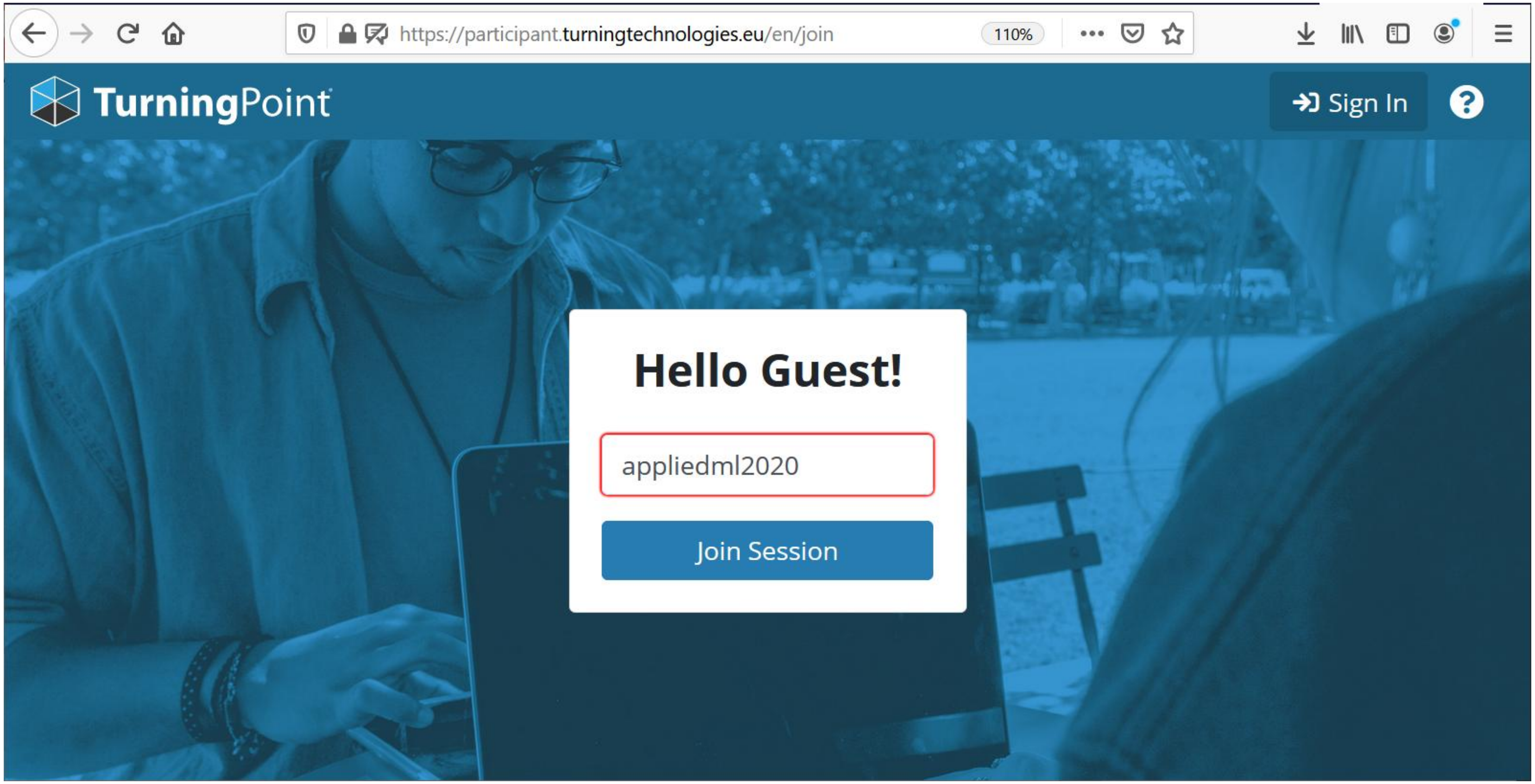
Pdf, GMM and E-M

Interactive Lecture

Launch polling system

https://participant.turningtechnologies.eu/en/join

Acces as GUEST and enter the session id: *appliedml2020*



Linear Correlation

Two variables x_1, x_2 are correlated if:

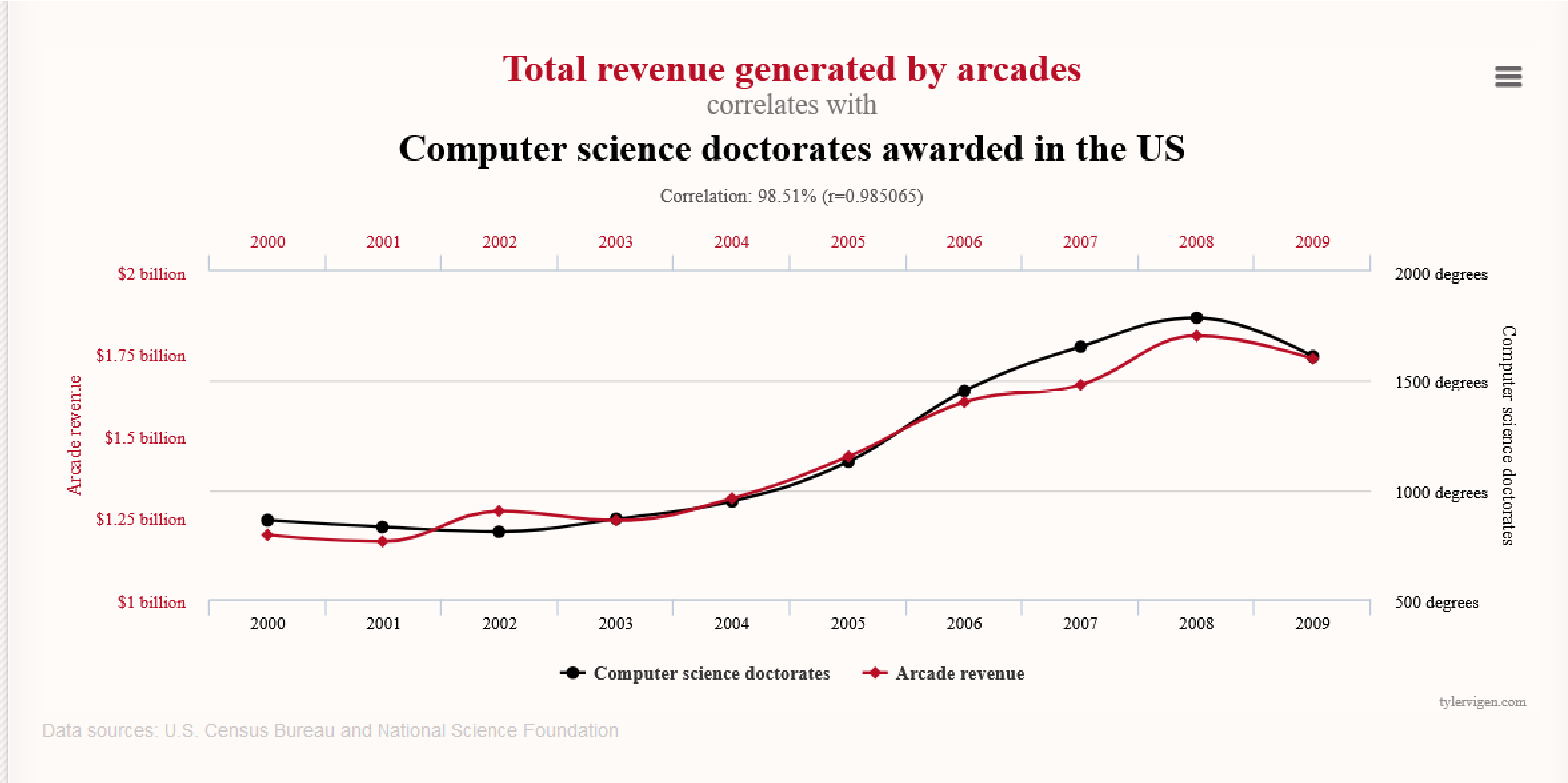
$$corr\left(x_1, x_2\right) = \frac{\operatorname{cov}\left(x_1, x_2\right)}{\operatorname{var}\left(x_1\right)\operatorname{var}\left(x_2\right)} \neq 0$$

