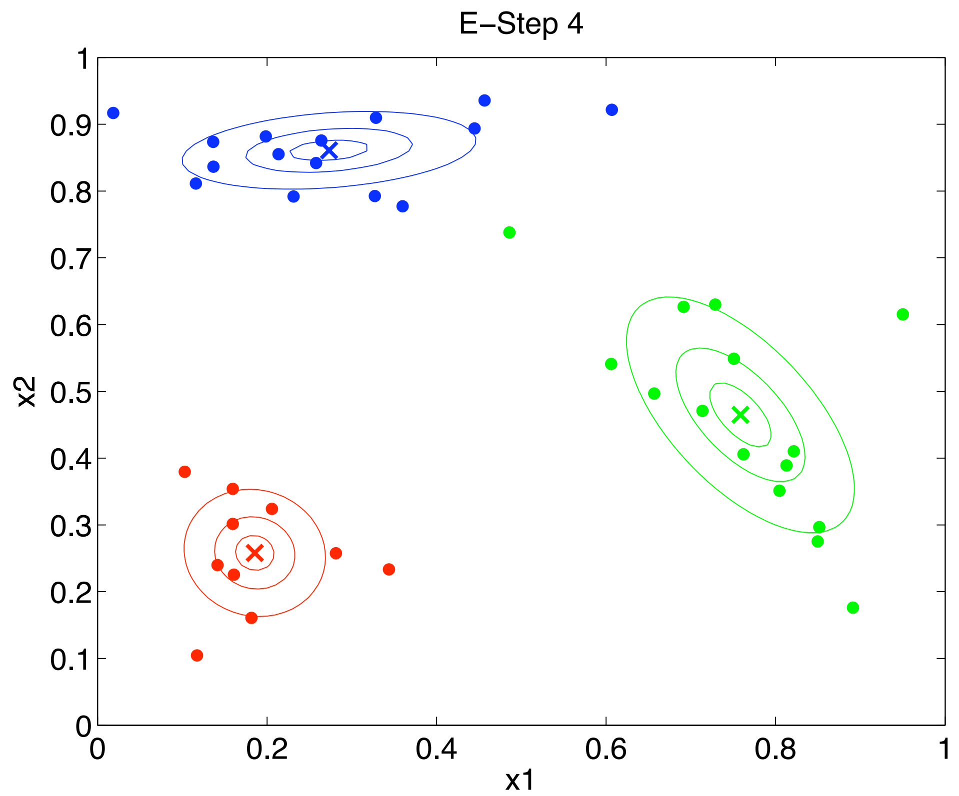
Hard EM for Mixture of Gaussians: Example



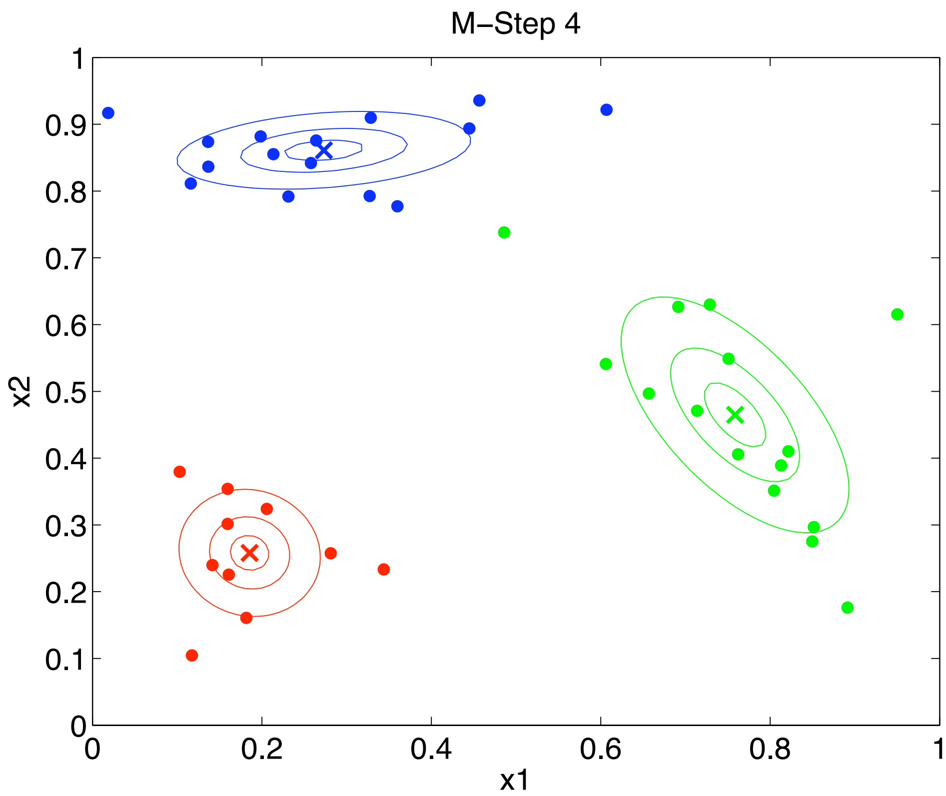
Hard EM for Mixture of Gaussians: Example



Hard EM for Mixture of Gaussians: Example



Hard EM for Mixture of Gaussians: Example



Soft EM for Mixture of Gaussians

1. Guess initial parameters p_k, μ_k, Σ_k for each class k
2. Repeat until convergence:
   (a) *E-step:* For each instance i and class k, compute the probabilities of class membership:

$$w_{ik} = P(y_i = k|\mathbf{x}_i) = \frac{P(\mathbf{x}_i|y_i = k)P(y_i = k)}{P(\mathbf{x}_i)}$$

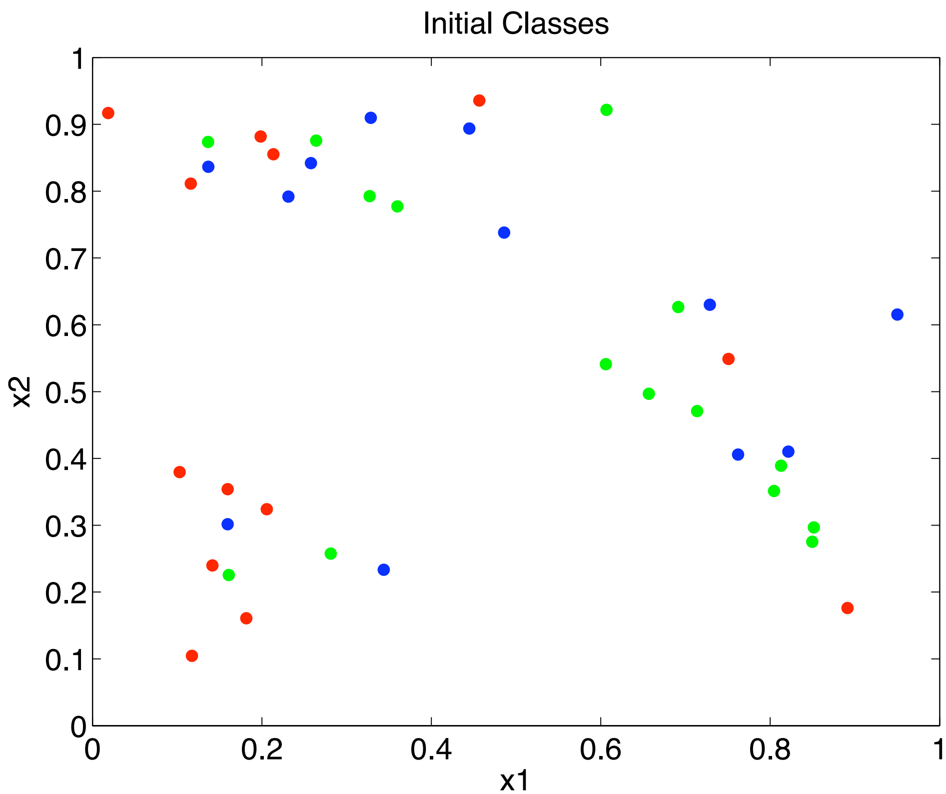
   I.e., instances are "partially assigned" to each class, according to w_{ik}

