E9:261

24-02-2016

Recap ...

- Modeling of Speech
  - Single Gaussian Model
    - Maximum likelihood estimation (MLE) of parameters [Sample mean and sample covariance]
    - Pros - Simplicity. Cons - Not capable of modeling speech like signals.
  - Gaussian Mixture Models
    - MLE is not straight forward

$$p(\mathbf{x}|\theta) = \sum_{i=1}^{M} \alpha_i p_i(\mathbf{x}|\theta_i)$$

    - Expectation Maximization (EM) algorithm.
      - Assuming a hidden variable (identity of the mixture component).

Matrix Differentiation Rules

The trace of a square matrix $\text{tr}(A)$ is equal to the sum of A's diagonal elements. The trace of a scalar equals that scalar. Also, $\text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$, and $\text{tr}(AB) = \text{tr}(BA)$ which implies that $\sum_i x_i^T A x_i = \text{tr}(AB)$ where $B = \sum_i x_i x_i^T$. Also note that $|A|$ indicates the determinant of a matrix, and that $|A^{-1}| = 1/|A|$.

We'll need to take derivatives of a function of a matrix $f(A)$ with respect to elements of that matrix. Therefore, we define $\frac{\partial f(A)}{\partial A}$ to be the matrix with i, j^{th} entry $[\frac{\partial f(A)}{\partial a_{i,j}}]$ where $a_{i,j}$ is the i, j^{th} entry of A. The definition also applies taking derivatives with respect to a vector. First, $\frac{\partial x^T A x}{\partial x} = (A + A^T)x$. Second, it can be shown that when A is a symmetric matrix:

$$\frac{\partial |A|}{\partial a_{i,j}} = \begin{cases} \mathcal{A}_{i,j} & \text{if } i = j \\ 2\mathcal{A}_{i,j} & \text{if } i \neq j \end{cases}$$

where $\mathcal{A}_{i,j}$ is the i, j^{th} cofactor of A. Given the above, we see that:

$$\frac{\partial \log |A|}{\partial A} = \left\{ \begin{array}{ll} \mathcal{A}_{i,j}/|A| & \text{if } i = j \\ 2\mathcal{A}_{i,j}/|A| & \text{if } i \neq j \end{array} \right\} = 2A^{-1} - \text{diag}(A^{-1})$$

by the definition of the inverse of a matrix. Finally, it can be shown that:

$$\frac{\partial \text{tr}(AB)}{\partial A} = B + B^T - \text{Diag}(B).$$

"A gentle tutorial on EM Algorithm and its application to parameter estimation of GMM and HMM", Jeff Bilmes

EM Algorithm

$$Q(\Theta, \Theta^{(i-1)}) = E\left[\log p(\mathcal{X}, \mathcal{Y}|\Theta)|\mathcal{X}, \Theta^{(i-1)}\right]$$

$$E\left[\log p(\mathcal{X}, \mathcal{Y}|\Theta)|\mathcal{X}, \Theta^{(i-1)}\right] = \int_{\mathbf{y}\in\Upsilon} \log p(\mathcal{X}, \mathbf{y}|\Theta) f(\mathbf{y}|\mathcal{X}, \Theta^{(i-1)}) d\mathbf{y}.$$

$$\Theta^{(i)} = \underset{\Theta}{\operatorname{argmax}}\ Q(\Theta, \Theta^{(i-1)}).$$

"A gentle tutorial on EM Algorithm and its application to parameter estimation of GMM and HMM", Jeff Bilmes

GMM Re-estimation Formulas

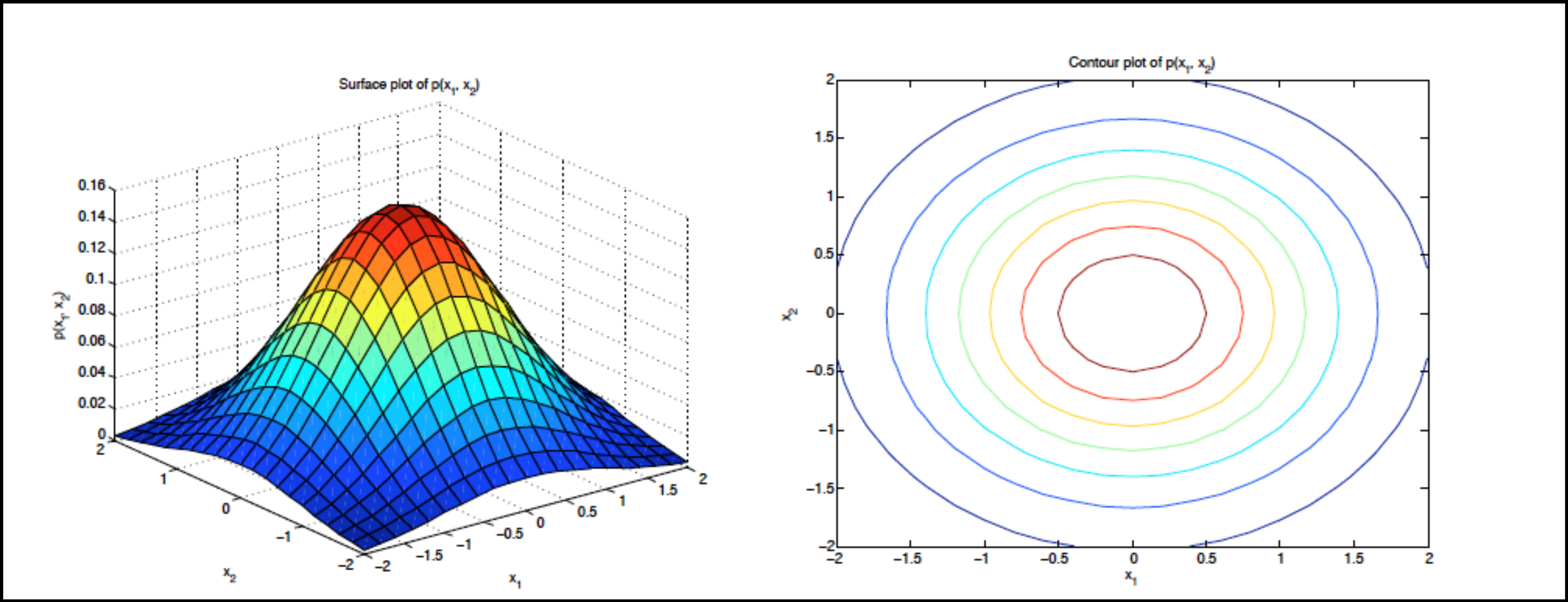
$$\alpha_\ell^{new} = \frac{1}{N}\sum_{i=1}^{N} p(\ell|x_i, \Theta^g)$$