$corr\left(x_1, x_2\right) >< 0$: positive / negative correlation

$\left|corr\left(x_1, x_2\right)\right| = 1$: perfectly correlated

$\left|corr\left(x_1, x_2\right)\right| < 0.5$ weakly correlated

Real and Spurious Correlations



Spurious Correlations, T. Vygen

Statistical Independence and uncorrelatedness

Are x,y correlated?
Are x,y statist. dependent?

Uncorrelated: $E\{x,y\} = E\{x\} E\{y\}$

Statistical Ind.: $p(x,y) = p(x)\, p(y)$

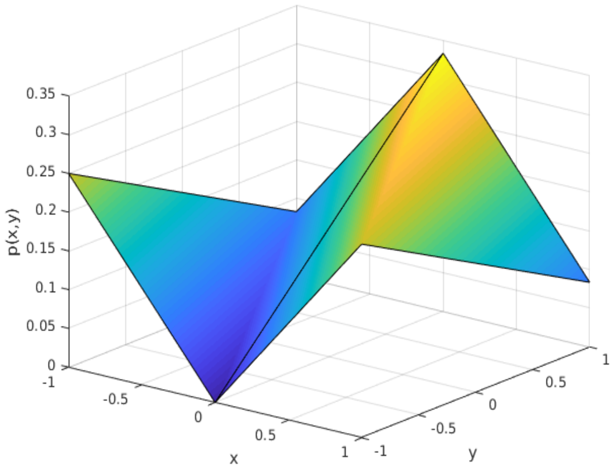
Statistical Independence and uncorrelatedness

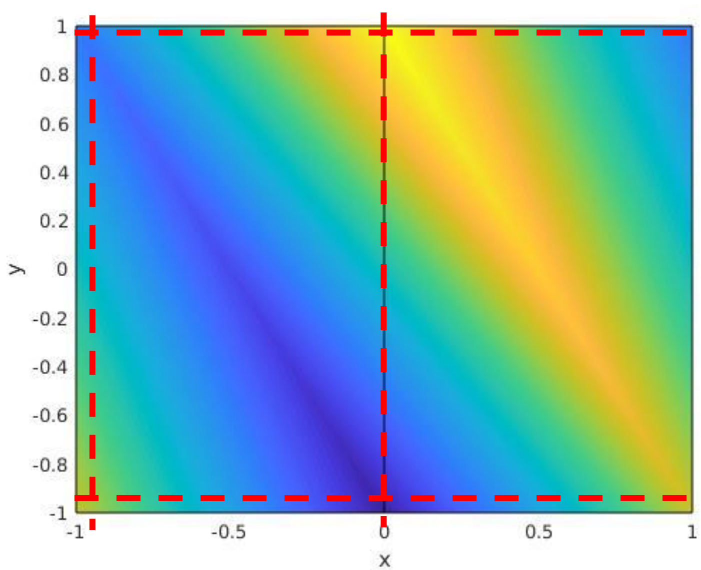
Are x,y correlated?
Are x,y statist. dependent?

