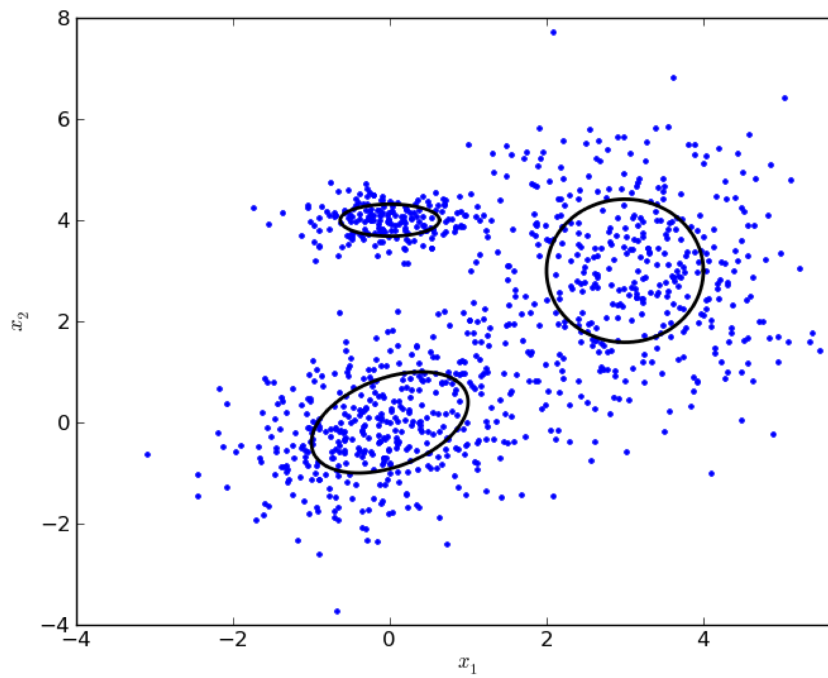
Symbols, Patterns and Signals: **Gaussian Mixture Methods**

Dr. Iván Palomares Carrascosa, office 319 Magg's House

Email: i.palomares@bristol.ac.uk

Gaussian mixture models

Approach to "soft" clustering where <u>each cluster is treated as a multivariate normal distribution with its own mean and covariance matrix</u>



Gaussian mixture models

Approach to "soft" clustering where <u>each cluster is treated as a multivariate normal distribution with its own mean and covariance matrix</u>

Would be easy if we knew which Gaussian does each data point come from, but then it would be a supervised classification problem (labelled)

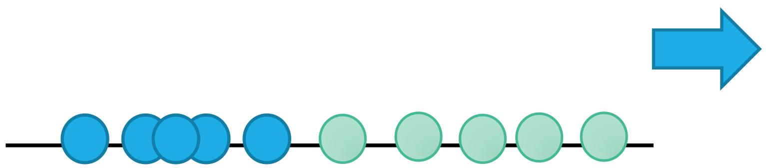
- In which we would use e.g. maximum-likelihood estimation rules

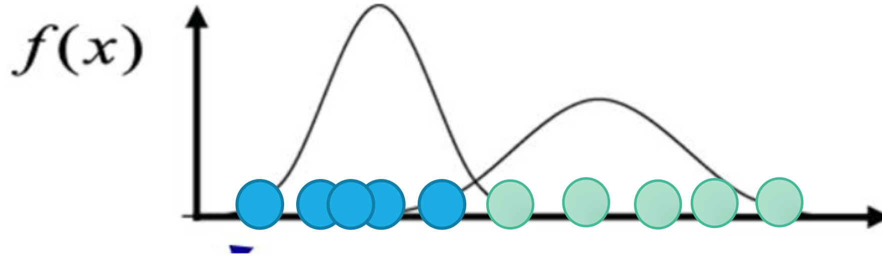
Idea: Treat cluster membership *(K clusters)* as continuous hidden variable

- Each observation $\boldsymbol{x}_i$ has a latent variable: $z_i = (z_{i1},\ldots,z_{iK})$
- **K**-means is special case: **0–1** cluster membership
- We will use a general algorithm called <u>**Expectation-Maximisation (EM)**</u>