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

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

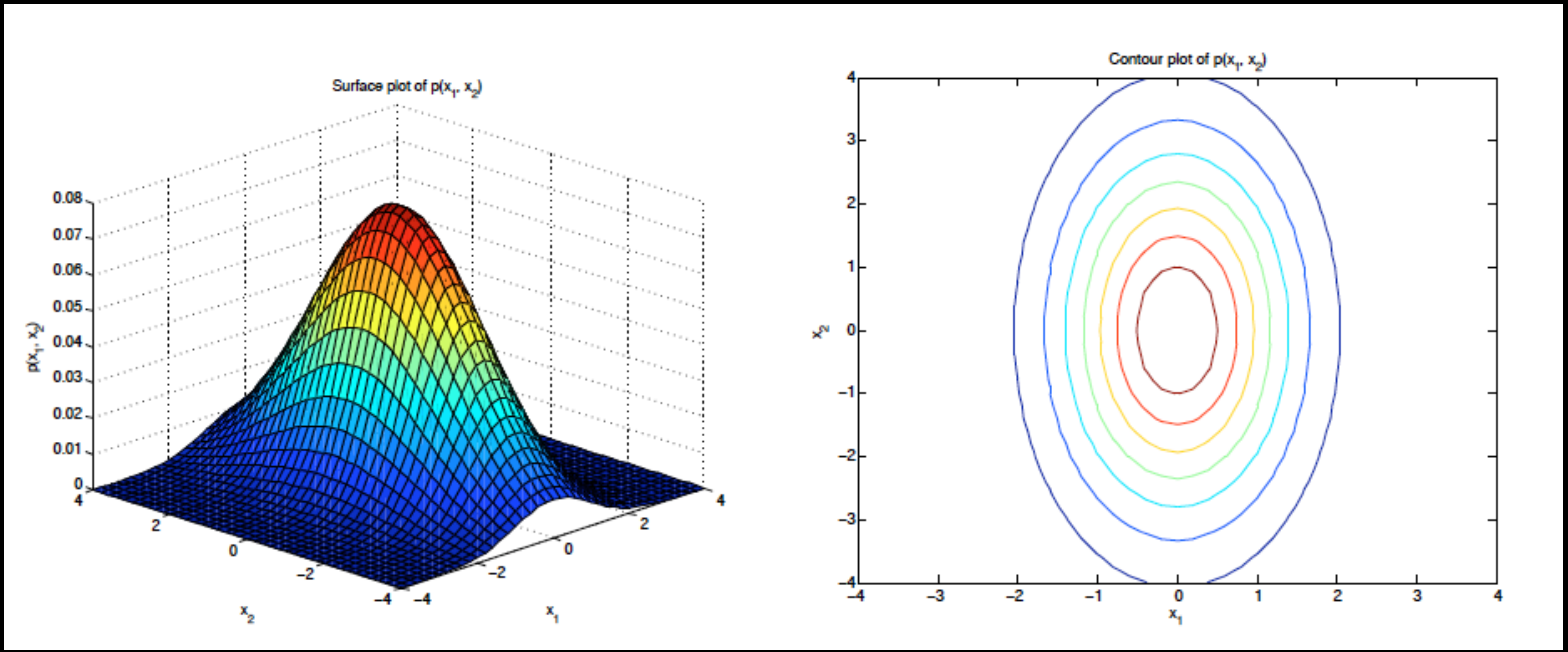
"A gentle tutorial on EM Algorithm and its application to parameter estimation of GMM and HMM", Jeff Bilmes

Some Insights - Gaussian Distributions



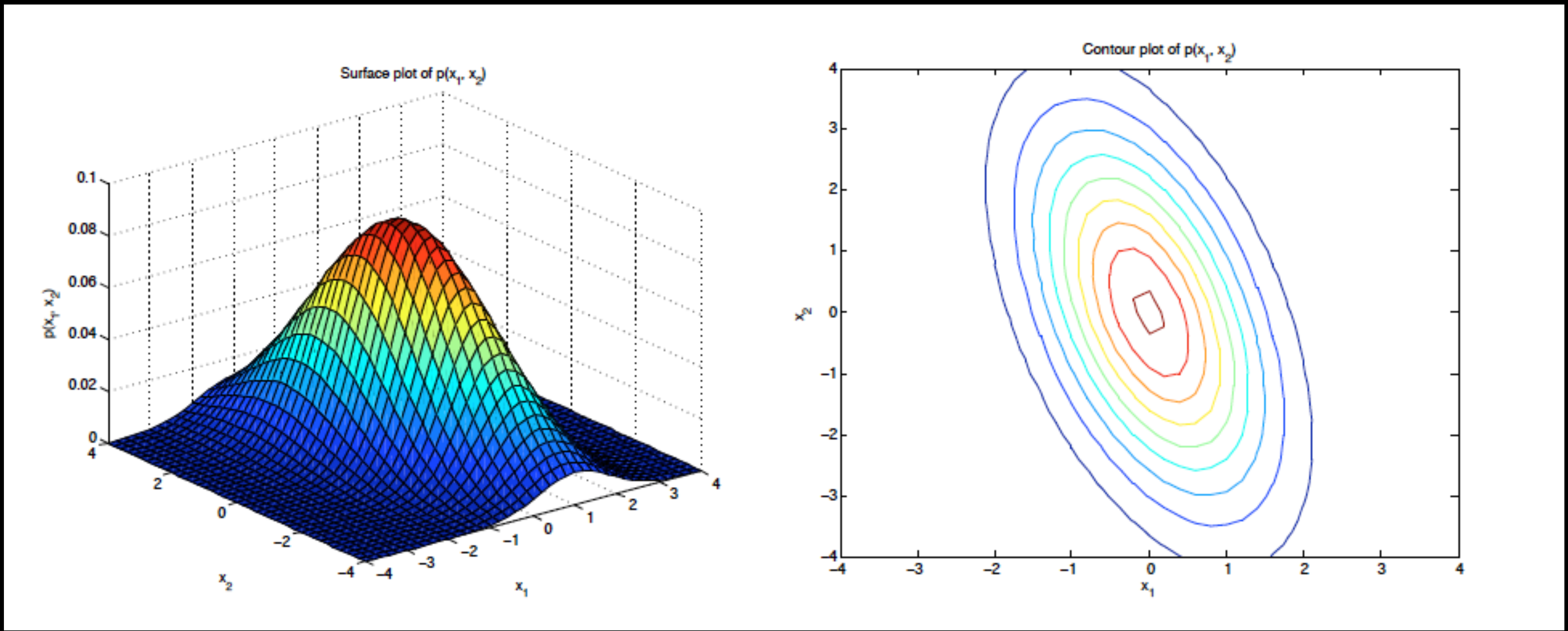
"Hidden Markov Models and Gaussian Mixture Models", ASR Lectures, Steve Renals and Peter Ball

