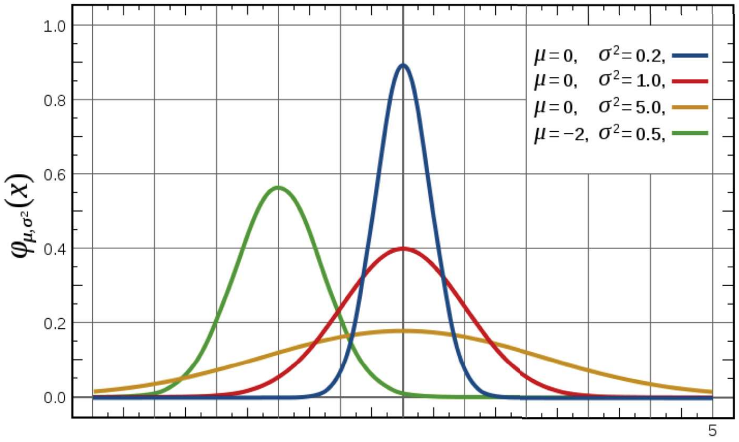
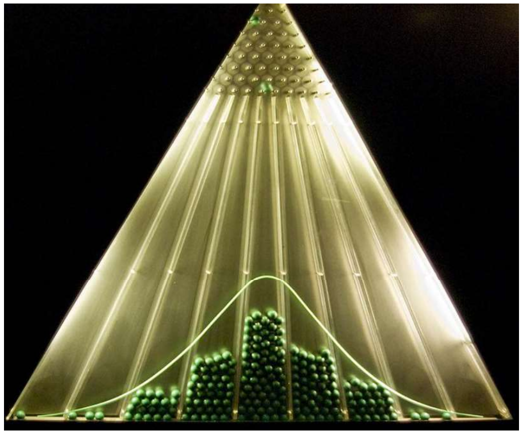
Gaussian Distribution

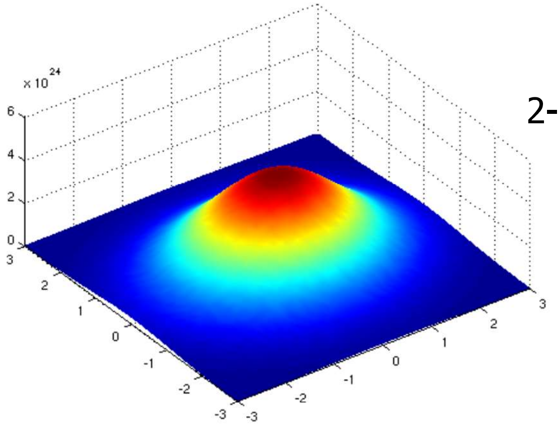


1-d Gaussian

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Bean machine: drop ball with pins





2-d Gaussian

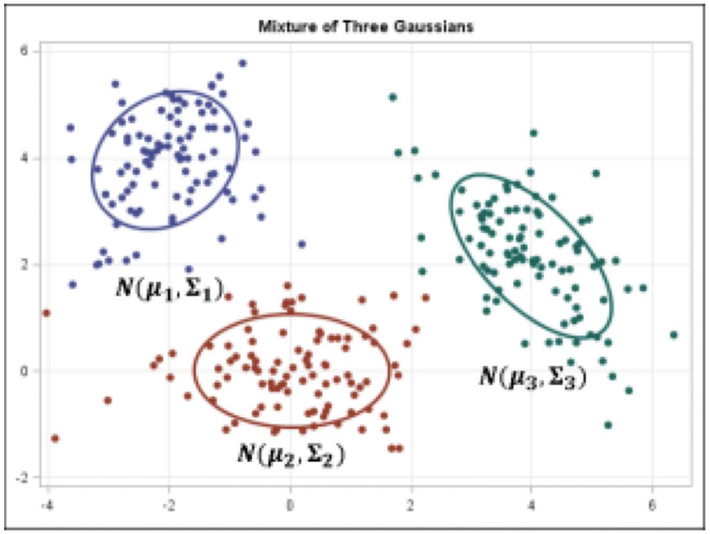
$$f_{\mathbf{X}}(x_1, \ldots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}$$

From wikipedia and http://home.dei.polimi.it

Gaussian Mixture Model

- Assumptions
  - Each data point comes from one of K classes.
  - The cluster prior distribution w_j is *unknown*.
  - Each class c_j follows a Gaussian

distribution: $P(x|c_j,\theta_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}}$



  - The parameters for each class μ_j,σ_j are *unknown*(need to be learned).
  - The probability of x_i is the sum over all classes, $P(x_i|\theta) = \sum_{j=1}^{K} P(x_i|\mathrm{c}_j,\theta_j)P(c_j)$

Soft Clustering with Gaussian Mixture Model

- ❑ Every object i is assigned to one cluster j with a probability
  - ❑ $P(z_i = j) \in [0,1]$ and $\Sigma_j P(z_i = j) = 1$
  - ❑ Where z_i is a hidden variable of which cluster x_i belongs to.