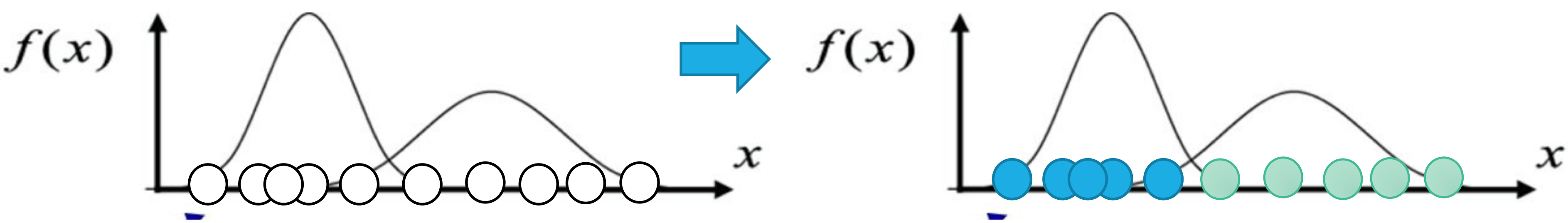
A chicken-and-egg problem

if we had the **class labels (cluster membership)**, we could find the model parameters (e.g. mean, covariance r

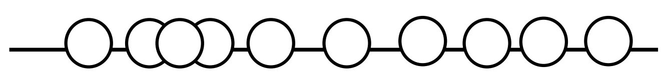




… and if we had the **model parameters**, we could compute the labels (using ML estimation).



But what if we do not have none of them? It's **a chicken-and-egg problem**!



EM Algorithm

Solution: let's iterate Expectation and Maximisation –> **EM algorithm**

- Powerful modelling approach under both hidden variables or missing data
- An iterative algorithm → "soft" version of K-means

Define

$\boldsymbol{\mu_j(t)}$: estimate of μ_j after the $\mathbf{t}^{th}$ iteration
$\mathbf{z_{ij}(t)}$: estimate of $\mathbf{z_{ij}}$ after the $\mathbf{t}^{th}$ iteration

E-step: assign expected value to hidden variables
M-step: re-estimate model parameters

EM for 1-D Gaussian mixtures

Given: data x_i ($1 \le i \le n$) drawn from K normal distributions with unknown means and equal variance

- i.e. variance doesn't influence the outcome and can be set to 1

Obtain: estimates of the means $\mu_1 \ldots \mu_K$

Approach: introduce hidden variables z_{ij} indicating the likelihood that x_i came from the *j-th* Gaussian

- E-step: for each data point $\mathbf{x_i}$ and each **j**

$$z_{ij}(t) = E[z_{ij} \mid x_i, \mu_j(t)] \propto e^{-(x_i - \mu_j(t))^2/2}, \text{ normalised such that } \sum_{j=1}^{K} z_{ij}(t) = 1$$

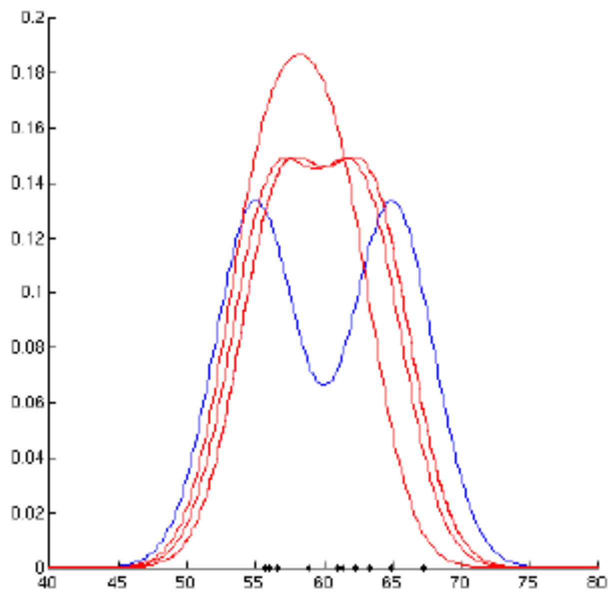
- M-step: for each **j**, estimate mean as weighted average

$$\mu_j(t+1) = \arg\max_{\mu} p(x_1 \ldots x_n, z_{1j} \ldots z_{nj} \mid \mu) = \ldots = \sum_{i=1}^{n} z_{ij}(t) x_i \Big/ \sum_{i=1}^{n} z_{ij}(t)$$