Some Insights - Gaussian Distributions



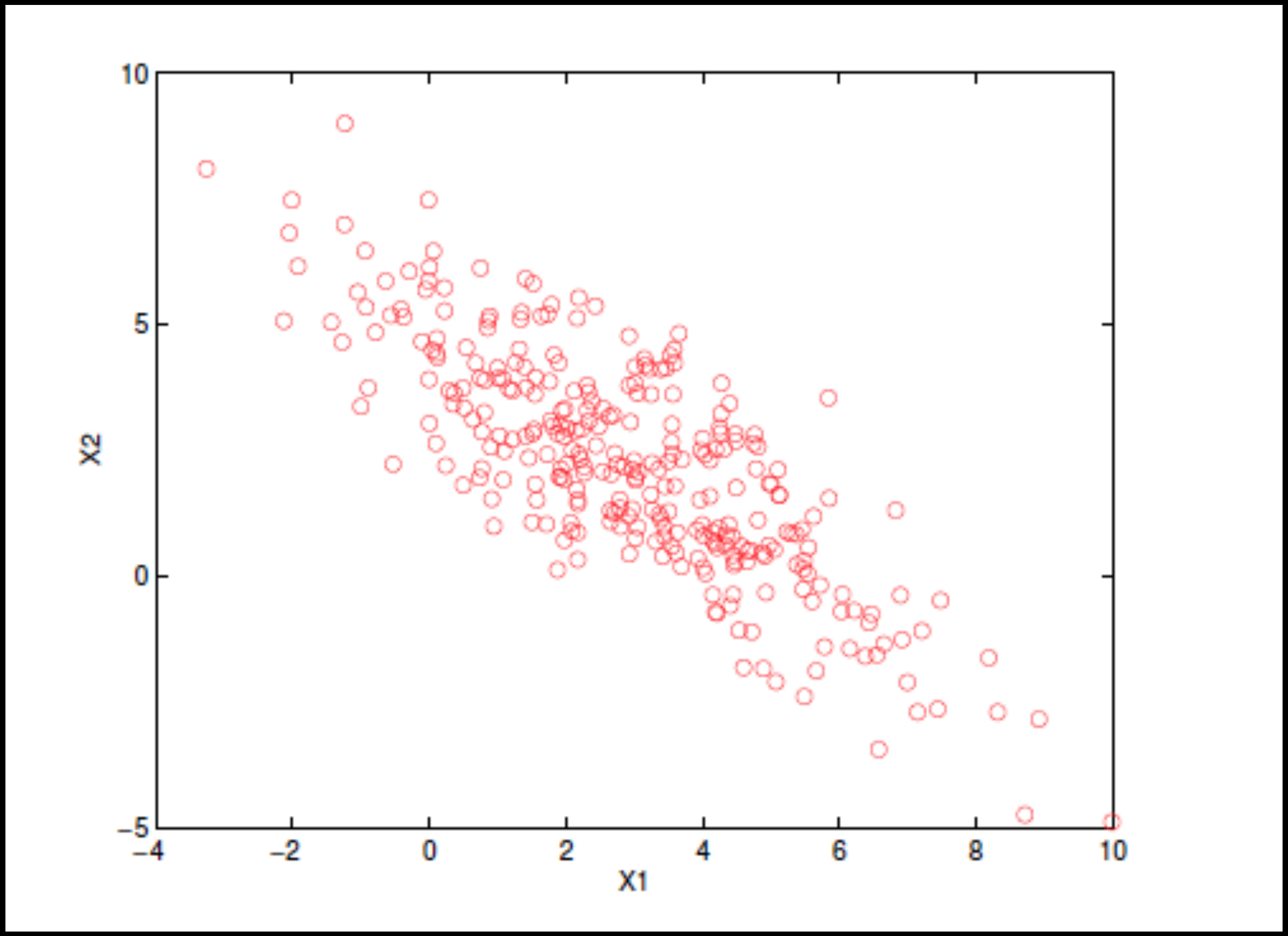
"Hidden Markov Models and Gaussian Mixture Models", ASR Lectures, Steve Renals and Peter Ball