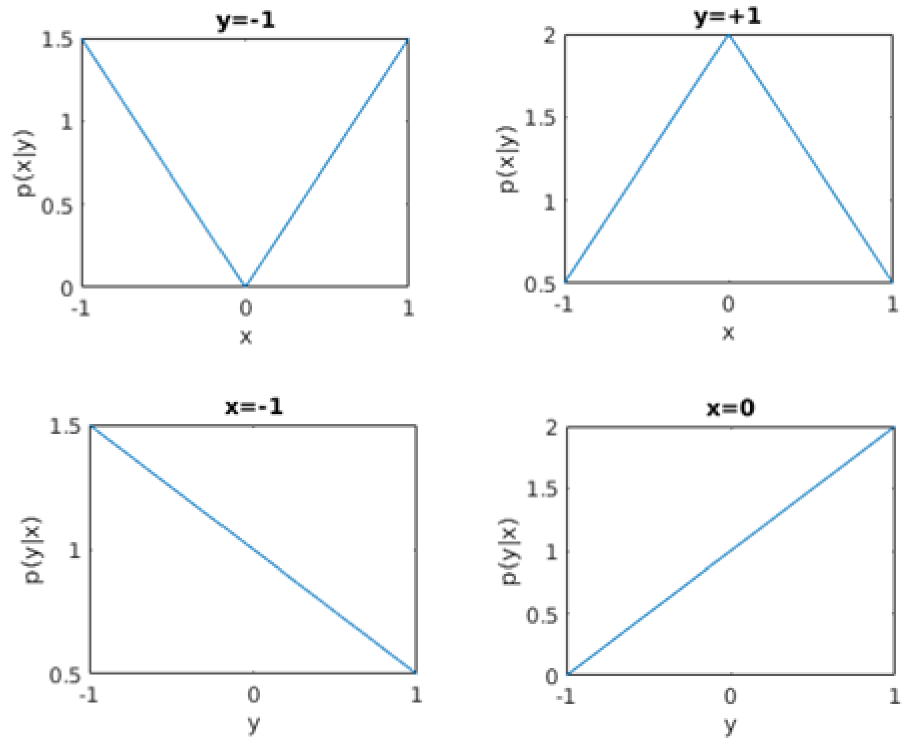
Uncorrelated: $E\{x, y\} = E\{x\} E\{y\}$

Statistical Ind.: $p(x, y) = p(x) p(y)$

$$E\{x, y\} = E\{x\} E\{y\} = 0$$







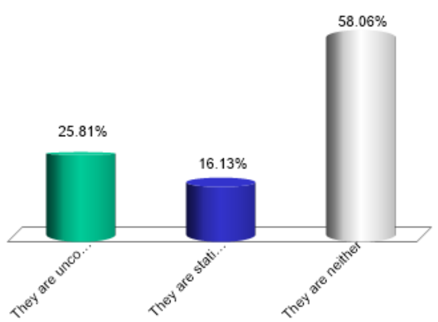
Statistical Independence and uncorrelatedness

Joint probabilities over two variables x_1, x_2

	x_2=-1	x_2=0	x_2=1	Total
x_1=-1	3/12	0	3/12	**1/2**
x_1=1	1/12	4/12	1/12	**1/2**
Total	**1/3**	**1/3**	**1/3**	

Are x_1 and x_2 uncorrelated and statistically independent?

A. They are uncorrelated
B. They are statistically independent
C. They are neither



Statistical Independence and uncorrelatedness

Joint probabilities over two variables x_1, x_2

	x_2=-1	x_2=0	x_2=1	Total
x_1=-1	3/12	0	3/12	**1/2**
x_1=1	1/12	4/12	1/12	**1/2**
Total	**1/3**	**1/3**	**1/3**	

$$E\left\{x_1, x_2\right\} = E\left\{x_1\right\} E\left\{x_2\right\}$$

$$\sum_{i,j=1}^{3} \left(x_1 = i, x_2 = j\right) p\left(\left(x_1 = i\right), \left(x_2 = j\right)\right) = \underbrace{\sum_{i=1}^{3} \left(x_1 = i\right) p\left(\left(x_1 = i\right)\right)}_{0} \underbrace{\sum_{i=1}^{3} \left(x_2 = i\right) p\left(\left(x_2 = i\right)\right)}_{0}$$

Both sums are zero $\Rightarrow$ x_1, x_2 : uncorrelated

$$p\left(x_1 = -1, x_2 = 1\right) = 3/12 = 0.25$$

$$\neq$$

$$p\left(x_1 = -1\right) p\left(x_2 = 1\right) = 1/2 * 1/3 = 0.1667$$

x_1 and x_2 are uncorrelated but not statistically independent.

