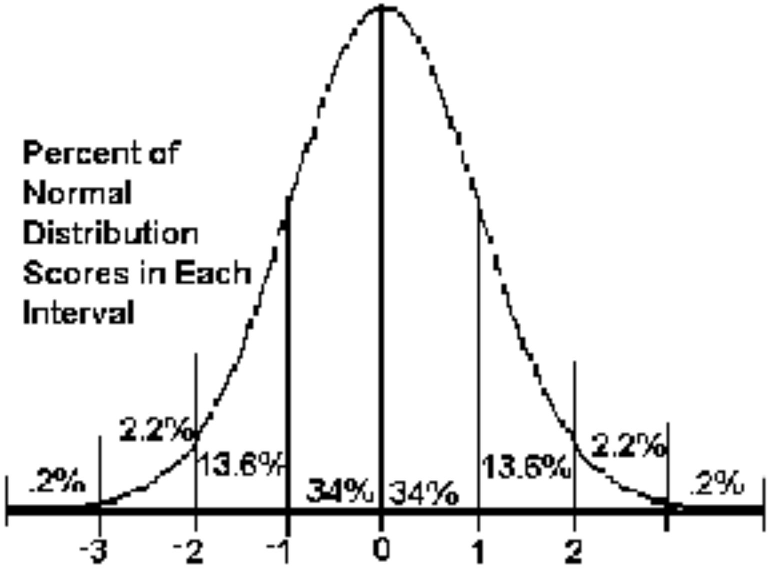
EXPECTATION MAXIMIZATION

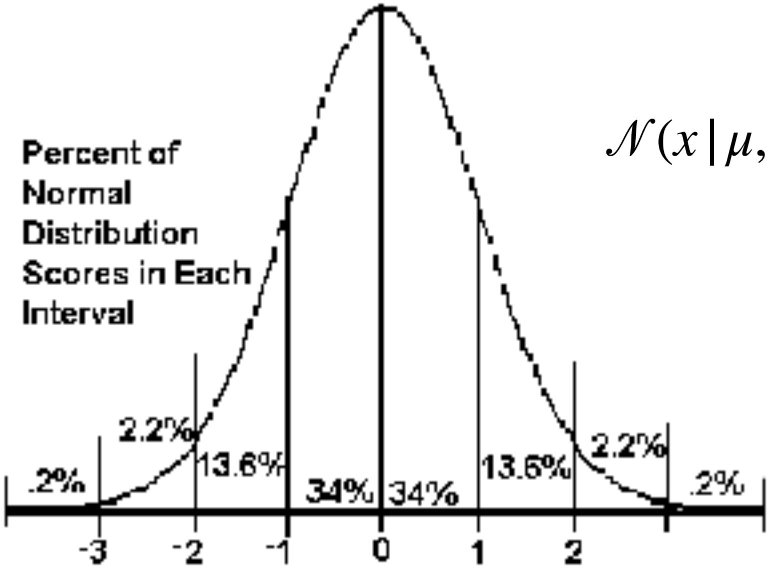
Notation: Normal distribution 1D case

N(μ , σ) is a 1D normal (Gaussian) distribution with mean μ and standard deviation σ (so the variance is σ^2.



Notation: Normal distribution 1D case

N(μ , σ) is a 1D normal (Gaussian) distribution with mean μ and standard deviation σ (so the variance is σ^2.



$$\mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\}$$

Multivariate Normal distribution

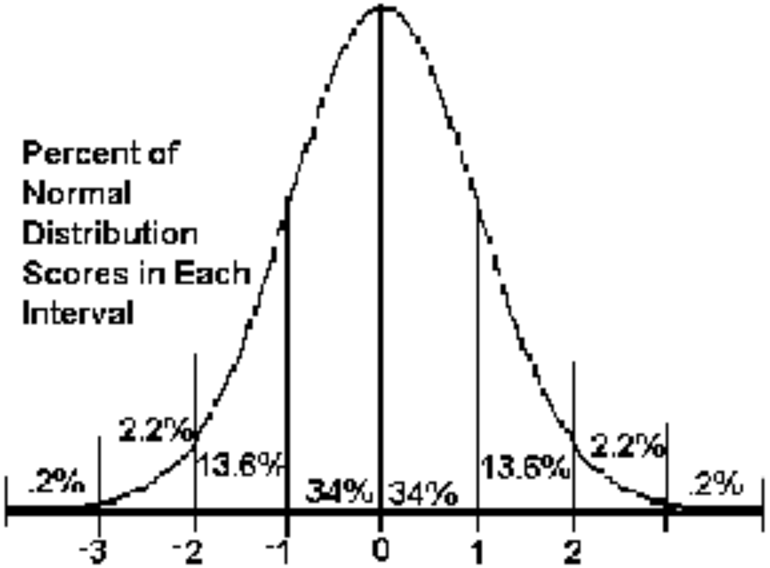
$$\mathcal{N}(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} exp\left\{ -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right\}$$