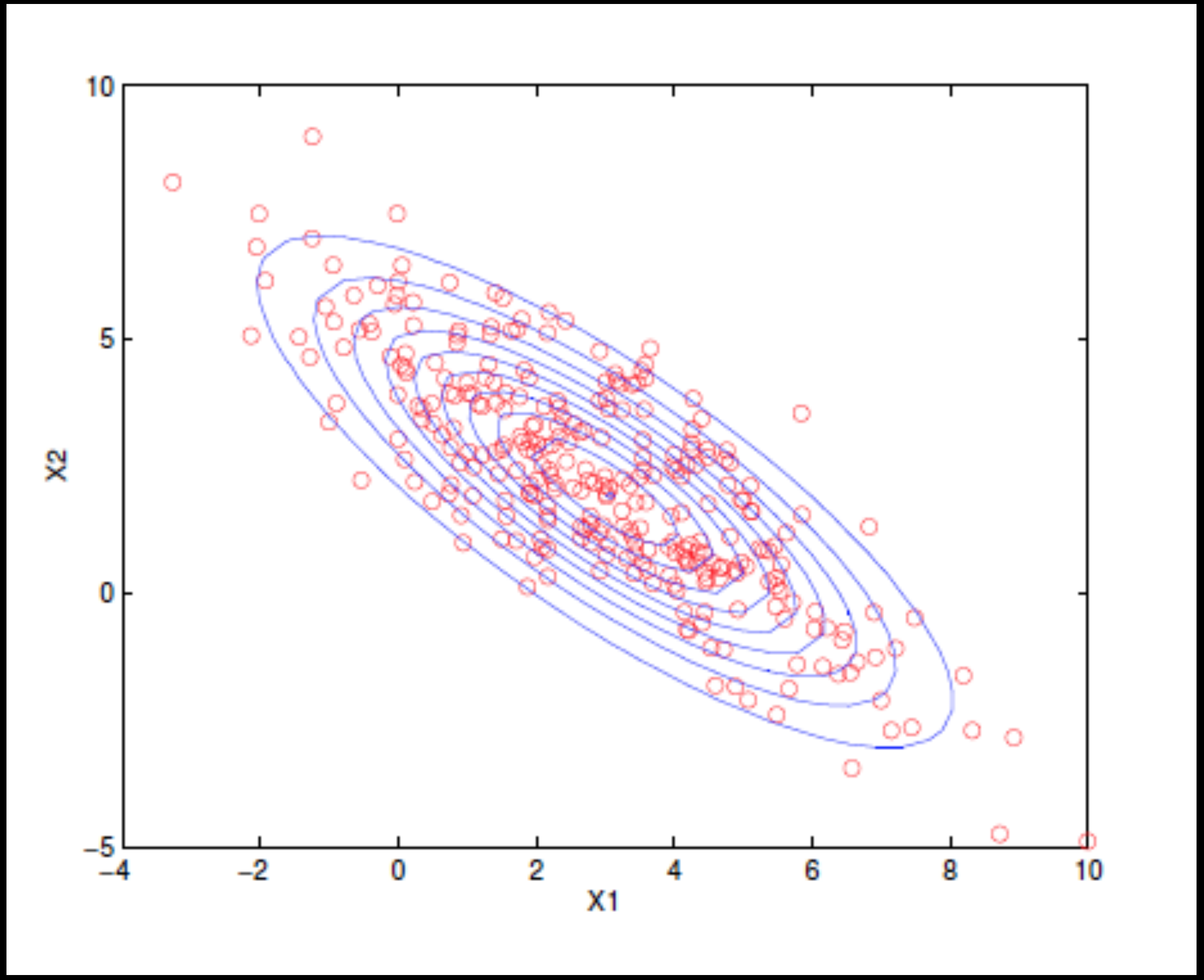
Some Insights - Gaussian Distributions



"Hidden Markov Models and Gaussian Mixture Models", ASR Lectures, Steve Renals and Peter Ball