Statistical Independence and uncorrelatedness

$$\textbf{Independent} \quad \substack{\longrightarrow \\ \not\longleftarrow} \quad \textbf{Uncorrelated}$$

$$p(x_1,x_2)=p(x_1)\,p(x_2) \quad \Rightarrow \quad E\{x_1,x_2\}=E\{x_1\}E\{x_2\}$$

$$p(x_1,x_2)=p(x_1)\,p(x_2) \quad \not\Leftarrow \quad E\{x_1,x_2\}=E\{x_1\}E\{x_2\}$$

Statistical independence ensures uncorrelatedness.
The converse is not true.

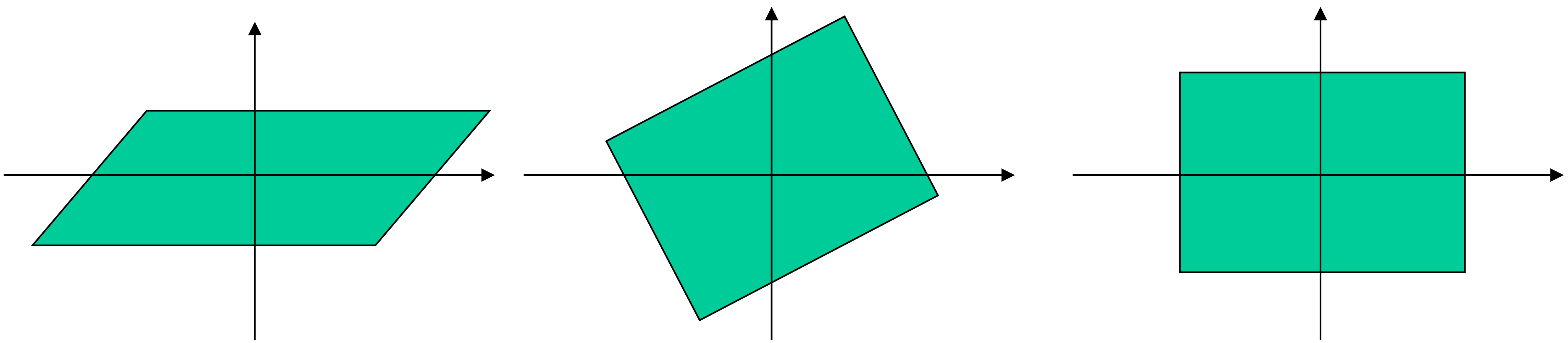
ICA: Preprocessing – Whitening & Independent Component Identification

Original Distribution

Uncorrelated distribution:

$$E\{x_1, x_2\} = E\{x_1\} E\{x_2\}$$

Statistically Indep. Distr.