1-D EM example

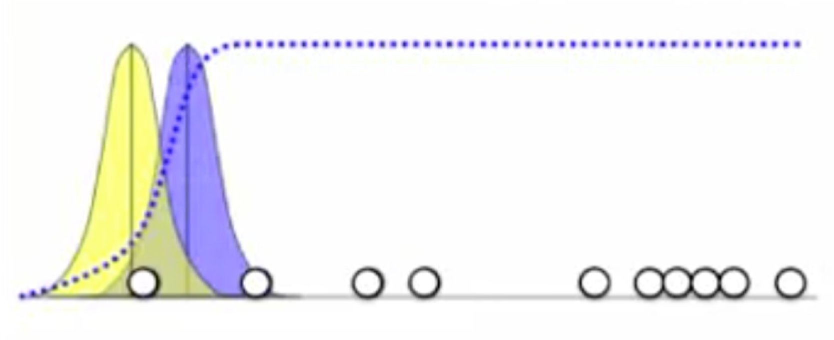
	x_i	55.6951	56.0631	56.5929	58.8639	61.0000	61.4035	62.2644	63.3310	64.9595	67.2668		
	z_{i1}											μ_1	40
	z_{i2}											μ_2	70
t=1	z_{i1}	1.0000	1.0000	0.9997	0.0372	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	μ_1	55.6951
	z_{i2}	0.0000	0.0000	0.0003	0.9628	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	μ_2	60.7440
t=2	z_{i1}	1.0000	1.0000	1.0000	0.9794	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	μ_1	56.1507
	z_{i2}	0.0000	0.0000	0.0000	0.0206	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	μ_2	62.7474
t=3	z_{i1}	1.0000	1.0000	1.0000	0.9996	0.0023	0.0002	0.0000	0.0000	0.0000	0.0000	μ_1	56.7931
	z_{i2}	0.0000	0.0000	0.0000	0.0004	0.9977	0.9998	1.0000	1.0000	1.0000	1.0000	μ_2	63.3554
t=4	z_{i1}	1.0000	1.0000	1.0000	0.9997	0.0025	0.0002	0.0000	0.0000	0.0000	0.0000	μ_1	56.8062
	z_{i2}	0.0000	0.0000	0.0000	0.0003	0.9975	0.9998	1.0000	1.0000	1.0000	1.0000	μ_2	63.3716

1-D EM example



x_i	55.6951	56.0631	56.5929	58.8639	61.0000	61.4035	62.2644	63.3310	64.9595	67.2668		
z_{i1}											μ_1	40
z_{i2}											μ_2	70
z_{i1}	1.0000	1.0000	0.9997	0.0372	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	μ_1	55.6951
z_{i2}	0.0000	0.0000	0.0003	0.9628	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	μ_2	60.7440
z_{i1}	1.0000	1.0000	1.0000	0.9794	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	μ_1	56.1507
z_{i2}	0.0000	0.0000	0.0000	0.0206	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	μ_2	62.7474
z_{i1}	1.0000	1.0000	1.0000	0.9996	0.0023	0.0002	0.0000	0.0000	0.0000	0.0000	μ_1	56.7931
z_{i2}	0.0000	0.0000	0.0000	0.0004	0.9977	0.9998	1.0000	1.0000	1.0000	1.0000	μ_2	63.3554
z_{i1}	1.0000	1.0000	1.0000	0.9997	0.0025	0.0002	0.0000	0.0000	0.0000	0.0000	μ_1	56.8062
z_{i2}	0.0000	0.0000	0.0000	0.0003	0.9975	0.9998	1.0000	1.0000	1.0000	1.0000	μ_2	63.3716