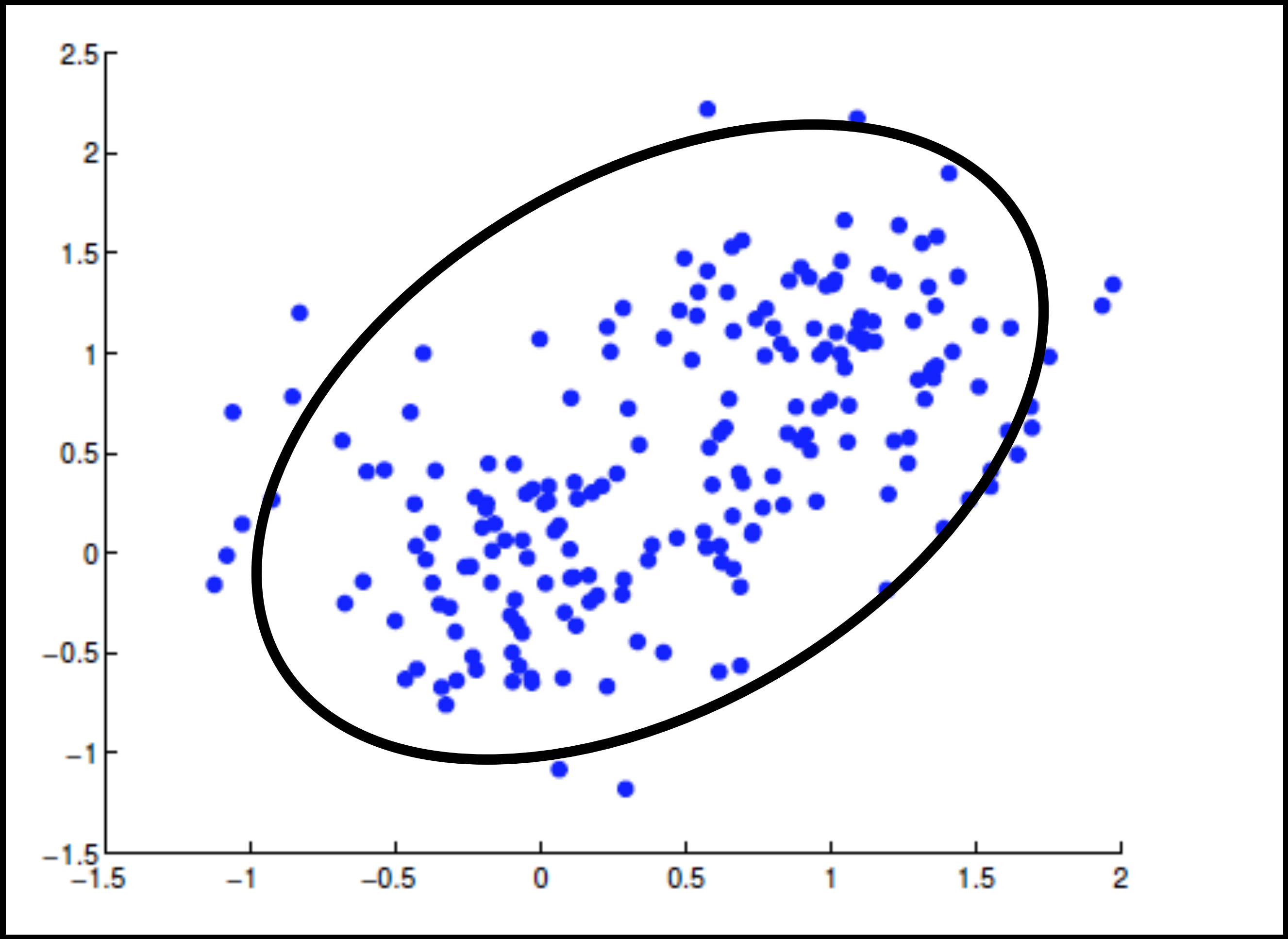
Some Insights - Gaussian Distributions





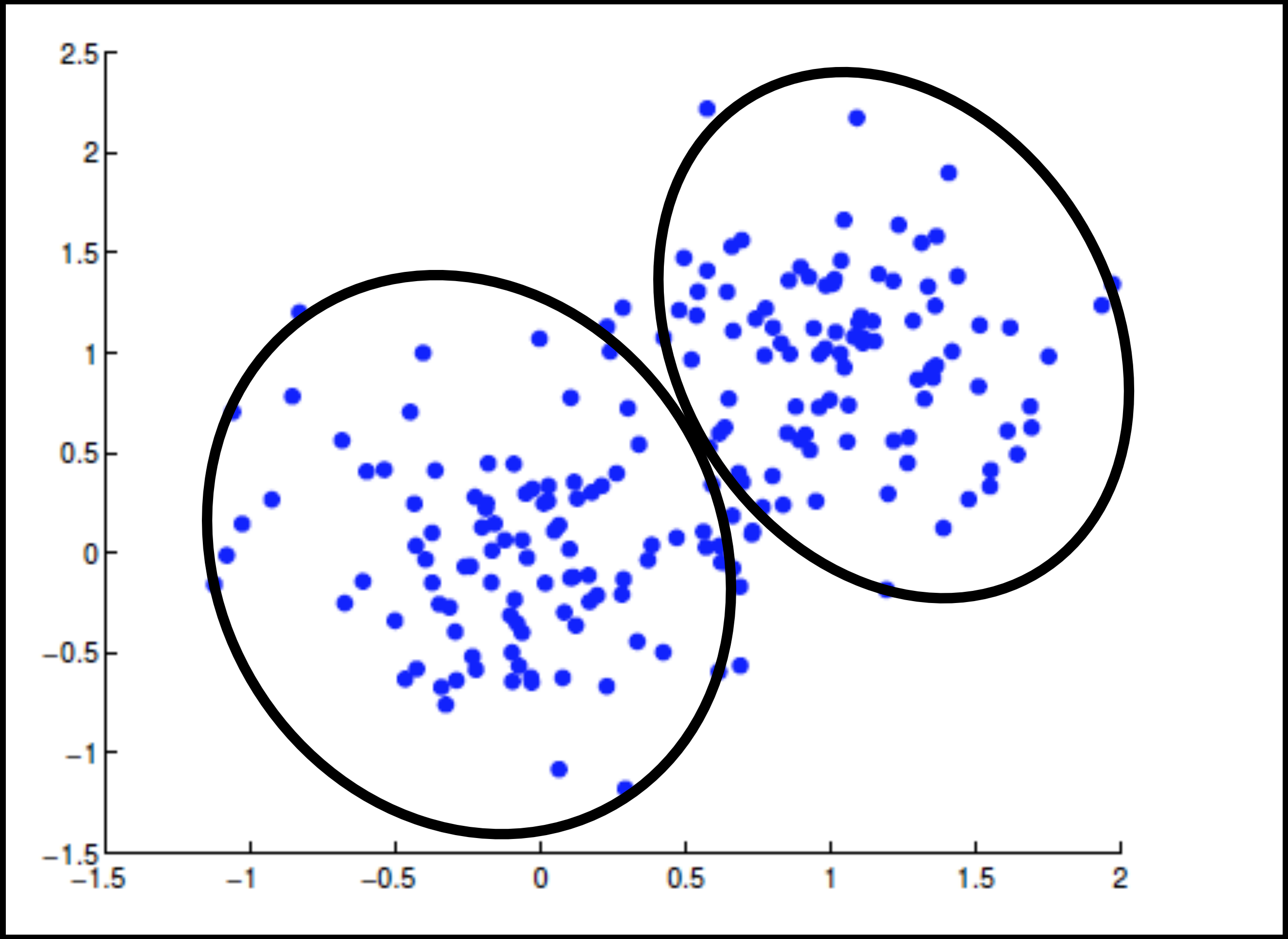
"Hidden Markov Models and Gaussian Mixture Models", ASR Lectures, Steve Renals and Peter Ball