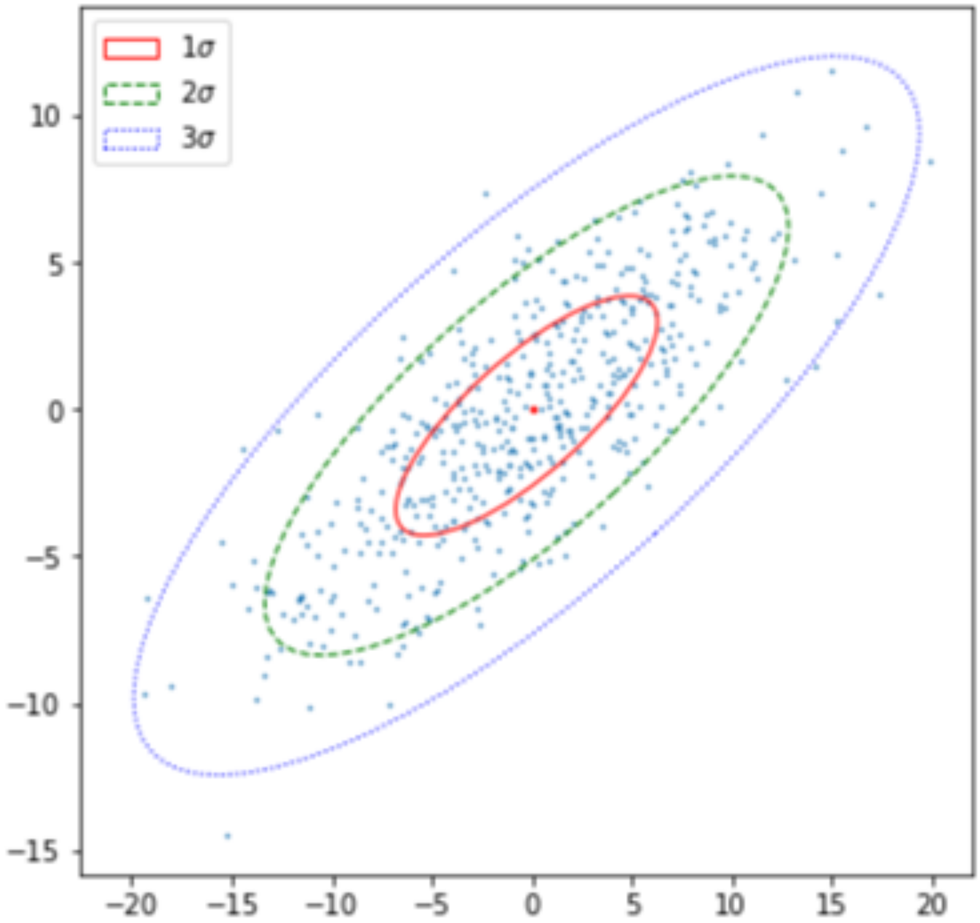
x is a D dimensional vector

μ is a D-dimensinal mean vector

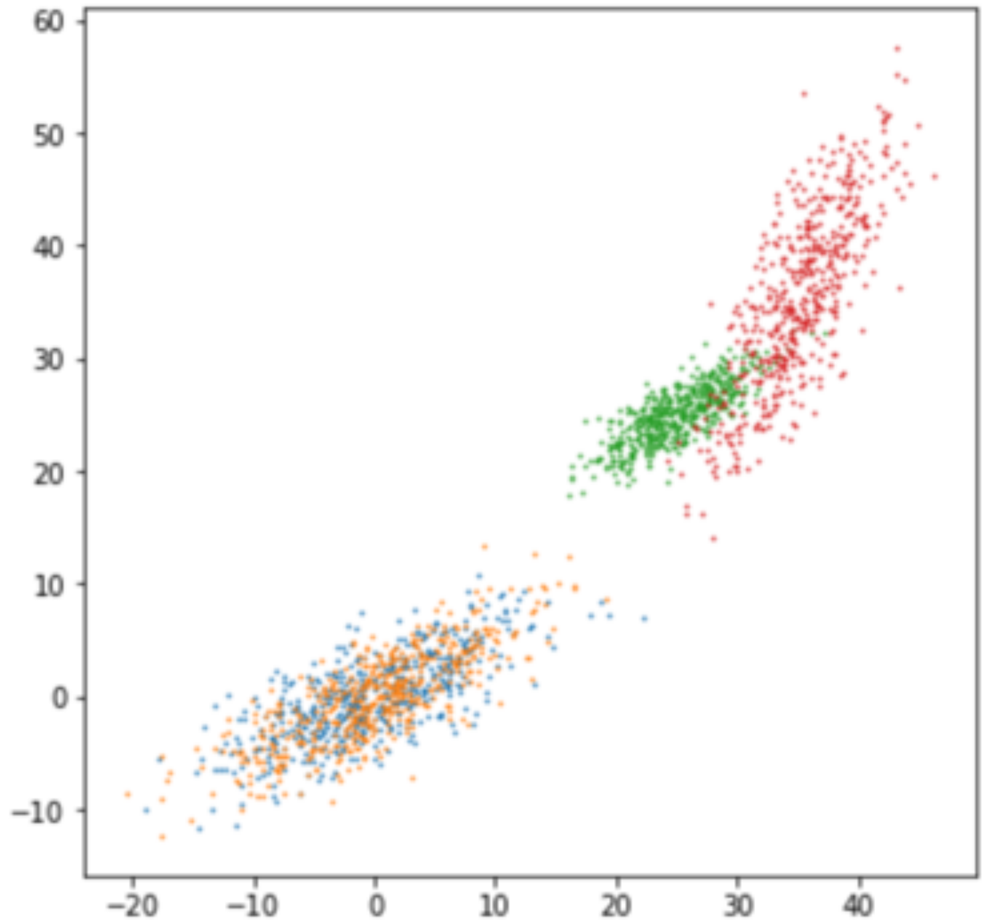
Σ is a D x D covariance matrix

Uni-modal dataset

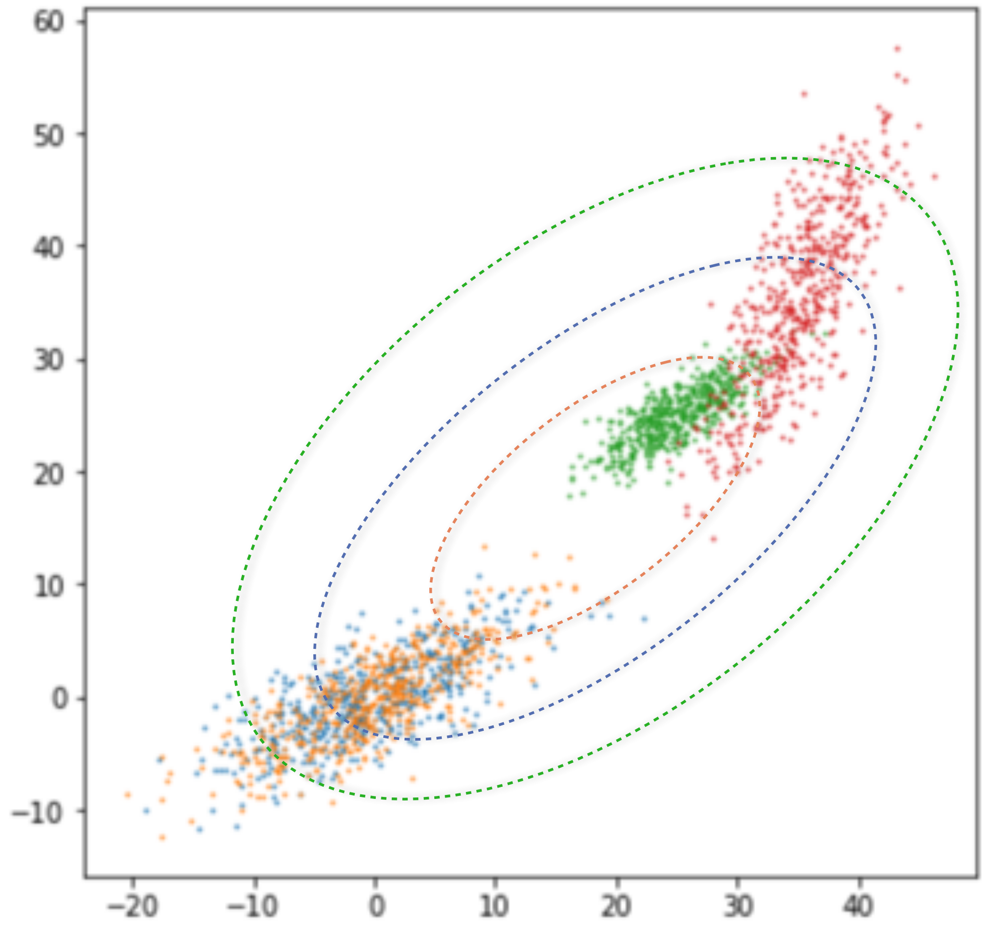




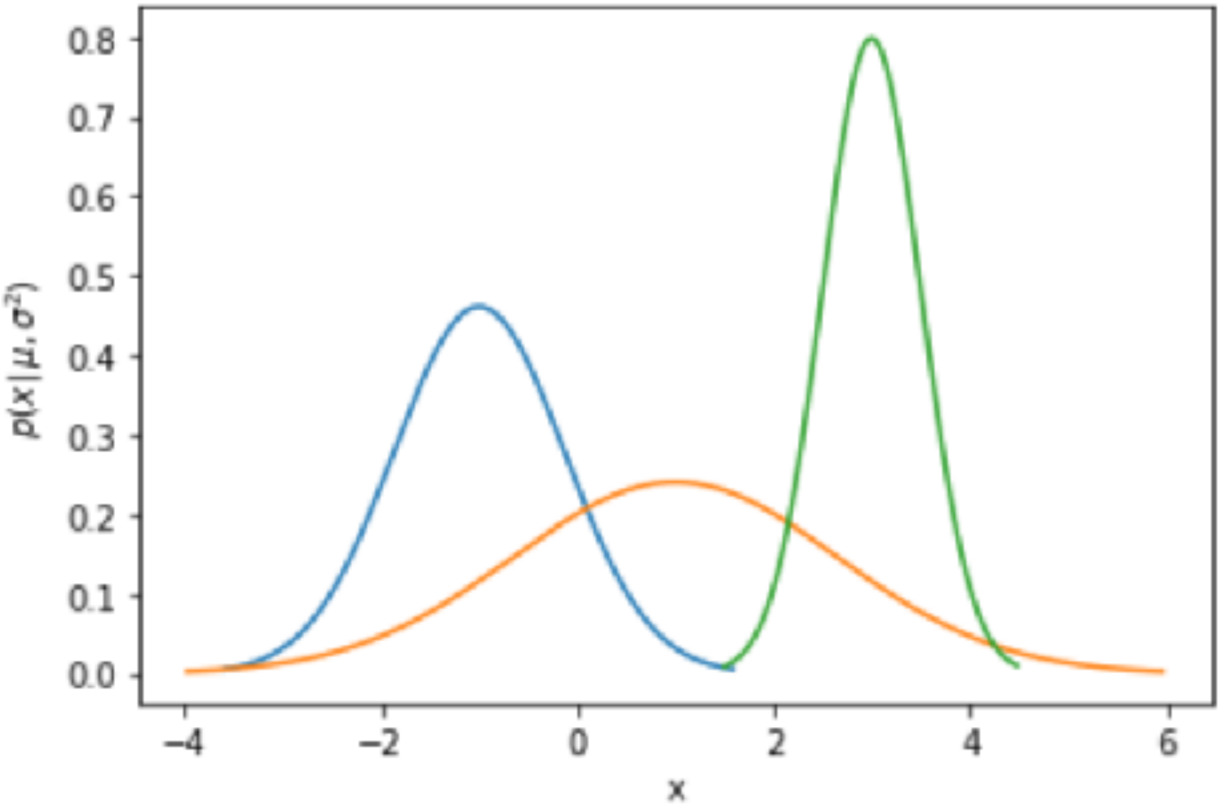
Multi-modal dataset



Multi-modal dataset

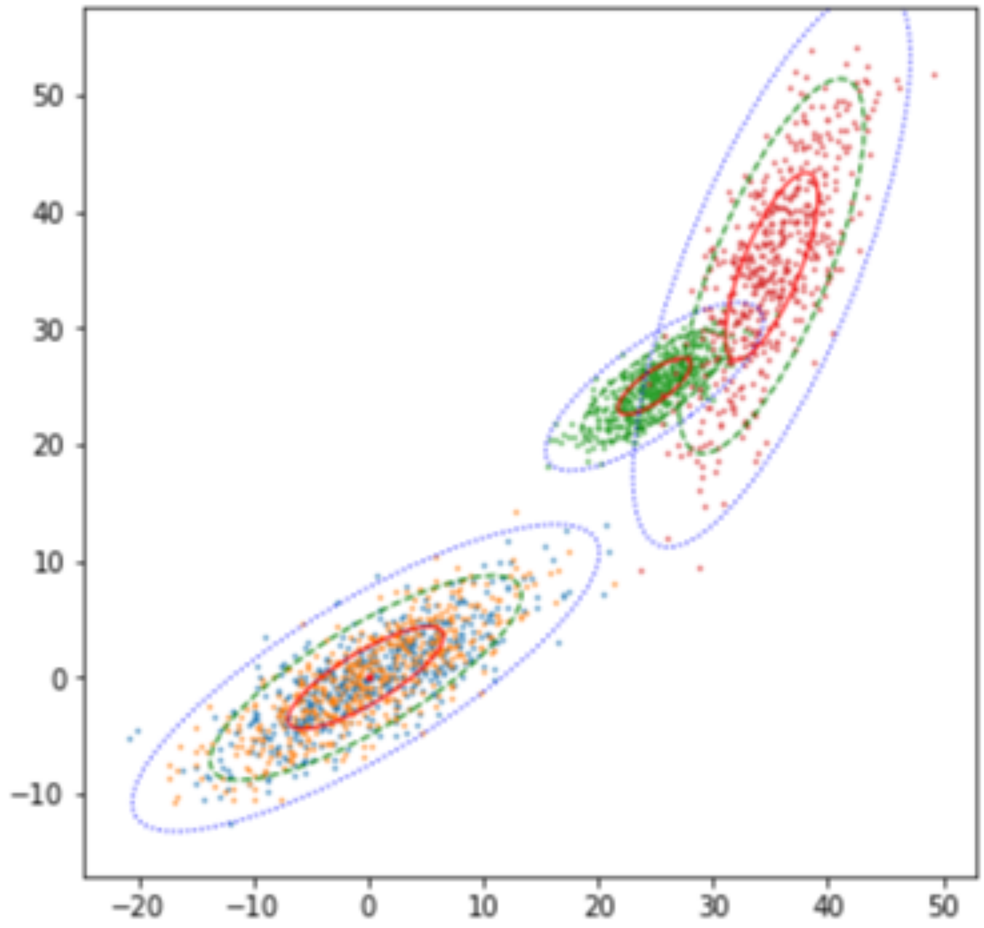


Multi-modal dataset



$$\mathcal{N}(x\,|\,\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\}$$

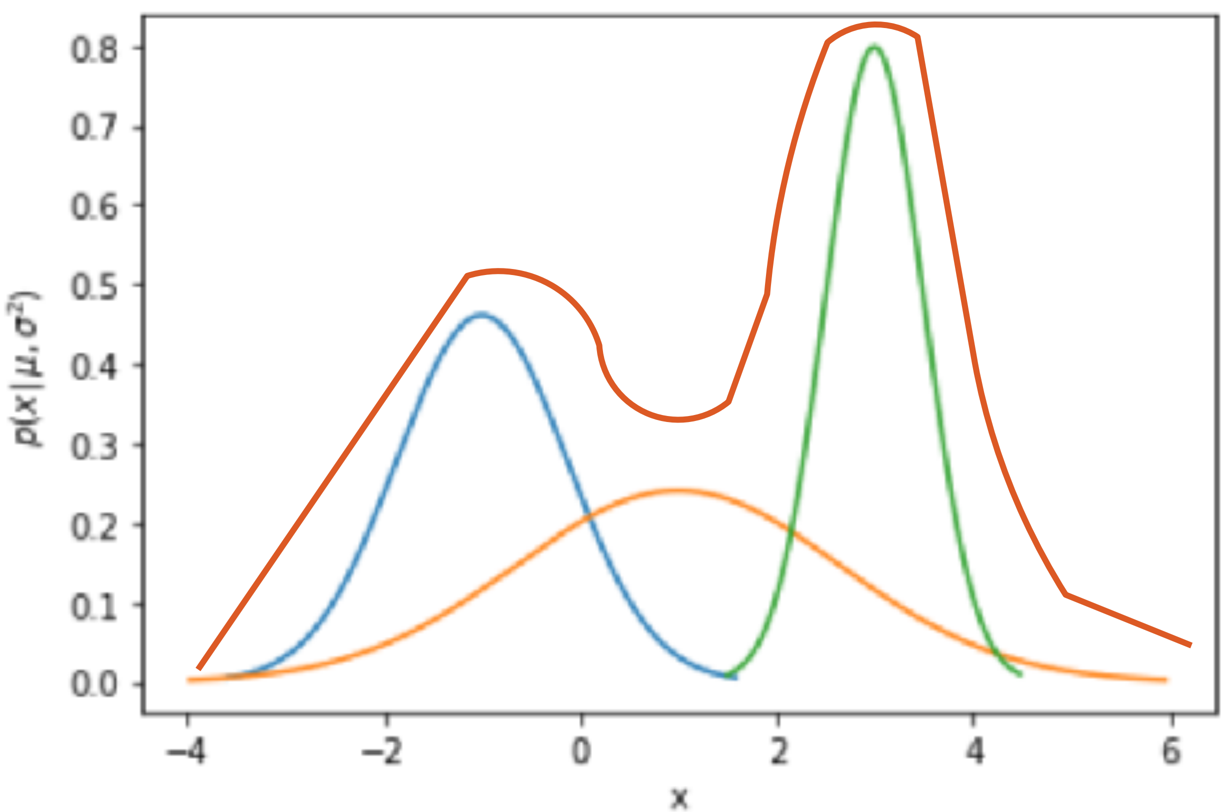
Multi-modal dataset



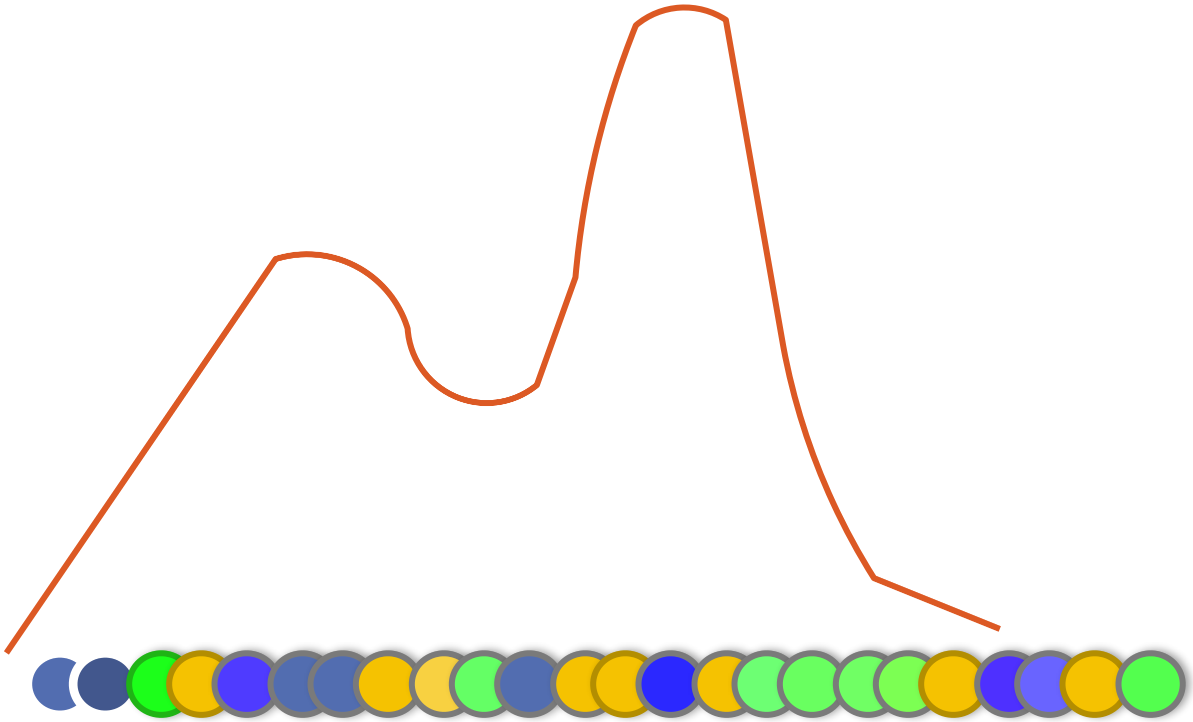
Gaussian Mixtures Model