   (b) *M-step:* Update the parameters of the model to maximize the likelihood of the data:

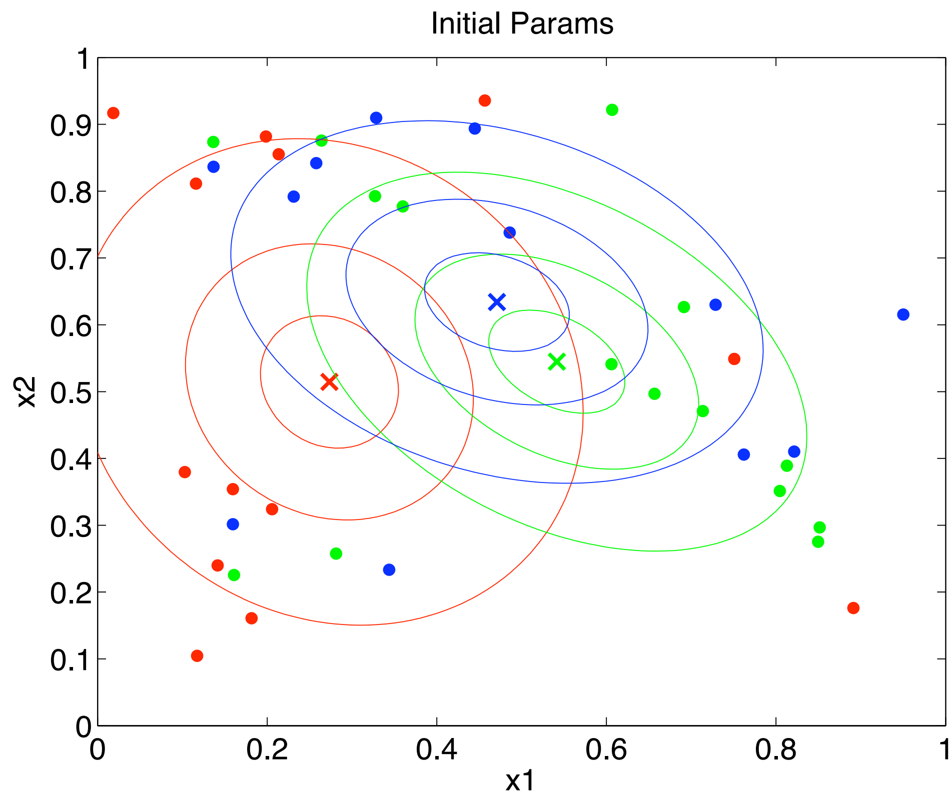
$$p_k = \frac{1}{m}\sum_{i=1}^{m} w_{ik} \qquad \mu_k = \frac{\sum_{i=1}^{m} w_{ik}\mathbf{x}_i}{\sum_{i=1}^{m} w_{ik}}$$

$$\Sigma_k = \frac{\sum_{i=1}^{m} w_{ik}\left(\mathbf{x}_i - \mu_k\right)\left(\mathbf{x}_i - \mu_k\right)^T}{\sum_{i=1}^{m} w_{ik}}$$

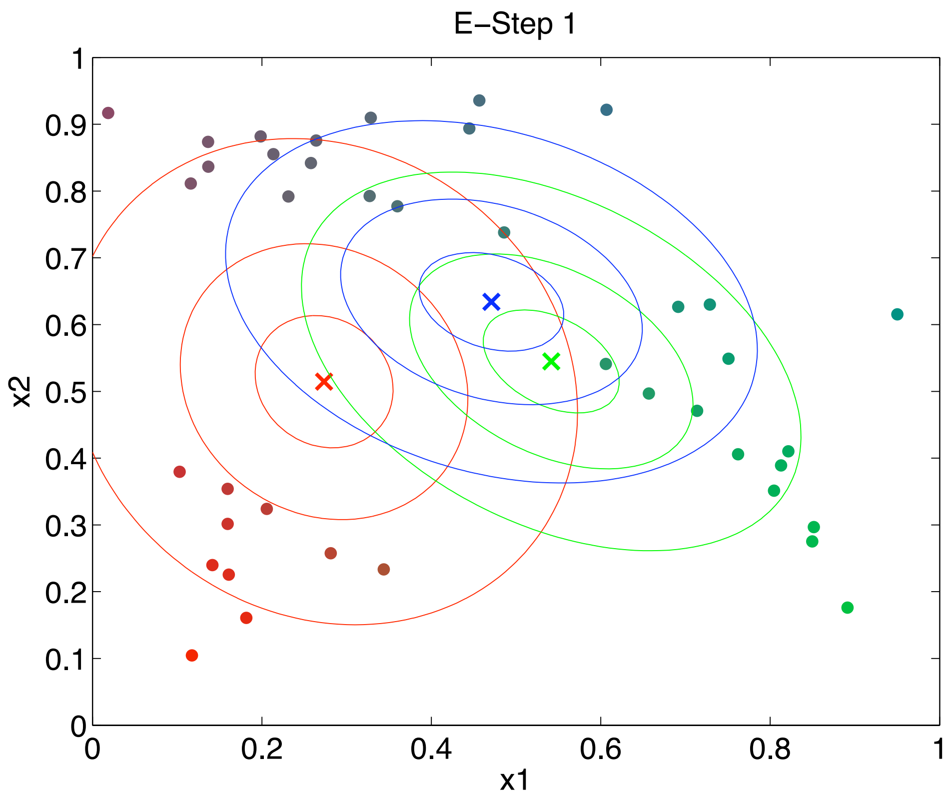
Soft EM for Mixture of Gaussians: Example



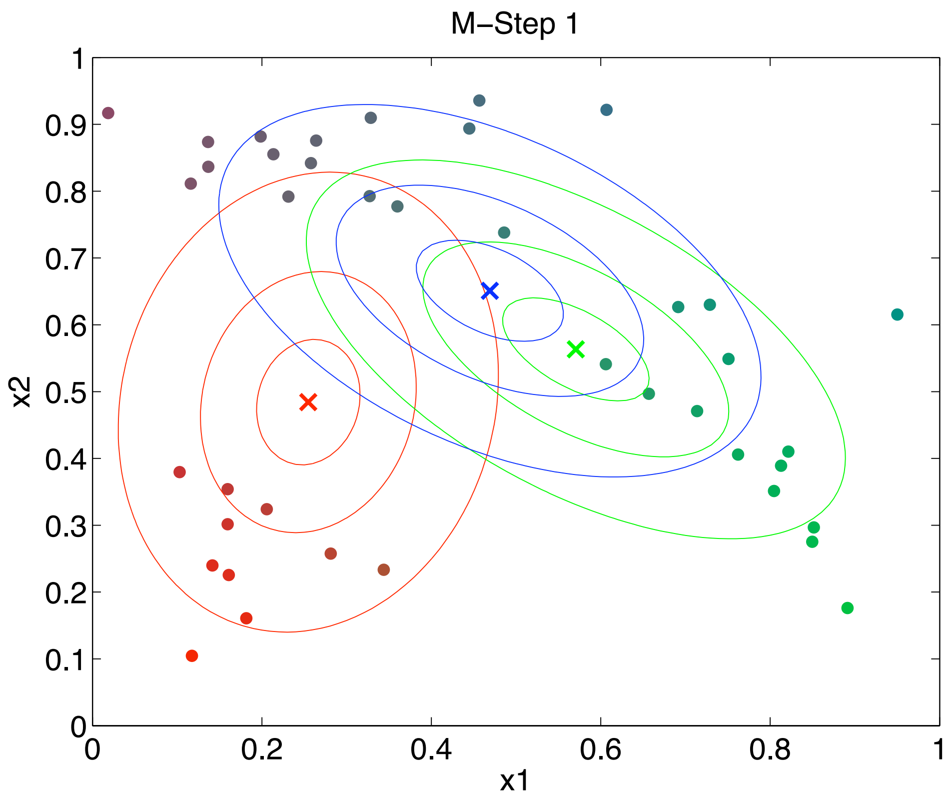
Soft EM for Mixture of Gaussians: Example