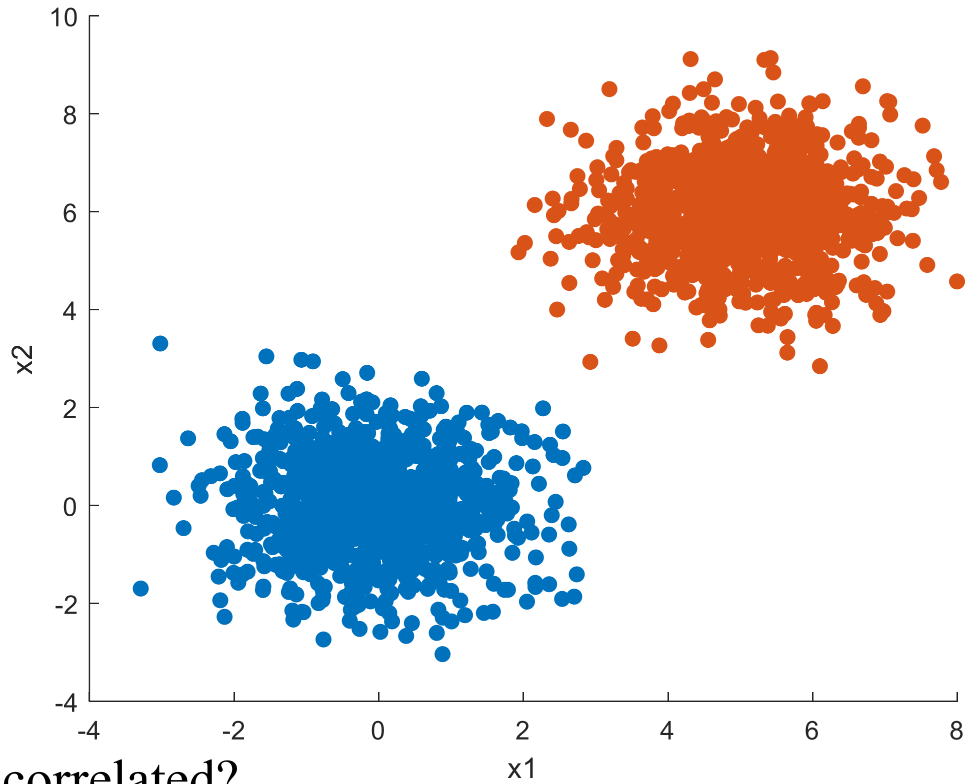


Whitening preprocessing:

$$E\{XX^T\} = I$$

After projection on independent components

:00

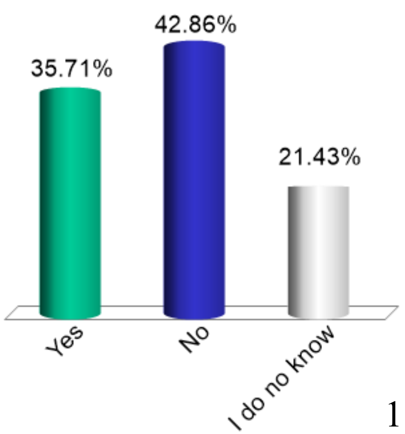


Are x_1 and x_2 correlated?