A linear combination of Gaussian distributions forms a superposition

Formulated as a probabilistic model known as mixture distribution

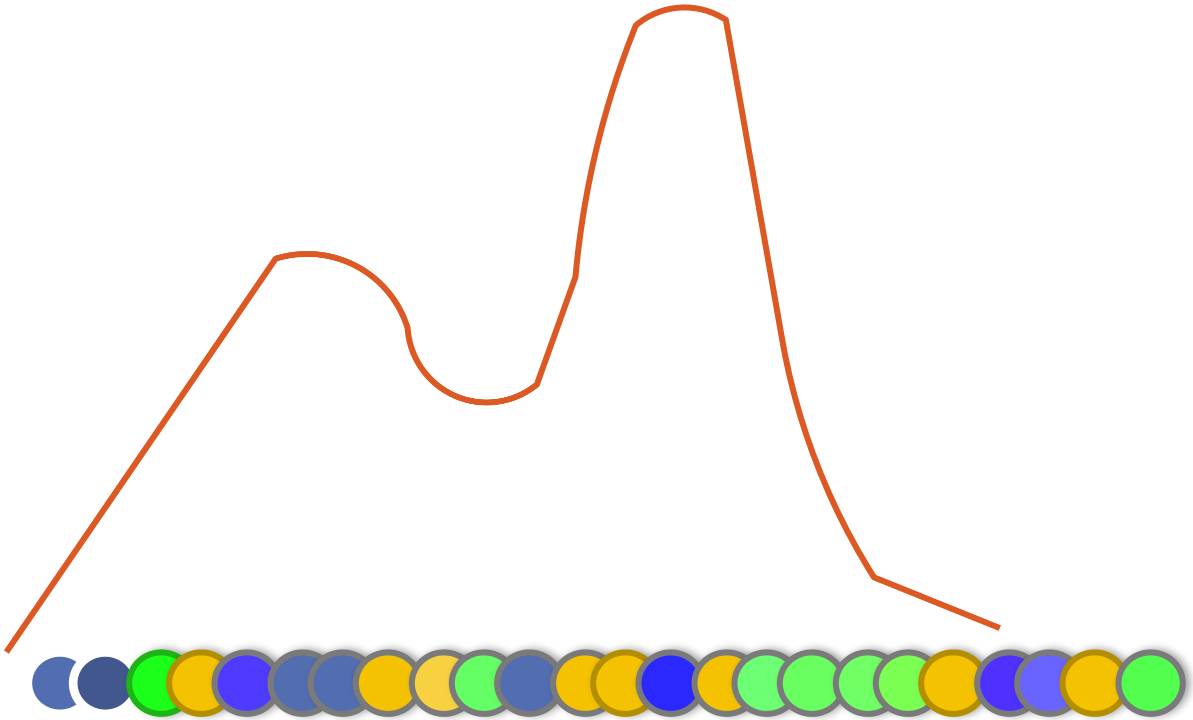


Gaussian Mixtures Model

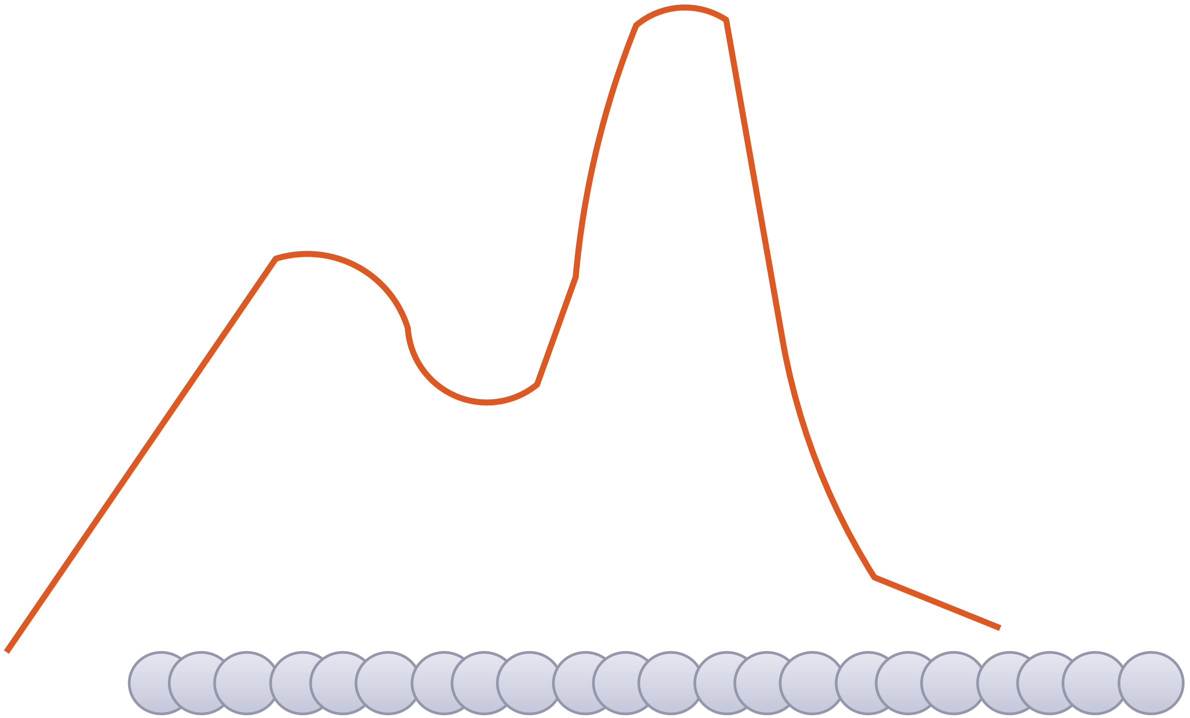


Gaussian Mixtures Model

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x \,|\, \mu_k, \Sigma_k)$$



Gaussian Mixtures Model



Gaussian Mixtures Model

- We have a linear combination of several Gaussians

- Each Gaussian is a cluster, one of K clusters

- Each cluster has a mean and covariance

- Mixing probability,

Gaussian Mixtures Model

Parameters - μ, Σ, π