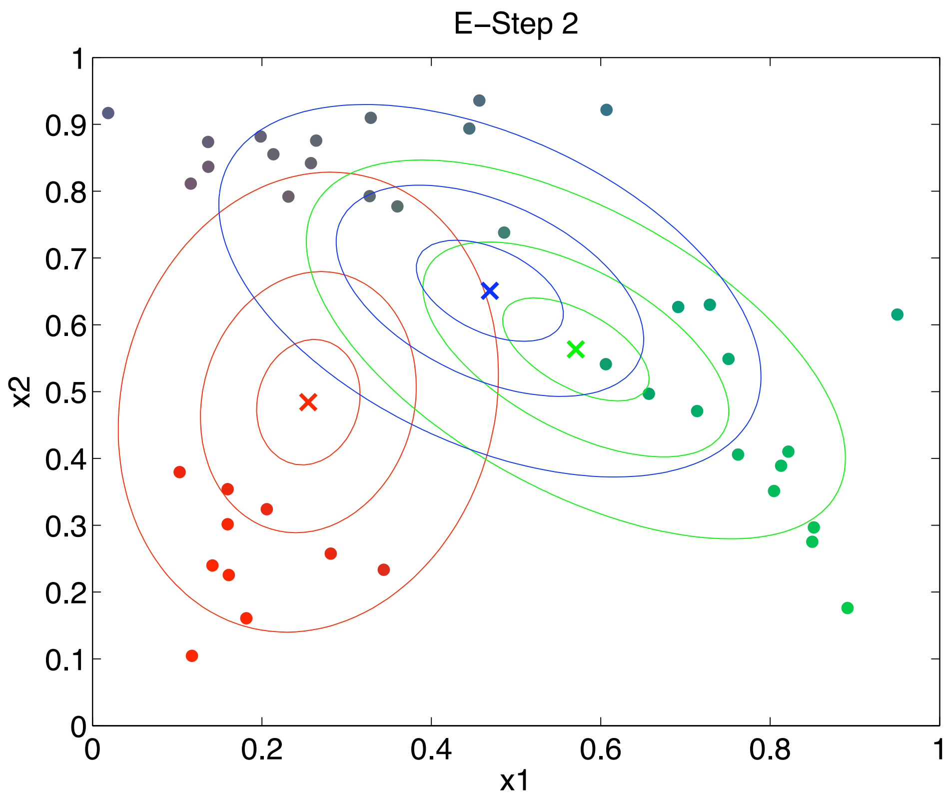
Soft EM for Mixture of Gaussians: Example



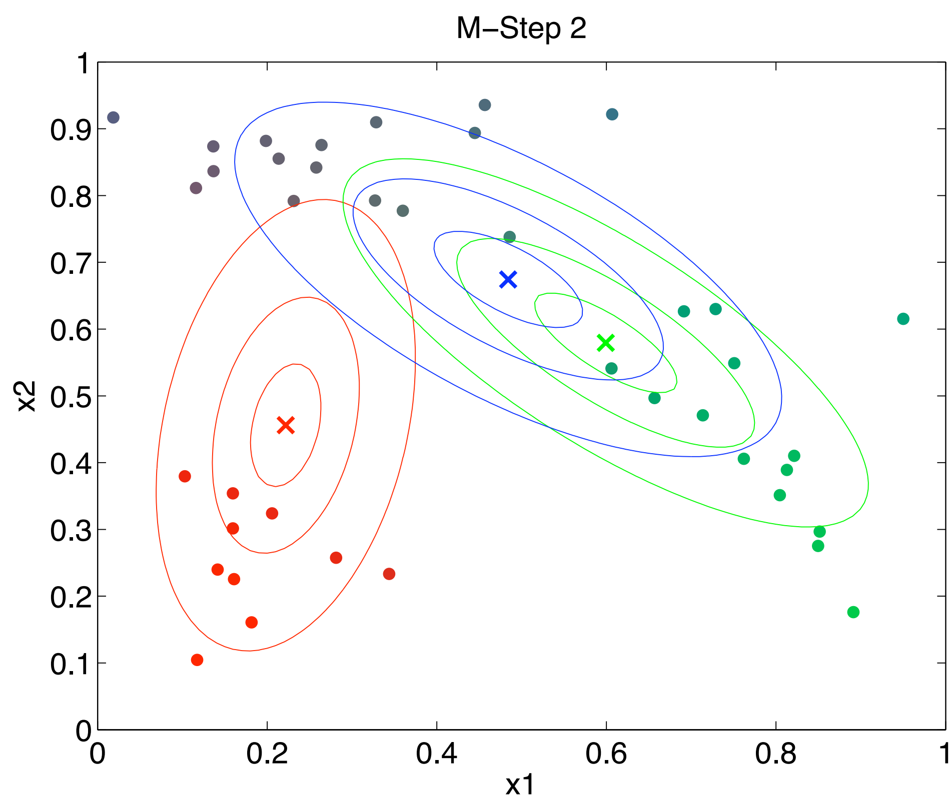
Soft EM for Mixture of Gaussians: Example



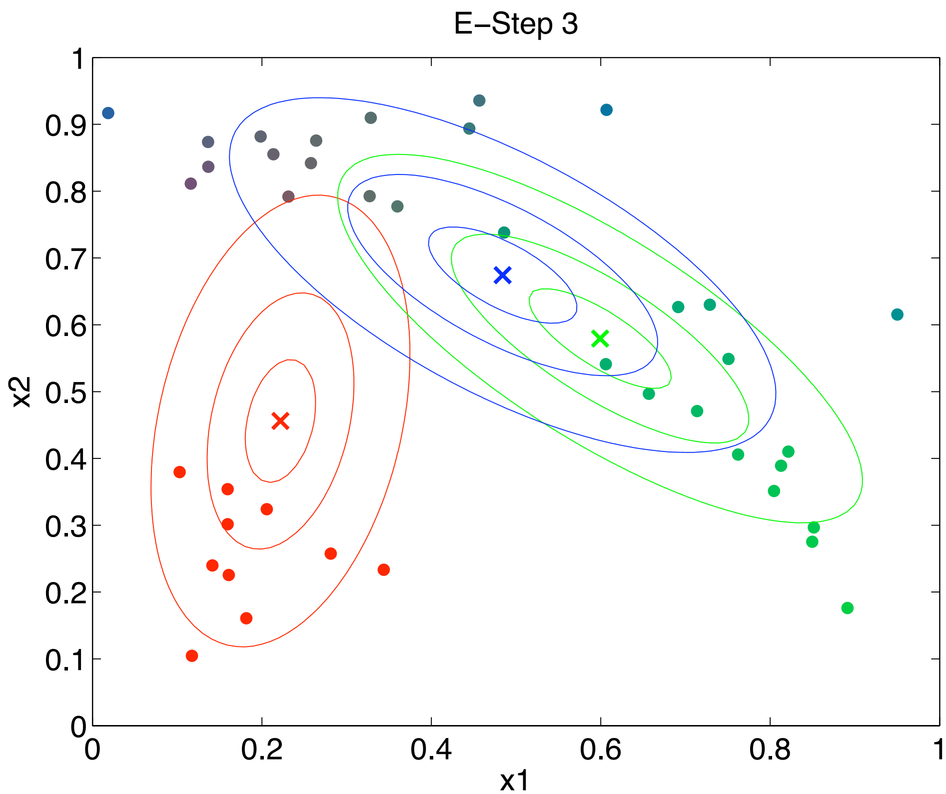
Soft EM for Mixture of Gaussians: Example