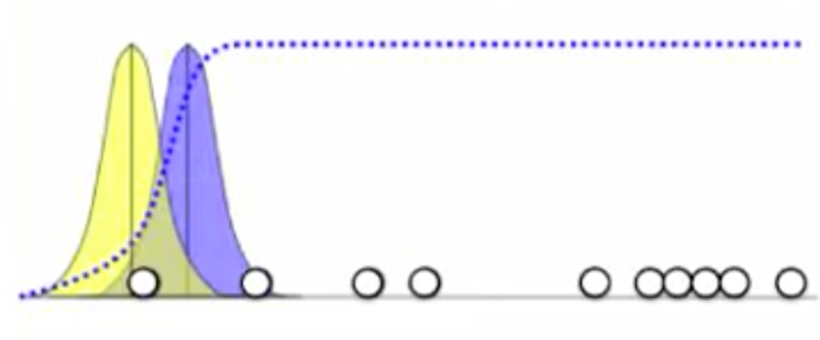
1-D EM: a visual example



INITIALIZATION (K=2)
Two centroids (means of gaussian distributions) are initialized randomly

CREDITS: Check out this EM video-tutorial: https://youtu.be/iQoXFmbXRJA

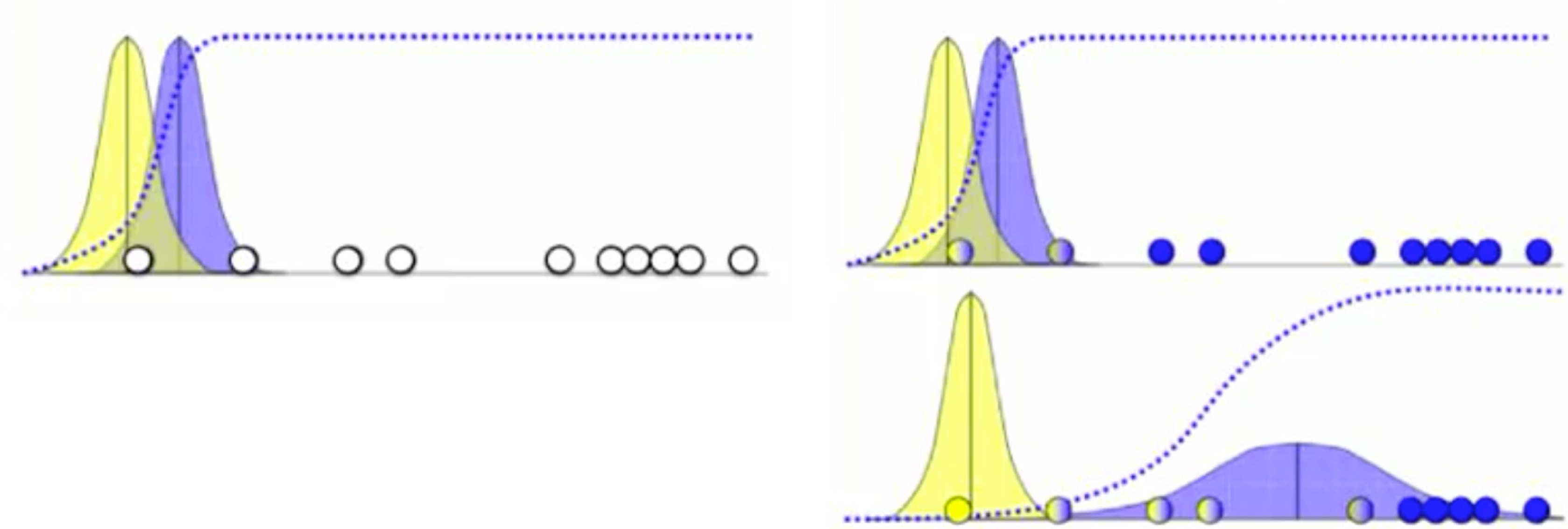
1-D EM: a visual example



FIRST ITERATION – E-STEP
How likely is each data point to belong to each of the two Gaussians?