$$\sum_{k=1}^{K} \pi_k = 1 \quad ; \quad 0 \leq \pi_k \leq 1$$

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k)$$

$$\mathcal{N}(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} exp\left\{ -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right\}$$

x is a D dimensional vector

μ is a D-dimensinal mean vector

Σ is a D x D covariance matrix

Maximum Likelihood Estimate

$$\mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

$$ln\mathcal{N}(x|\mu,\Sigma) = -\frac{D}{2}ln\,2\pi - \frac{1}{2}ln\,\Sigma - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)$$

Once Optimal values of the parameters are found,

the solution will correspond to the Maximum Likelihood Estimate (MLE)

Maximum Likelihood Estimate

For a point, x_i, let the cluster to which that point belongs be labeled z_k

values of z_k, satisfy

$$z_k \in \{0,1\} \qquad \sum_k z_k = 1$$

z is a K-dimensional binary random variable having 1-of-K representation,

A particular element z_k is equal to 1 and all other elements are equal to 0

$$p(z) = \prod_{k=1}^{K} \pi_k^{z_k}$$

Maximum Likelihood Estimate

The conditional distribution of x, given a particular value for z,

is a Gaussian

$$p(x|z_k = 1) = \mathcal{N}(x|\mu_k, \Sigma_k)$$