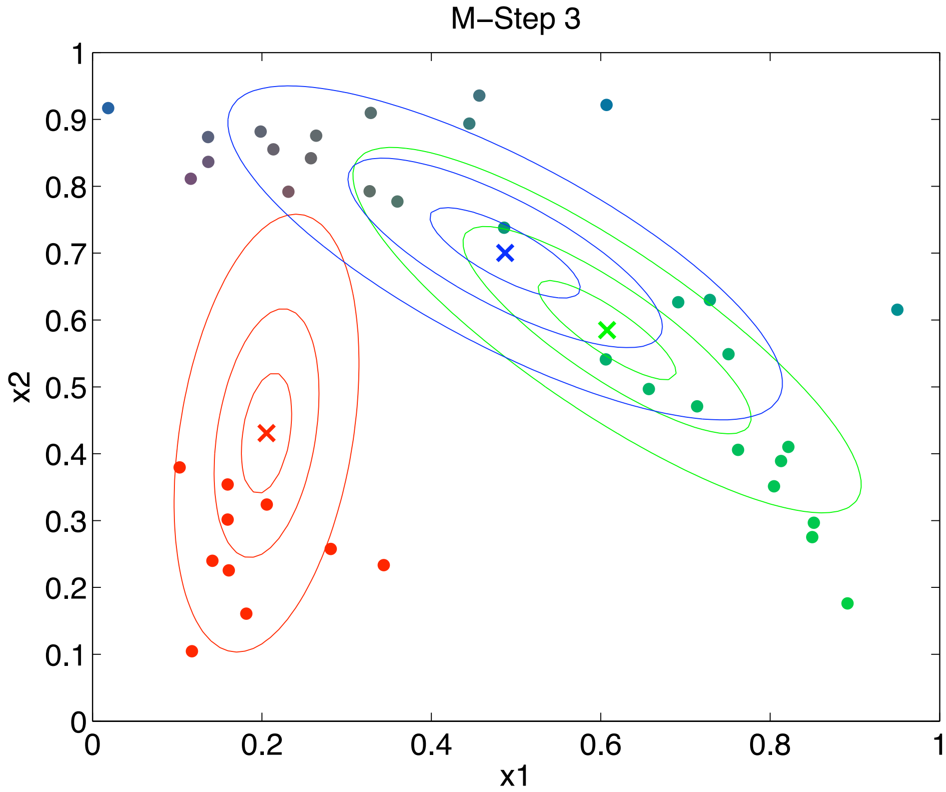
Soft EM for Mixture of Gaussians: Example



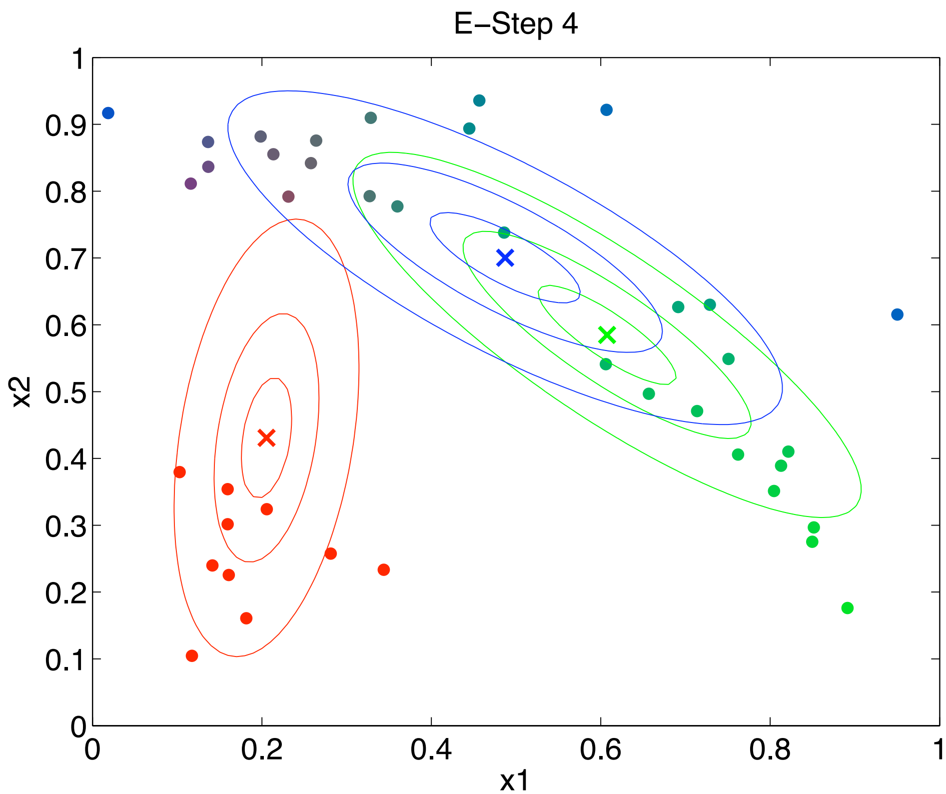
Soft EM for Mixture of Gaussians: Example



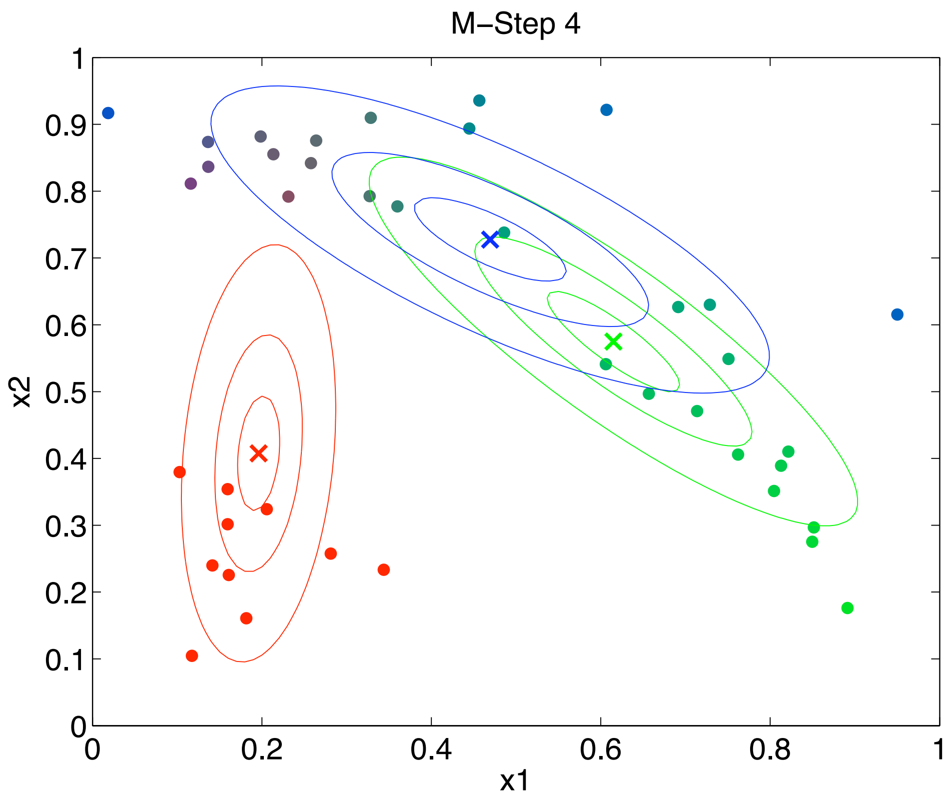
Soft EM for Mixture of Gaussians: Example