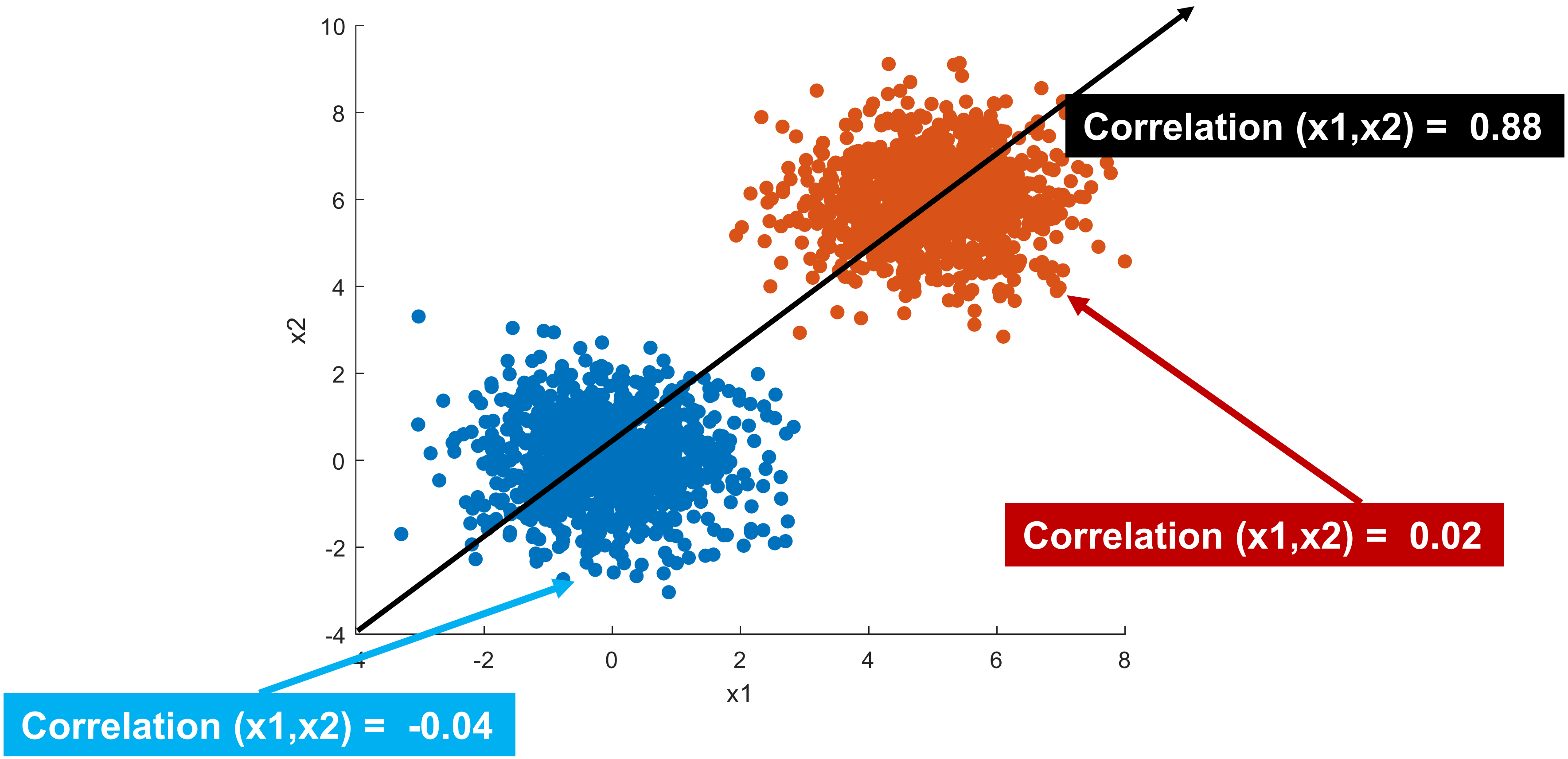
A. Yes

B. No

C. I do no know



Correlations

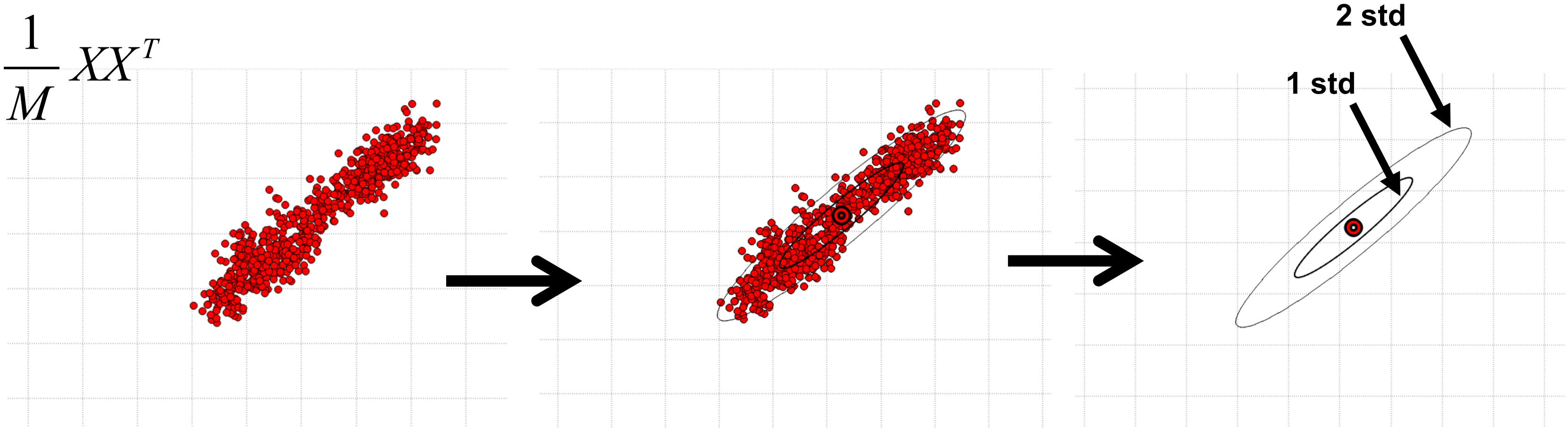


Spurious correlations as we compare two groups of data that come from two different distributions

Modeling Data with a Gaussian Function

Construct covariance matrix from (centered) set of datapoints $X = \left\{x^i\right\}^{i=1\ldots M}$:

$$\Sigma = \frac{1}{M} XX^T$$



$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} e^{\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)}$$

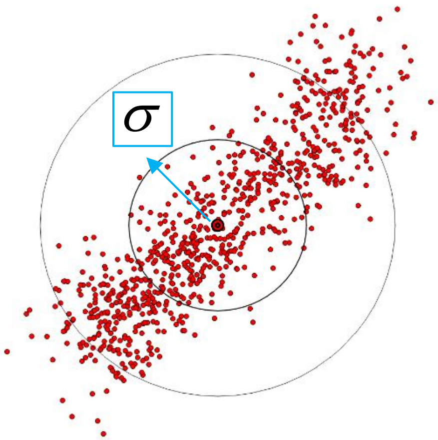