Some Insights - Gaussian Distributions



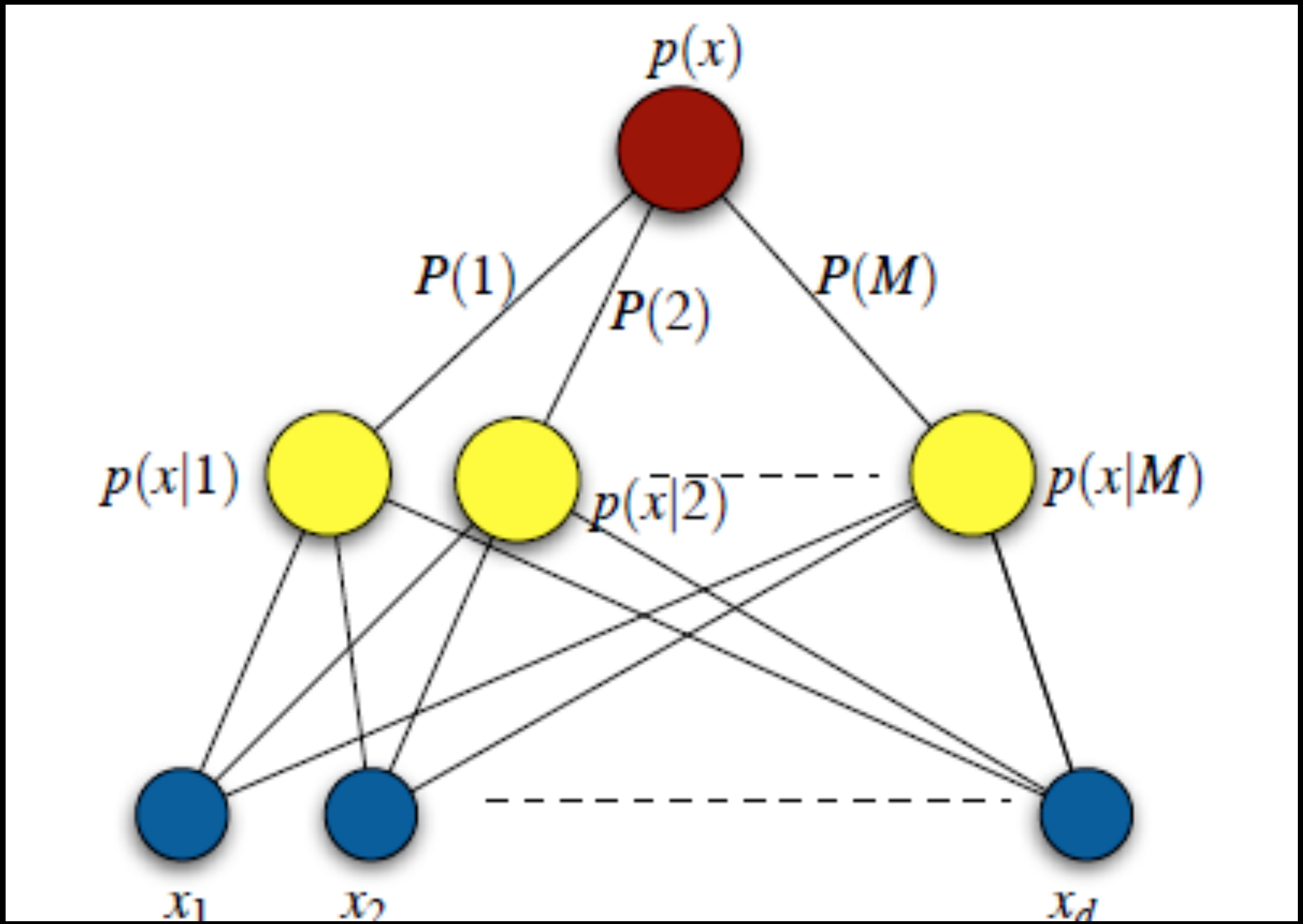
"Hidden Markov Models and Gaussian Mixture Models", ASR Lectures, Steve Renals and Peter Ball