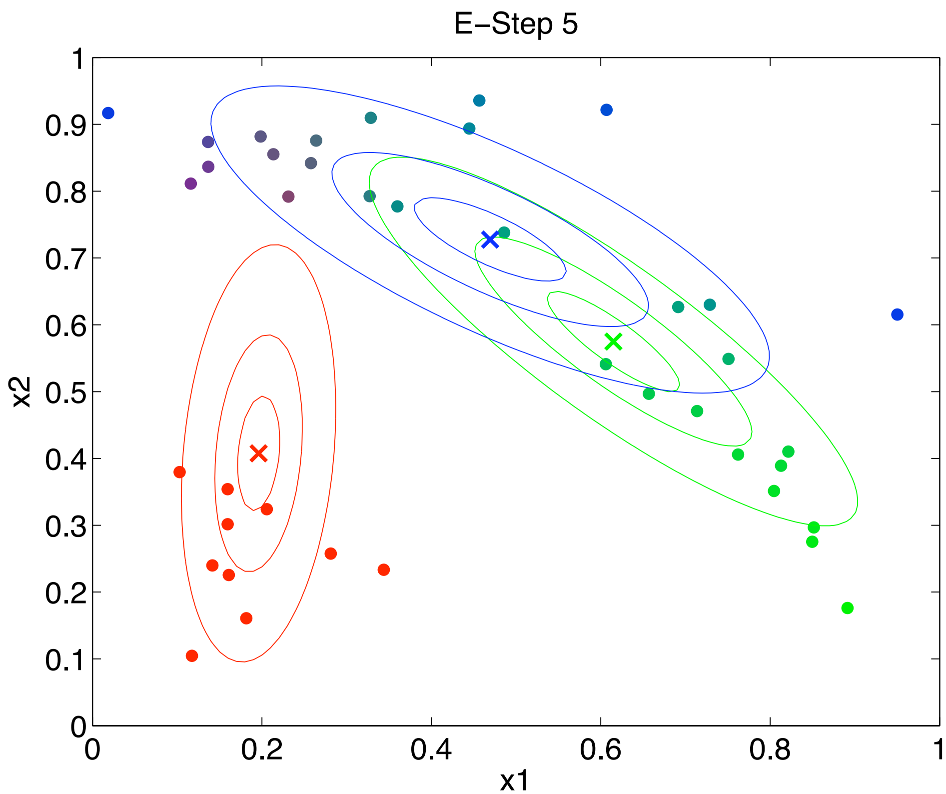
Soft EM for Mixture of Gaussians: Example



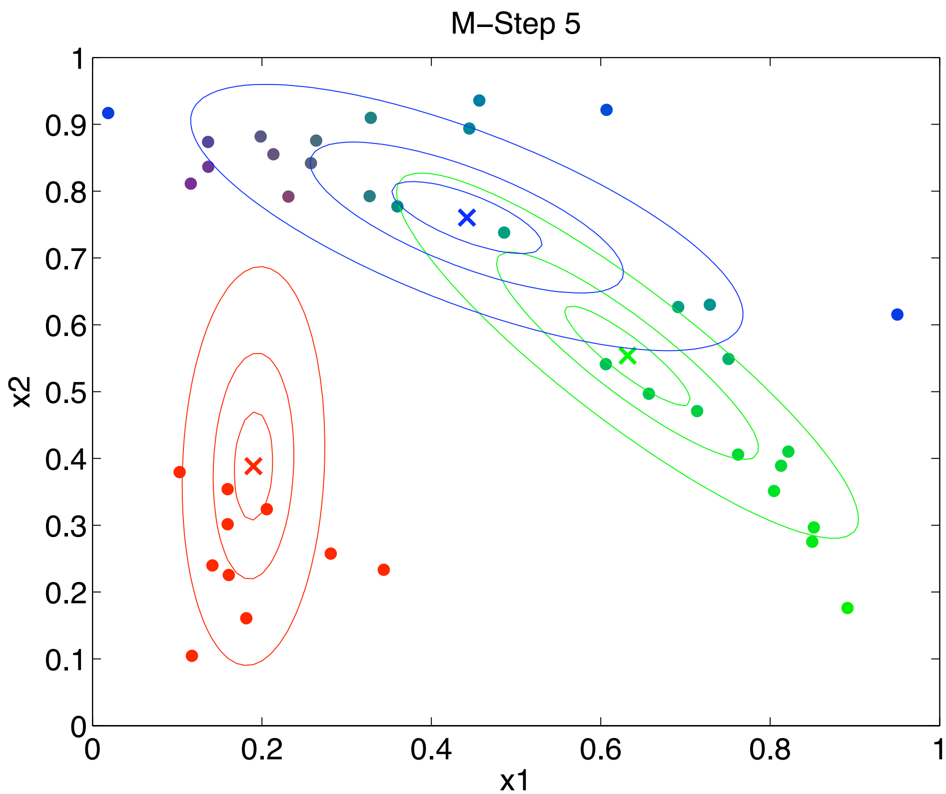
Soft EM for Mixture of Gaussians: Example



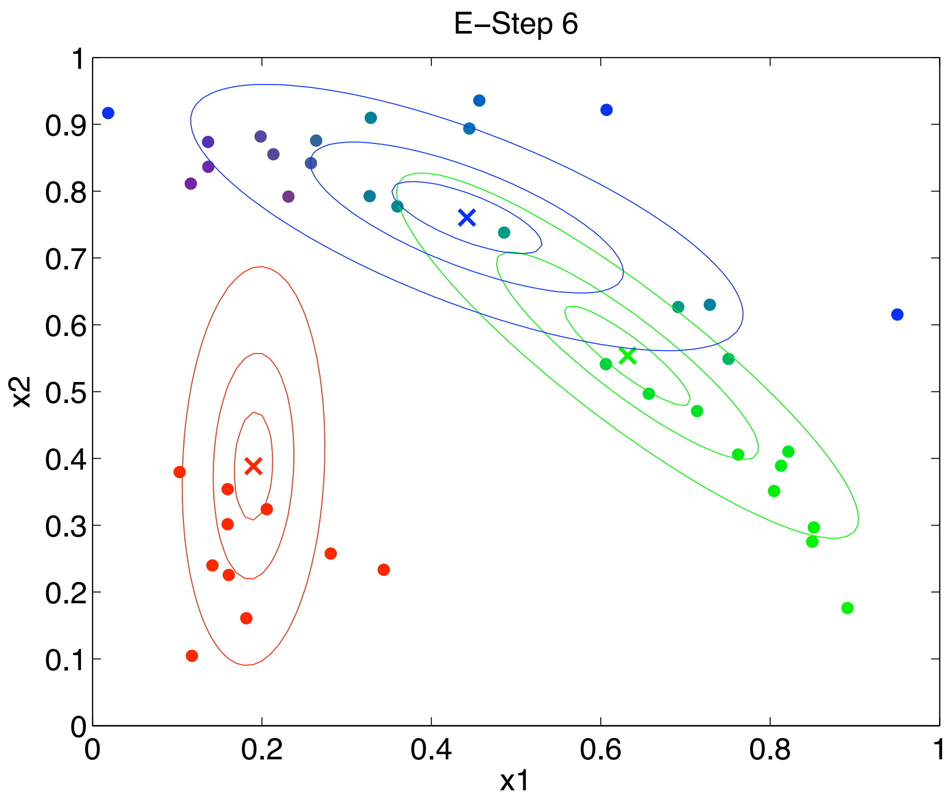
Soft EM for Mixture of Gaussians: Example