CREDITS: Check out this EM video-tutorial: https://youtu.be/iQoXFmbXRJA

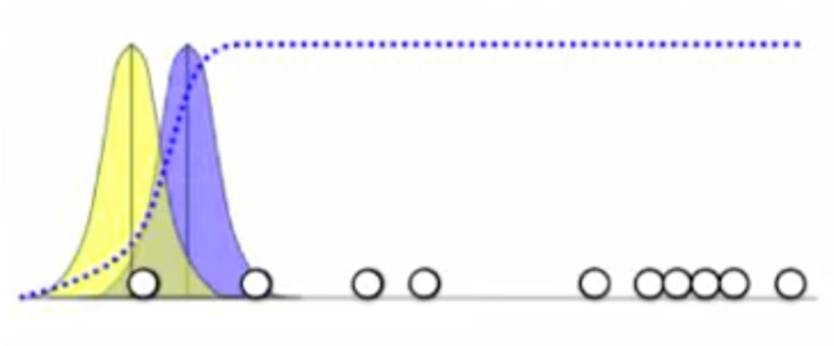
1-D EM: a visual example

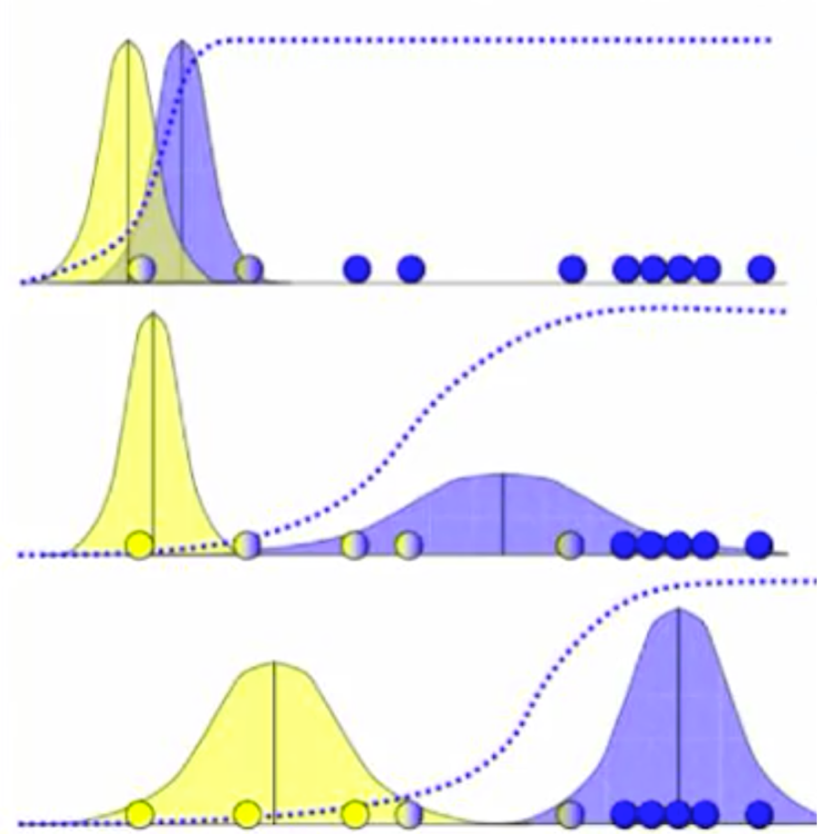


FIRST ITERATION – M-STEP
Recomputing the parameters of each Gaussian distribution.
NOTE: This example also estimates variances – not only means.