$$p(x|z) = \prod_{k=1}^{K} \mathcal{N}(x|\mu_k, \Sigma_k)^{z_k}$$

Our goal: what is the probability of z given our observation x?

$$p(z|x)?$$

Maximum Likelihood Estimate

Our goal: what is the probability of z given our observation x?

The joint distribution, $P(x, z)$, is given by $p(z)p(x \mid z)$

The marginal distribution of x, is obtained by summing the joint distribution over all possible states of z, to give

$$p(x) = \sum_{z} p(z)p(x \mid z) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k)$$

It means that, for every observed data point x_i, there is a corresponding latent variable z_i

Maximum Likelihood Estimate

$$p(x) = \sum_{z} p(z)p(x\,|\,z) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x\,|\,\mu_k, \Sigma_k)$$

Conditional probability of, z_k given x_k, is represented by Bayes' theorem

$$p(z_k = 1\,|\,x) = \frac{p(z_k = 1)p(x\,|\,z_k = 1)}{\sum_{i=1}^{K} p(z_i = 1)p(x\,|\,z_i = 1)}$$

$$= \frac{\pi_k \mathcal{N}(x\,|\,\mu_k, \Sigma_k)}{\sum_{i=1}^{K} \pi_i \mathcal{N}(x\,|\,\mu_i, \Sigma_i)}$$

π_k is the prior probability of $z_k = 1$

Maximum Likelihood Estimate

$$p(x) = \sum_z p(z)p(x|z) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$$