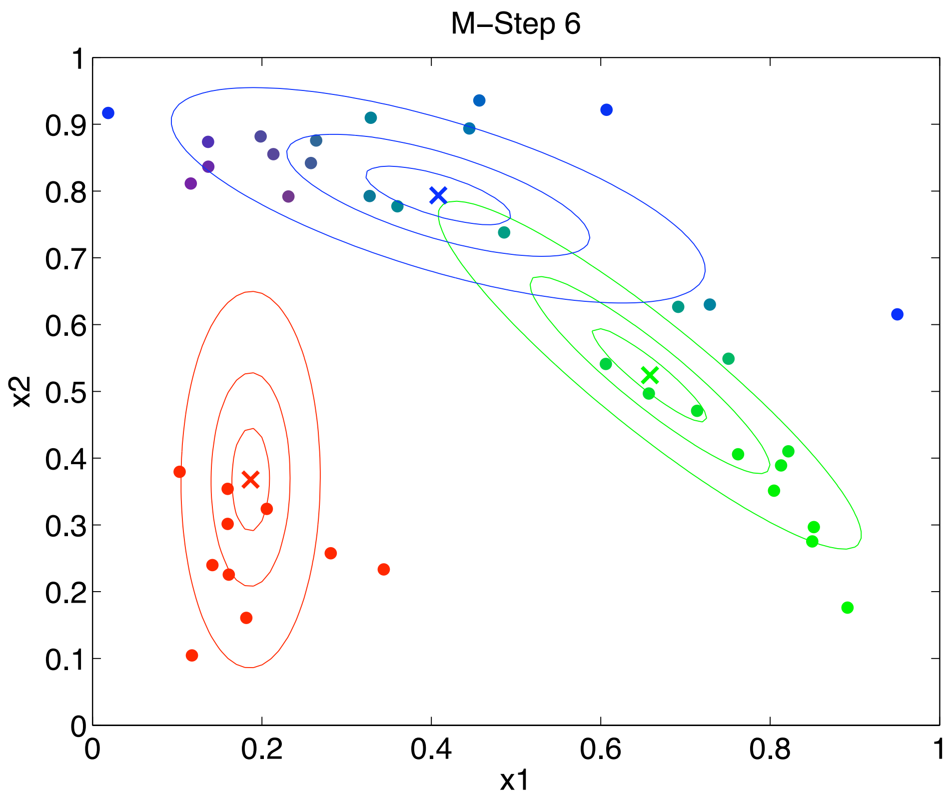
Soft EM for Mixture of Gaussians: Example



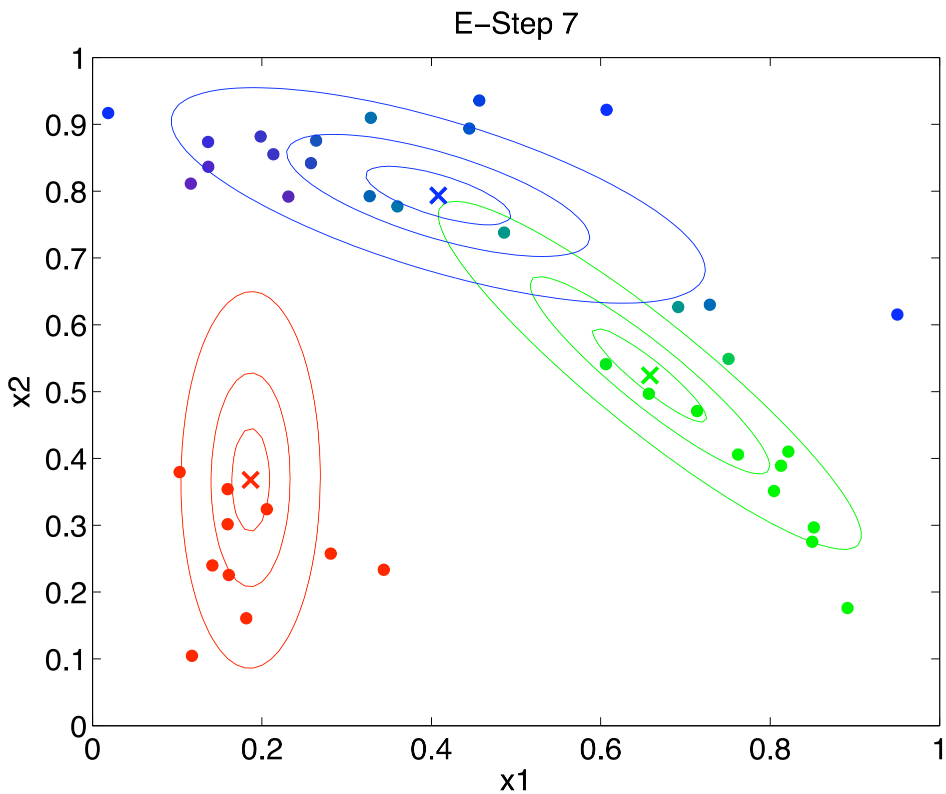
Soft EM for Mixture of Gaussians: Example



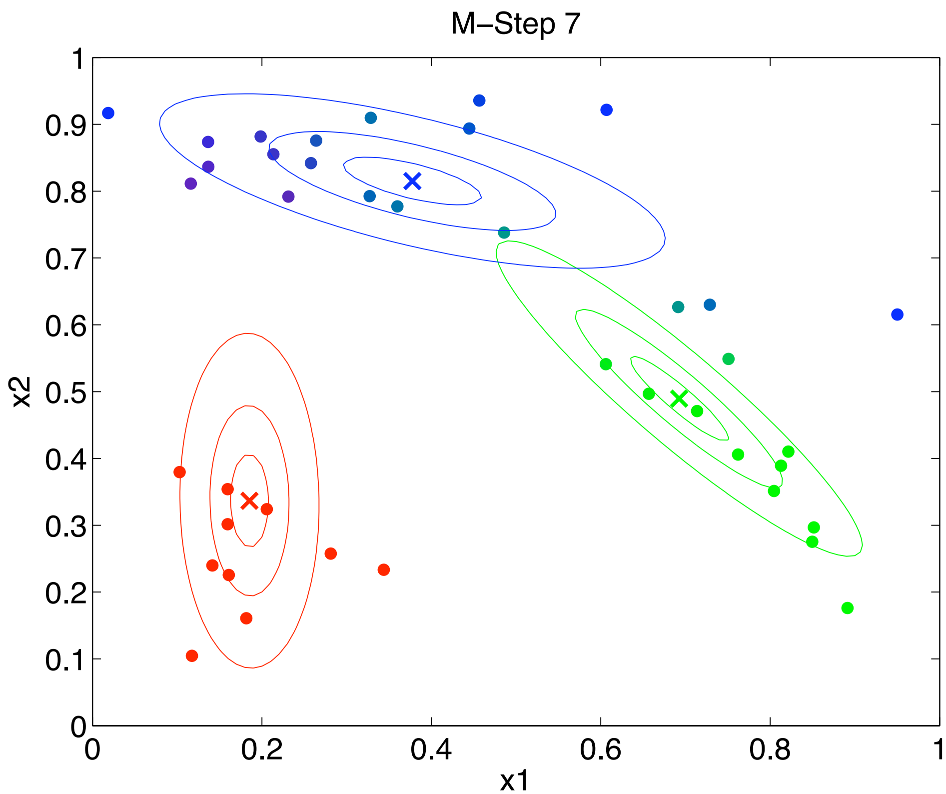
Soft EM for Mixture of Gaussians: Example