CREDITS: Check out this EM video-tutorial: https://youtu.be/iQoXFmbXRJA

1-D EM: a visual example



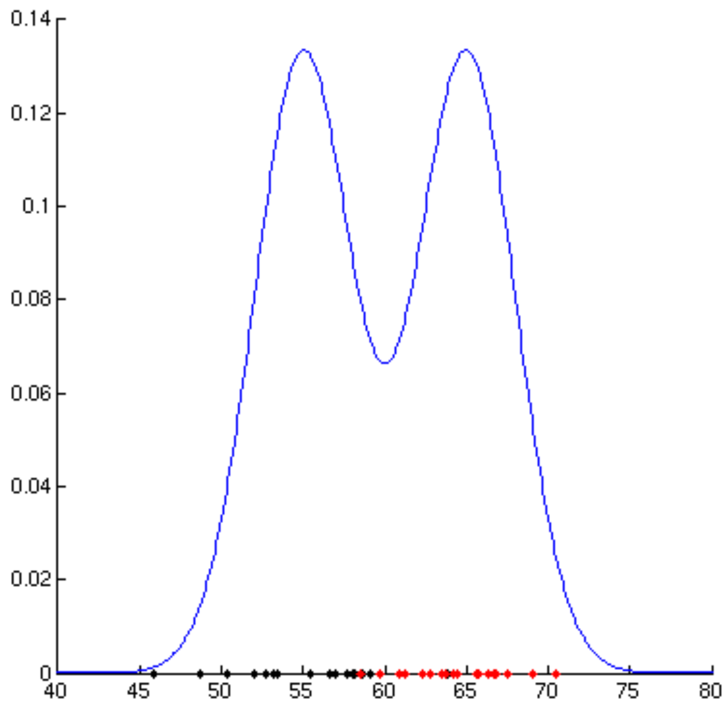


SECOND ITERATION – E-STEP

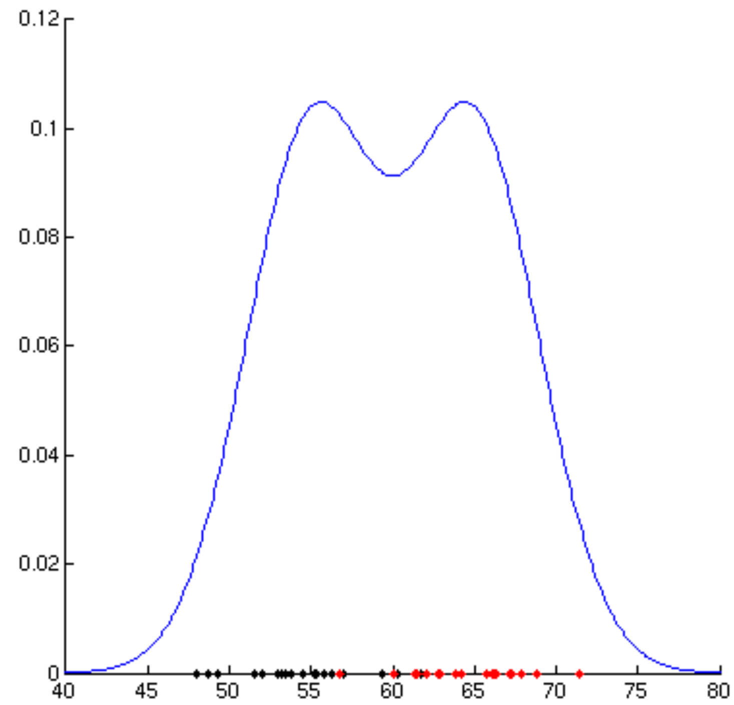
CREDITS: Check out this EM video-tutorial: https://youtu.be/iQoXFmbXRJA

1-D Gaussian mixtures

- μ_1=55, μ_2=65, σ_1=σ_2=3
- some overlap between clusters



- μ_1=55, μ_2=65, σ_1=σ_2=4
- more overlap between clusters



Discussion

If the Gaussians have equal variance, Gaussian mixture models are very similar to *K*-means

- main difference: ‘soft’ cluster membership

But GMMs are more general: can also estimate covariances

- The usual caveats apply: local maxima, dependence on initial configuration, etc.
- **Expectation-Maximization is a very general technique for estimating hidden variables**

Practice yourself: Problem sheet – exercise 9