Some Insights - Gaussian Distributions



"Hidden Markov Models and Gaussian Mixture Models", ASR Lectures, Steve Renals and Peter Ball

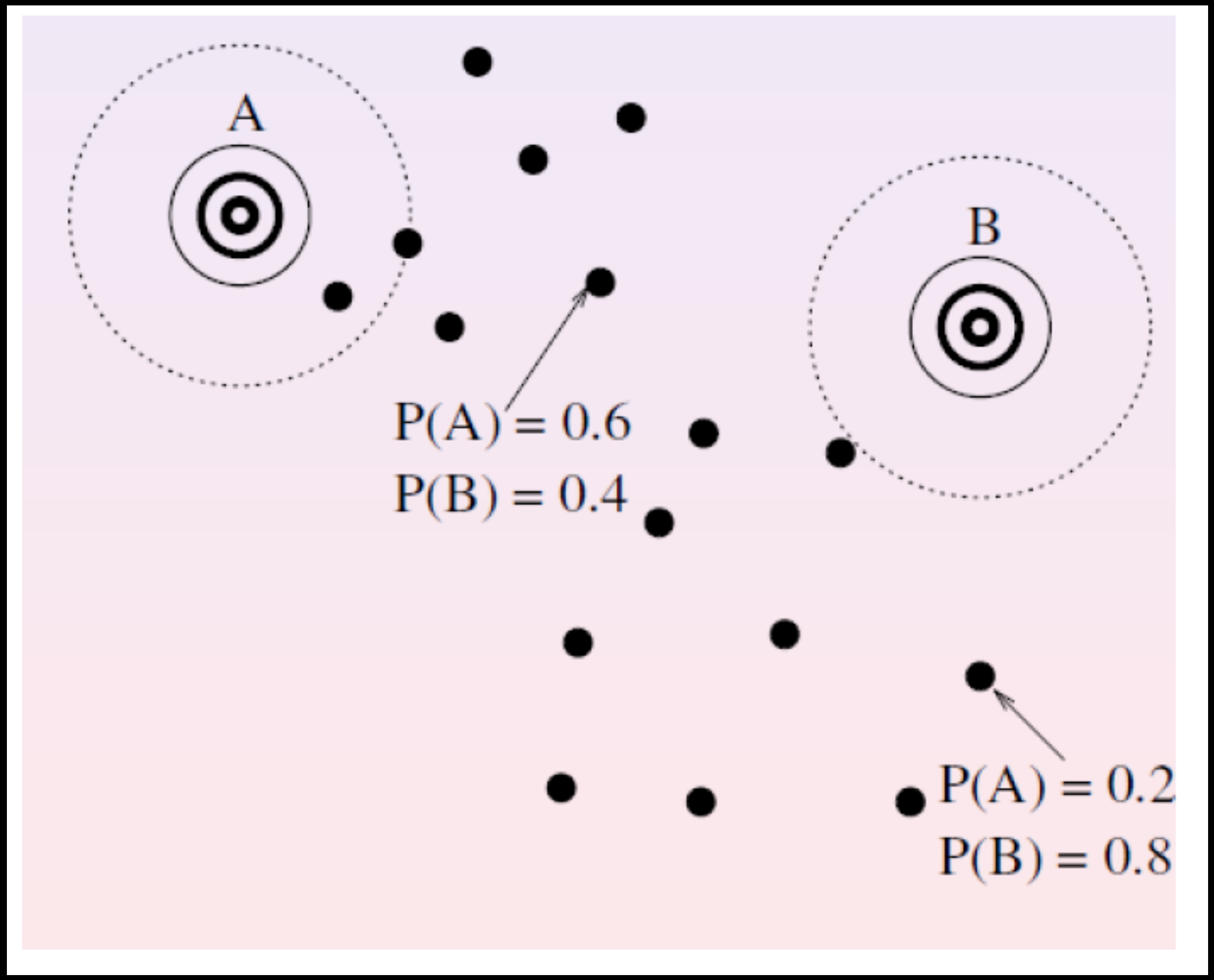
Gaussian Mixture Models



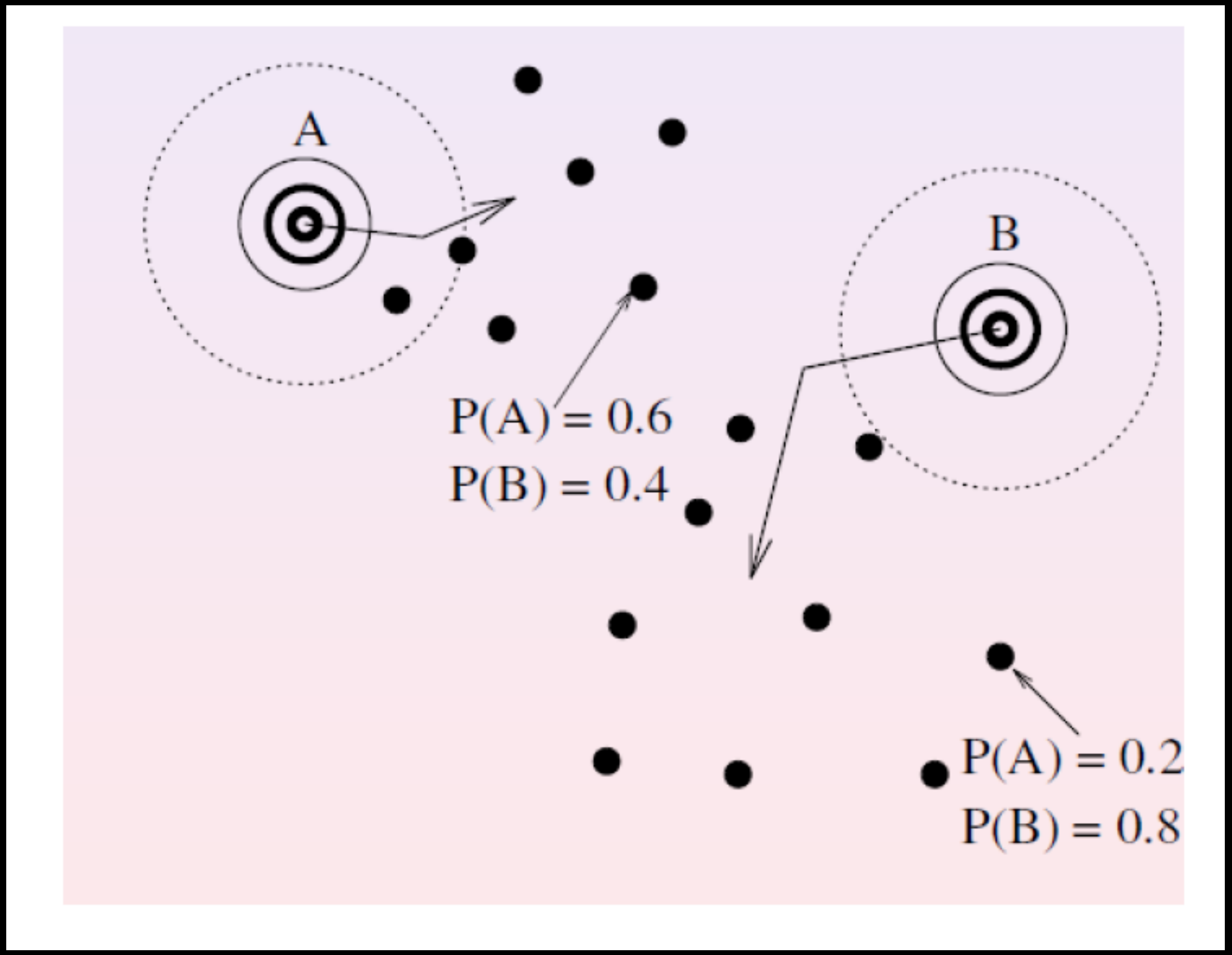
"Hidden Markov Models and Gaussian Mixture Models", ASR Lectures, Steve Renals and Peter Ball