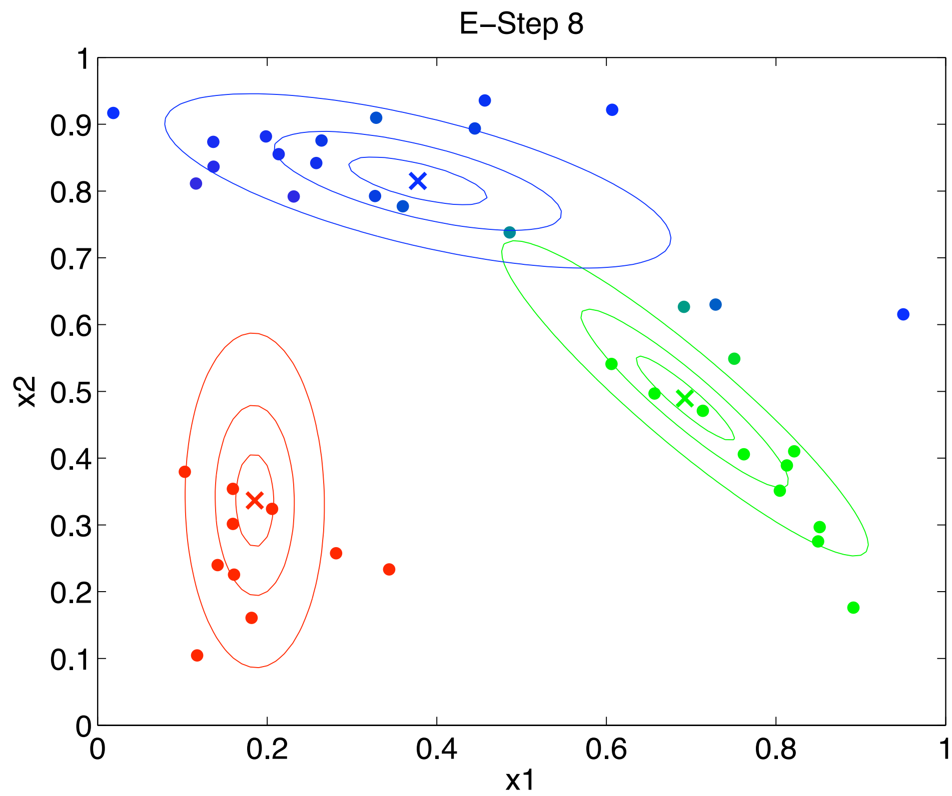
Soft EM for Mixture of Gaussians: Example



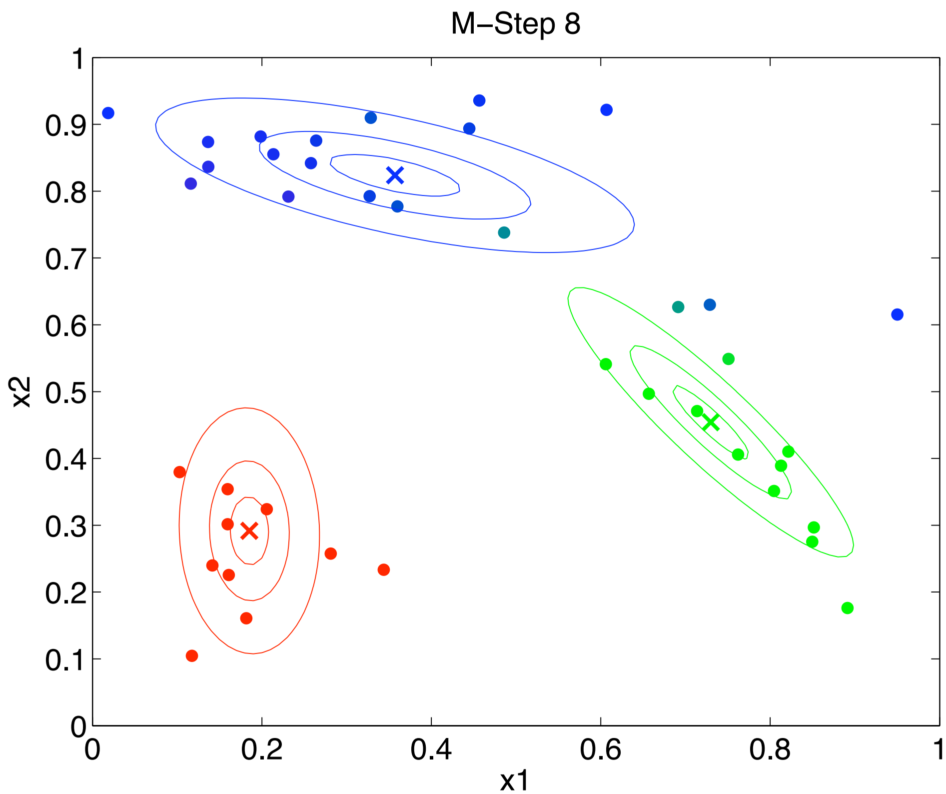
Soft EM for Mixture of Gaussians: Example



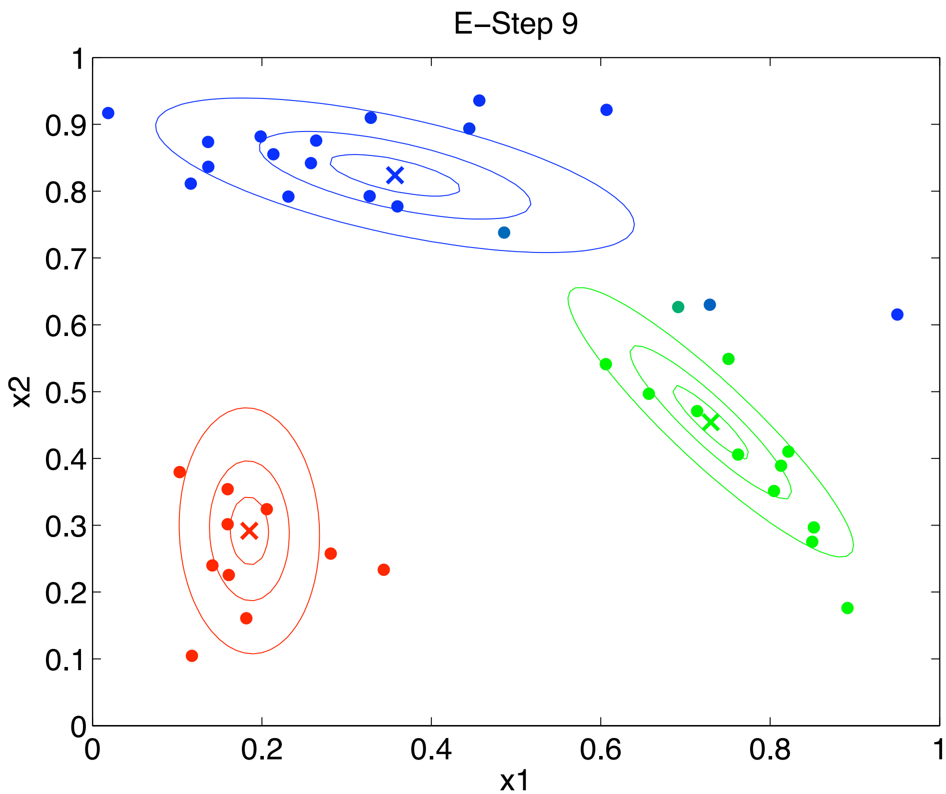
Soft EM for Mixture of Gaussians: Example