Conditional probability of, z_k given x_k, is represented by Bayes' theorem

$$p(z_k = 1|x) = \frac{p(z_k = 1)p(x|z_k = 1)}{\sum_{i=1}^{K} p(z_i = 1)p(x|z_i = 1)}$$

$$= \frac{\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)}{\sum_{i=1}^{K} \pi_i \mathcal{N}(x|\mu_i, \Sigma_i)}$$

π_k is the prior probability of $z_k = 1$

$p(z_k = 1|x)$ is the posterior probability

Maximum Likelihood Estimate

$$p(x) = \sum_z p(z)p(x|z) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$$

Data set of observations $\{x_1, \ldots, x_N\}$

X is an $N \times D$ matrix

Z is an $N \times K$ matrix of latent variables

Assumption: Data points are drawn independently from the distribution

$$p(X|\pi, \mu, \Sigma) = \prod_{i=1}^{n} p(x_i) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)$$

The log likelihood is given by:

$$ln\, p(X|\pi, \mu, \Sigma) = \sum_{i=1}^{n} ln \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)$$

Maximum Likelihood Estimate

The log likelihood is given by:

$$ln\, p(X|\pi,\mu,\Sigma) = \sum_{i=1}^{n} ln \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k,\Sigma_k)$$

$$\mathcal{N}(x_i|\mu_k,\Sigma_k) = \frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma_k|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)\right\}$$

Expectation Maximization

For lack of a closed form solution

$$ln\, p(X \mid \pi, \mu, \Sigma) = \sum_{i=1}^{n} ln \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)$$

We will use an iterative technique

$Step1$ – Choose some initial values for the means, covariances and mixing coefficients, evaluate log likelihood

$Step2$

E-step: Use current values for the parameters to evaluate the posterior probabilities

$$\gamma(z_{ik}) = p(z_k = 1 \mid x_i) = \frac{p(z_k = 1)p(x_i \mid z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(x_i \mid z_j = 1)}$$

$$= \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i \mid \mu_j, \Sigma_j)}$$

Expectation Maximization

$Step 3$

M-step: re-estimate means, covariances and mixing coefficients

$$\mu_k^{new} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_{ik}) x_i$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_{ik})(x_i - \mu_k^{new})(x_i - \mu_k^{new})^T$$

$$\pi_k^{new} = \frac{N_k}{N} \qquad \text{where } N_k = \sum_{i=1}^{N} \gamma(z_{ik})$$

$Step 4$

– Evaluate the log likelihood

$$ln\, p(X \mid \pi, \mu, \Sigma) = \sum_{i=1}^{n} ln \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)$$

Expectation Maximization

$Step1$ – Choose some initial values for the means, covariances and mixing coefficients, evaluate log likelihood

mean - pick a random value between minimum and maximum data value , twice
mean1 and mean2

std - pick a random value, may be just 1

mixing-coeff - pick equal values - 0.5 for two clusters

$Step2$

E-step: Use current values for the parameters to evaluate the posterior probabilities

$$\gamma(z_{ik}) = p(z_k = 1 \mid x_i) = \frac{p(z_k = 1)p(x_i \mid z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(x_i \mid z_j = 1)}$$

$$= \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i \mid \mu_j, \Sigma_j)}$$

Expectation Maximization

Step2

E-step: Use current values for the parameters to evaluate the posterior probabilities

$$\gamma(z_{ik}) = p(z_k = 1 \mid x_i) = \frac{p(z_k = 1)p(x_i \mid z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(x_i \mid z_j = 1)}$$

$$= \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i \mid \mu_j, \Sigma_j)}$$

For each data point - find pdf value for each distribution

$$\mathcal{N}(x_i \mid \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma_k|^{\frac{1}{2}}} exp\left\{ -\frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right\}$$

For two clusters - we will get two values for each data point γ_1, γ_2

$$\gamma_1 = \frac{\gamma_1}{\gamma_1 + \gamma_2} \qquad \gamma_2 = \frac{\gamma_2}{\gamma_1 + \gamma_2}$$

Expectation Maximization

(γ_1, γ_2)

$Step3$

M-step: re-estimate means, covariances and mixing coefficients

$$\mu_k^{new} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_{ik}) x_i$$

N_k – sum of weights (probabilities) – $\gamma_1' s, \gamma_2' s$

where $N_k = \sum_{i=1}^{N} \gamma(z_{ik})$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_{ik})(x_i - \mu_k^{new})(x_i - \mu_k^{new})^T$$

$$\pi_k^{new} = \frac{N_k}{N}$$

N - length of data

$Step4$

– Evaluate the log likelihood

$$ln\, p(X \mid \pi, \mu, \Sigma) = \sum_{i=1}^{n} ln \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)$$

Expectation Maximization

```
Input Gaussian 1: μ = 1.5e+01, σ = 7.1
Input Gaussian 2: μ = 5.5, σ = 2.9
Gaussian 1: μ = 5.9, σ = 3.0, weight = 0.8
Gaussian 2: μ = 2e+01, σ = 4.4, weight = 0.2
```