GMM EM Algorithm

E-step

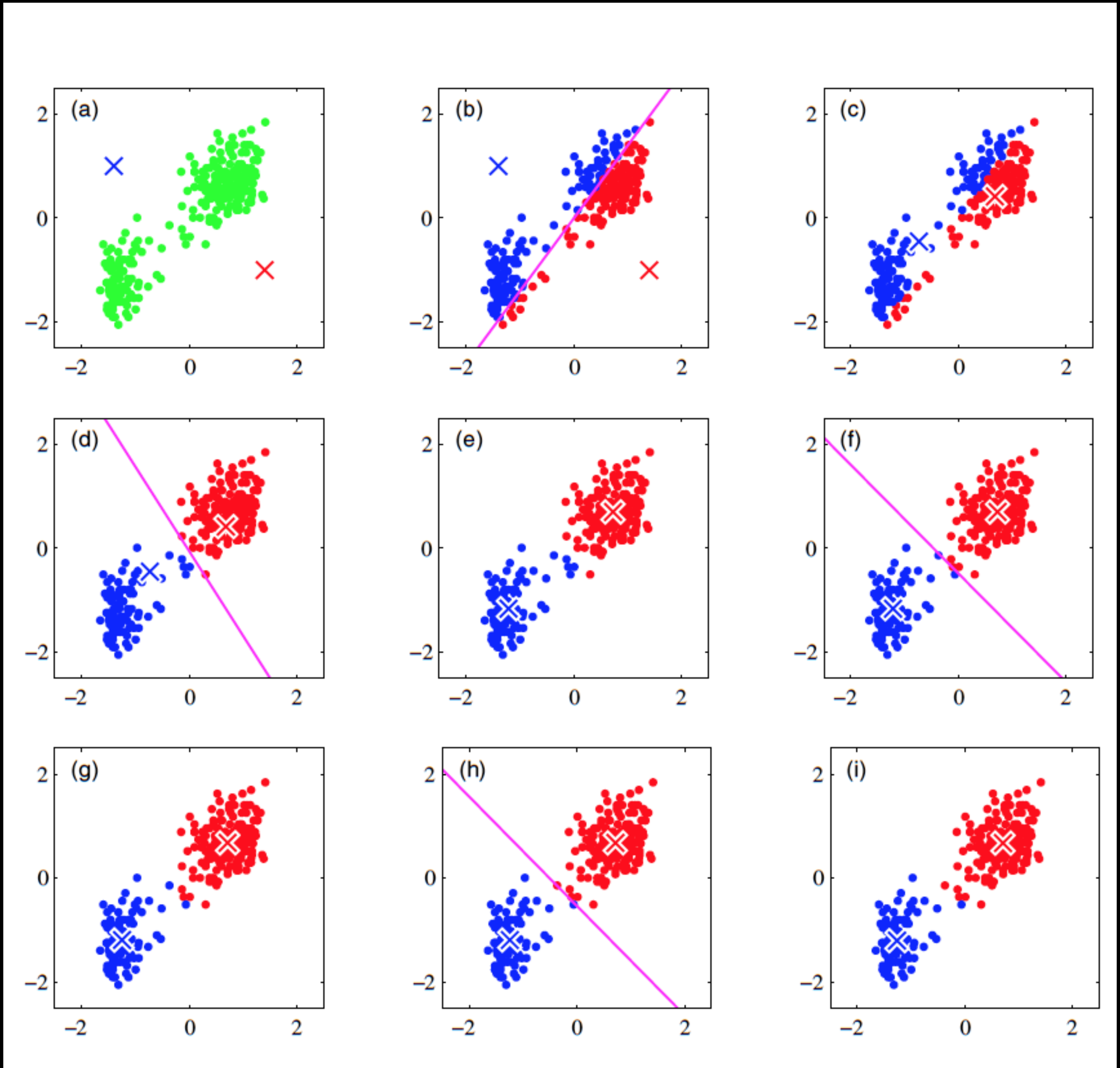


M-step



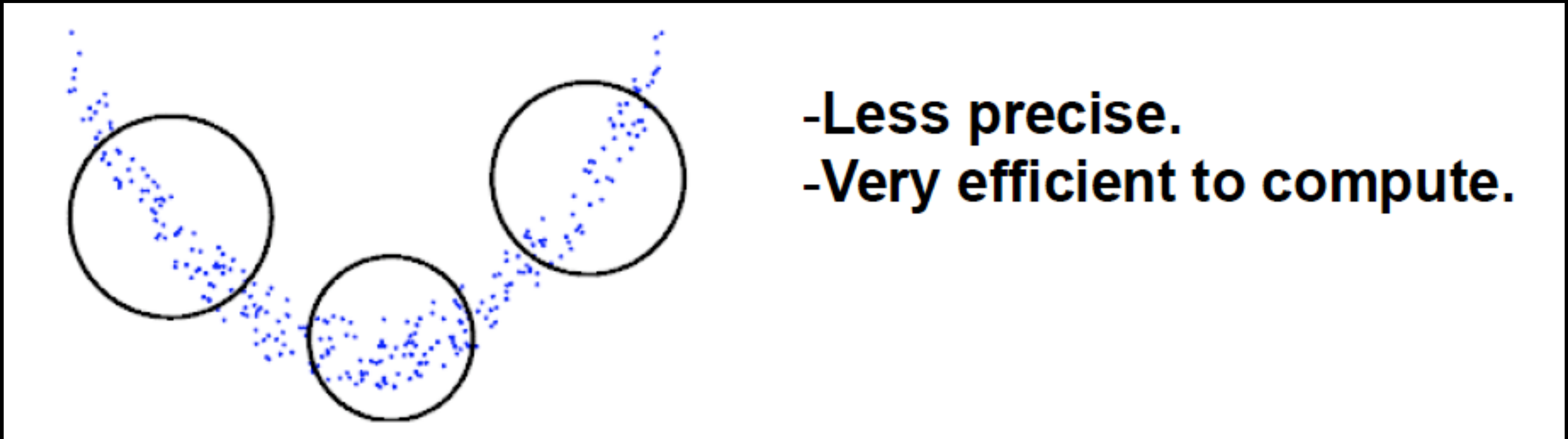
"A short tutorial on GMM models" - Mohand Said, Ali, University of Quebec

K-means Algorithm



Other Considerations

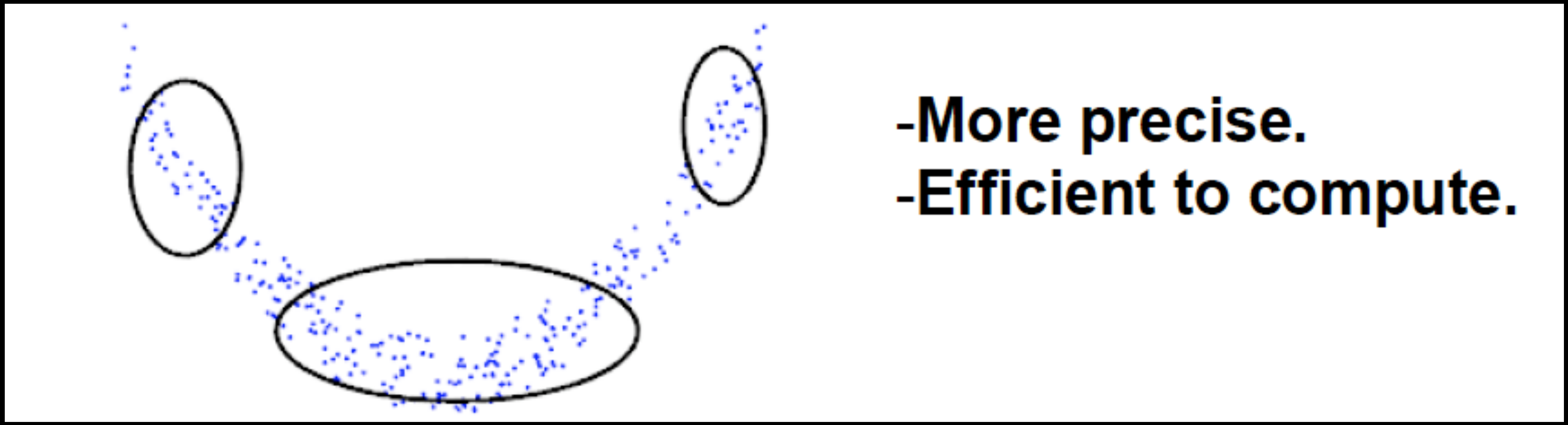
- Initialization
  - Random Initialization, K-means algorithm, ….
  - Number of Gaussians (Based on the data availability)
  - Type of covariance structure
    - Spherical covariance



“A short tutorial on GMM models” - Mohand Said, Ali, University of Quebec

Other Considerations

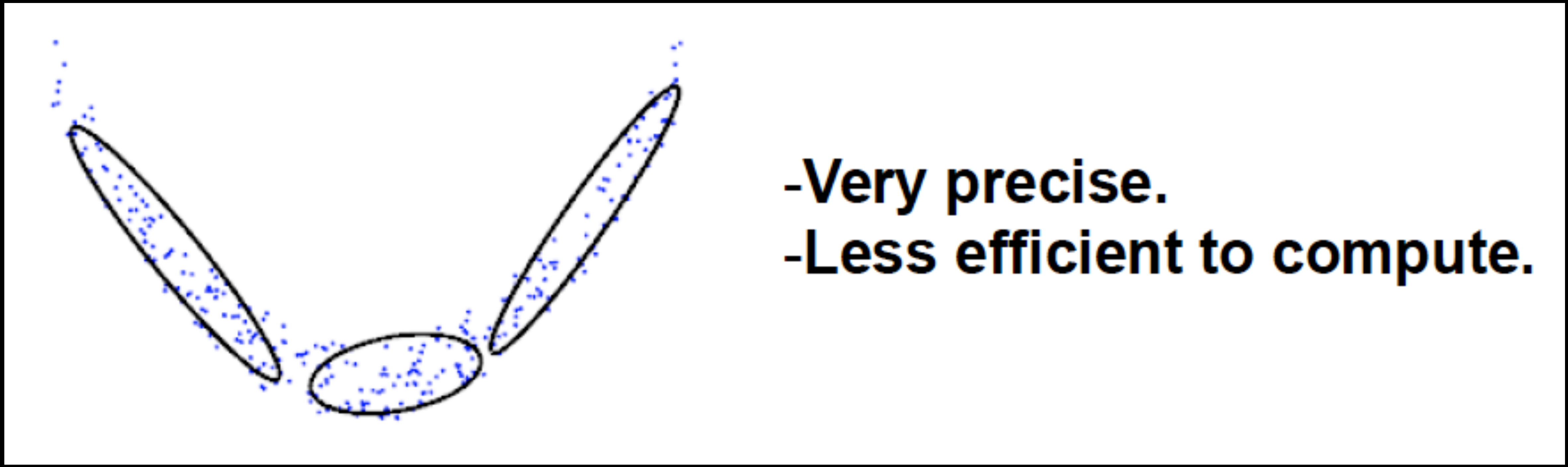
- Initialization
  - Random Initialization, K-means algorithm, ….
  - Number of Gaussians (Based on the data availability)
  - Type of covariance structure
    - Diagonal covariance



"A short tutorial on GMM models" - Mohand Said, Ali, University of Quebec

Other Considerations

- Initialization
  - Random Initialization, K-means algorithm, ….
  - Number of Gaussians (Based on the data availability)
  - Type of covariance structure
    - Full covariance



“A short tutorial on GMM models” - Mohand Said, Ali, University of Quebec