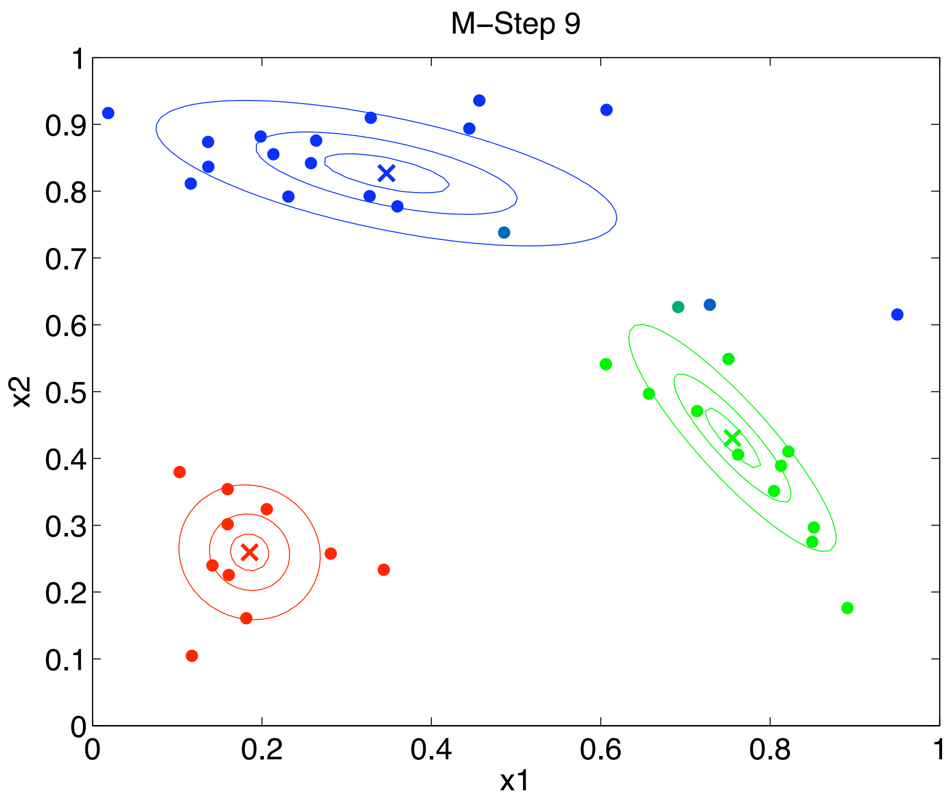
Soft EM for Mixture of Gaussians: Example



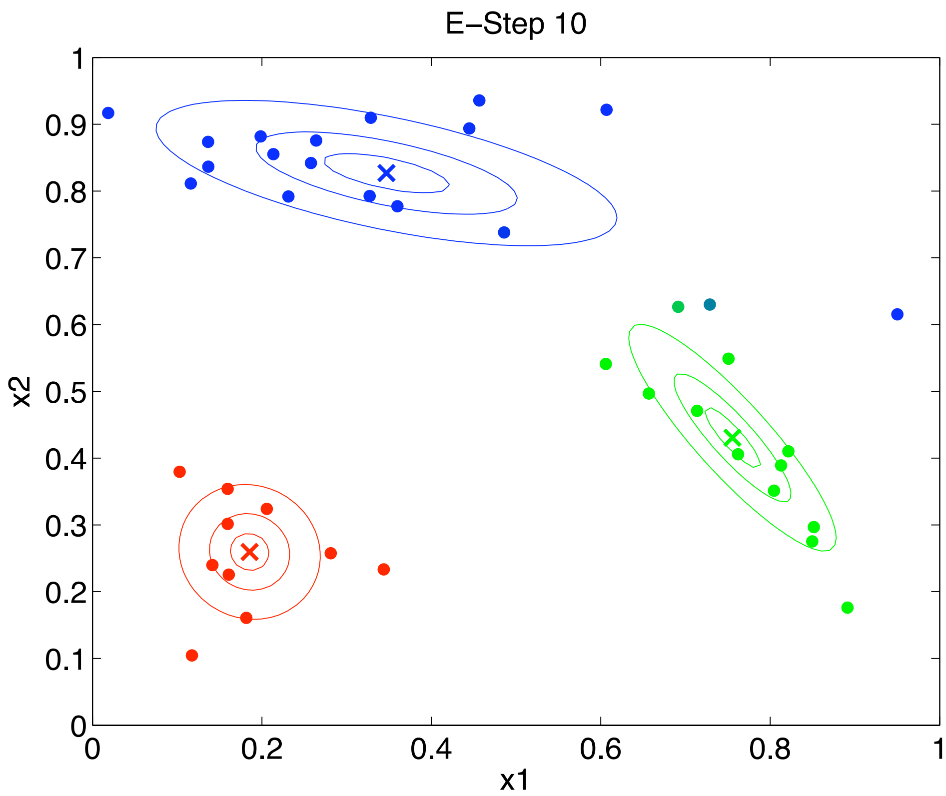
Soft EM for Mixture of Gaussians: Example



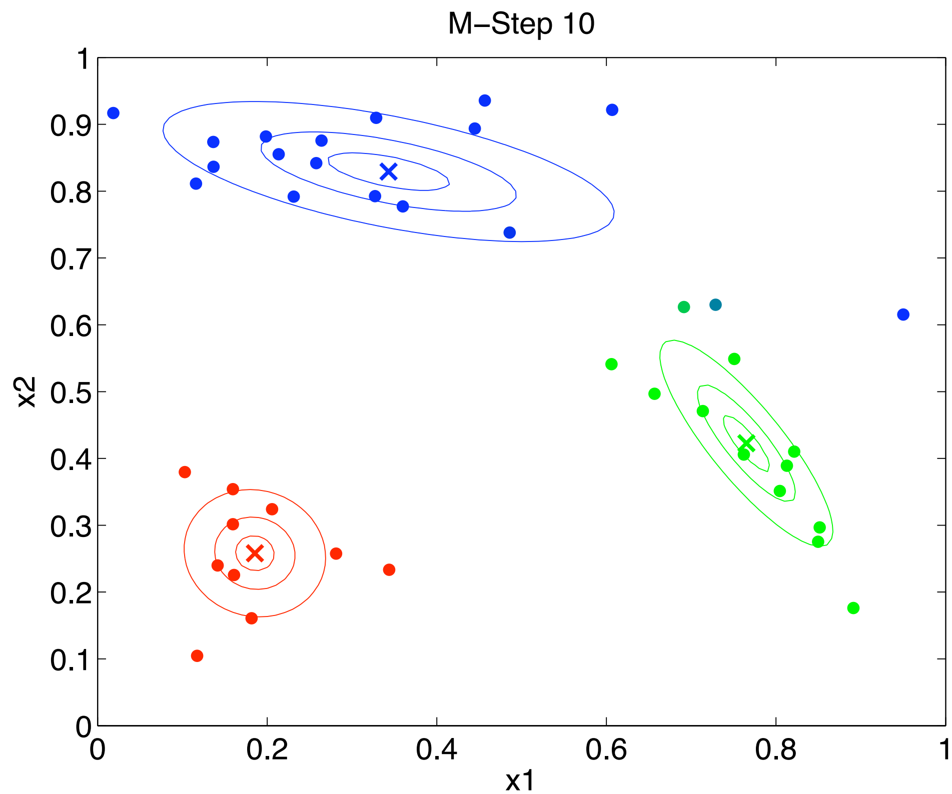
Soft EM for Mixture of Gaussians: Example



Soft EM for Mixture of Gaussians: Example



Soft EM for Mixture of Gaussians: Example



Note that some points still have “mixed” colour, indicating that they belong with significant probability to more than one component/cluster.

Comparison of hard EM and soft EM

- Soft EM does not commit to a particular value of the missing item. Instead, it considers all possible values, with some probability
- This is a pleasing property, given the uncertainty in the value
- Soft EM is almost always the method of choice (and often when people say "EM", they mean the soft version)
- The complexity of each iteration of the two versions is pretty much the same.
- Soft EM might take more iterations, if we stop it based on having a change in the parameter values below a threshold