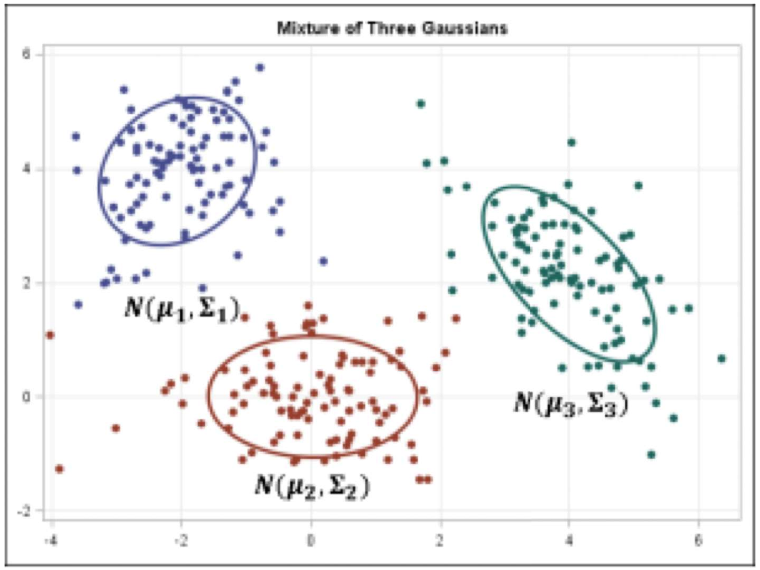
Assume the parameters of the GMM have been learned

- ❑ The probability of x_i belonging to cluster c_j :
  - ❑ $P(z_i = c_j | x_i) \ \propto P(x_i, z_i = c_j)$

$$= w_j P(x_i | z_i = c_j)$$

Cluster prior probabilities

Probability density function of each cluster



The E-M(Expectation Maximization) Algorithm

- A framework to approach maximum likelihood or maximum a posteriori estimates of parameters in statistical models.
- Expectation Step:
  - Assigns objects to clusters according to the current soft clustering or parameters of probabilistic clusters
  - $w_{ij}^{t+1} = P\left(z_i = j \middle| x_i, \theta_j^t\right) \propto w_j P\left(x_i \middle| z_i = j, \theta_j^t\right)$

Joint probability of x_i and its cluster c_j

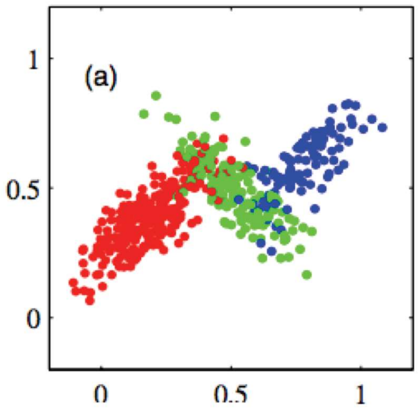
- Maximization Step:
  - finds the new parameters of each cluster that maximize the expected likelihood
  - $\theta_{t+1} = argmax_\theta \Sigma_i \Sigma_j w_{ij}^{t+1} logL(x_i, z_j|\theta)$

Example: Applying E-M algorithm to 1-D GMM

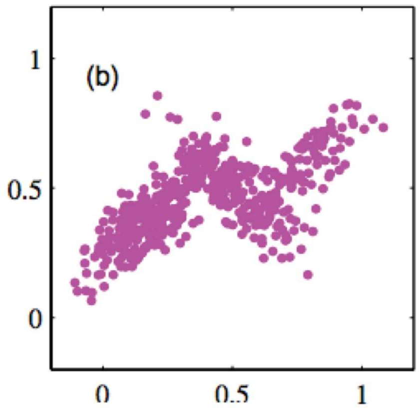
- Iteratively do the following two steps
  - E-Step: Evaluate the soft clustering probability according to $\mu_j^t, \sigma_j^t, w_j^t$
    - $w_{ij}^{t+1} = \frac{w_j^t P\left(x_i \middle| \mu_j^t, \sigma_j^t\right)}{\Sigma_k w_k^t P\left(x_i \middle| \mu_k^t, \sigma_k^t\right)}$
  - M-Step: Find the new parameters μ_j^t, σ_j^t that maximize log likelihood. In Gaussian distribution, this is equivalent to do parameter estimation when each data point has a weight.
    - $\mu_j^{t+1} = \frac{\Sigma_i w_{ij}^{t+1} x_i}{\Sigma_i w_{ij}^{t+1}}, \left(\sigma_j^2\right)^{t+1} = \frac{\Sigma_i w_{ij}^{t+1} (x_i - \mu_j^{t+1})^2}{\Sigma_i w_{ij}^{t+1}}$ ⇦ Weighted average means and variance
    - $w_j^{t+1} = \frac{\Sigma_i w_{ij}^{t+1}}{n}$

Gaussian Mixture Model

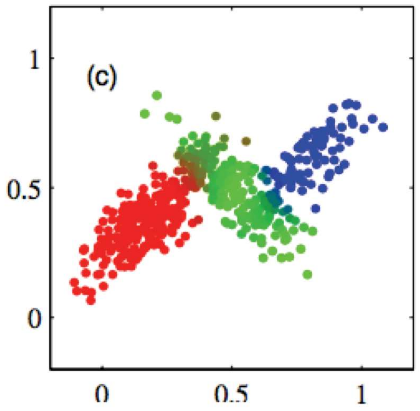
- Example of applying Gaussian Mixture Model



The data points belong to three classes. Each class follows a Gaussian distribution.



We hide the class information.



Class information inferred by GMM.

- We can use E-M algorithm to learn the parameters.

Example: Applying E-M algorithm to GMM