if x is N-dimensional, then

μ is a $N-$dimensional mean vector

Σ is a $N \times N$ covariance matrix

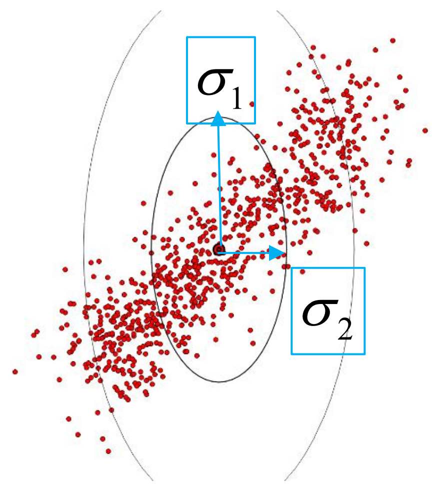
Which model?

$$p\left(x=\left[x_1\ x_2\right],;\mu,\Sigma\right)=\frac{1}{\left(2\pi\right)^{\frac{N}{2}}\left|\Sigma\right|^{\frac{1}{2}}}e^{\left(-\frac{1}{2}\left(x-\mu\right)^T\Sigma^{-1}\left(x-\mu\right)\right)}$$



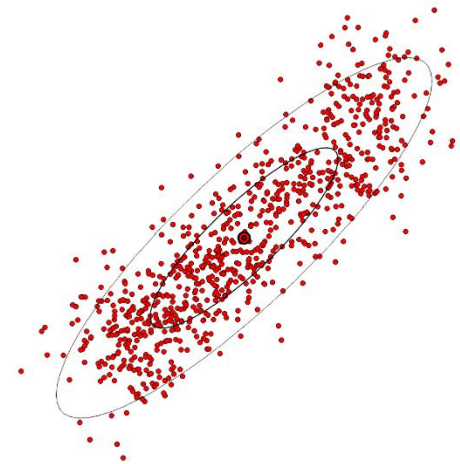
Spherical

$$\Sigma=\begin{bmatrix}\sigma & 0\\ 0 & \sigma\end{bmatrix}$$



Diagonal

$$\Sigma=\begin{bmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{bmatrix}$$



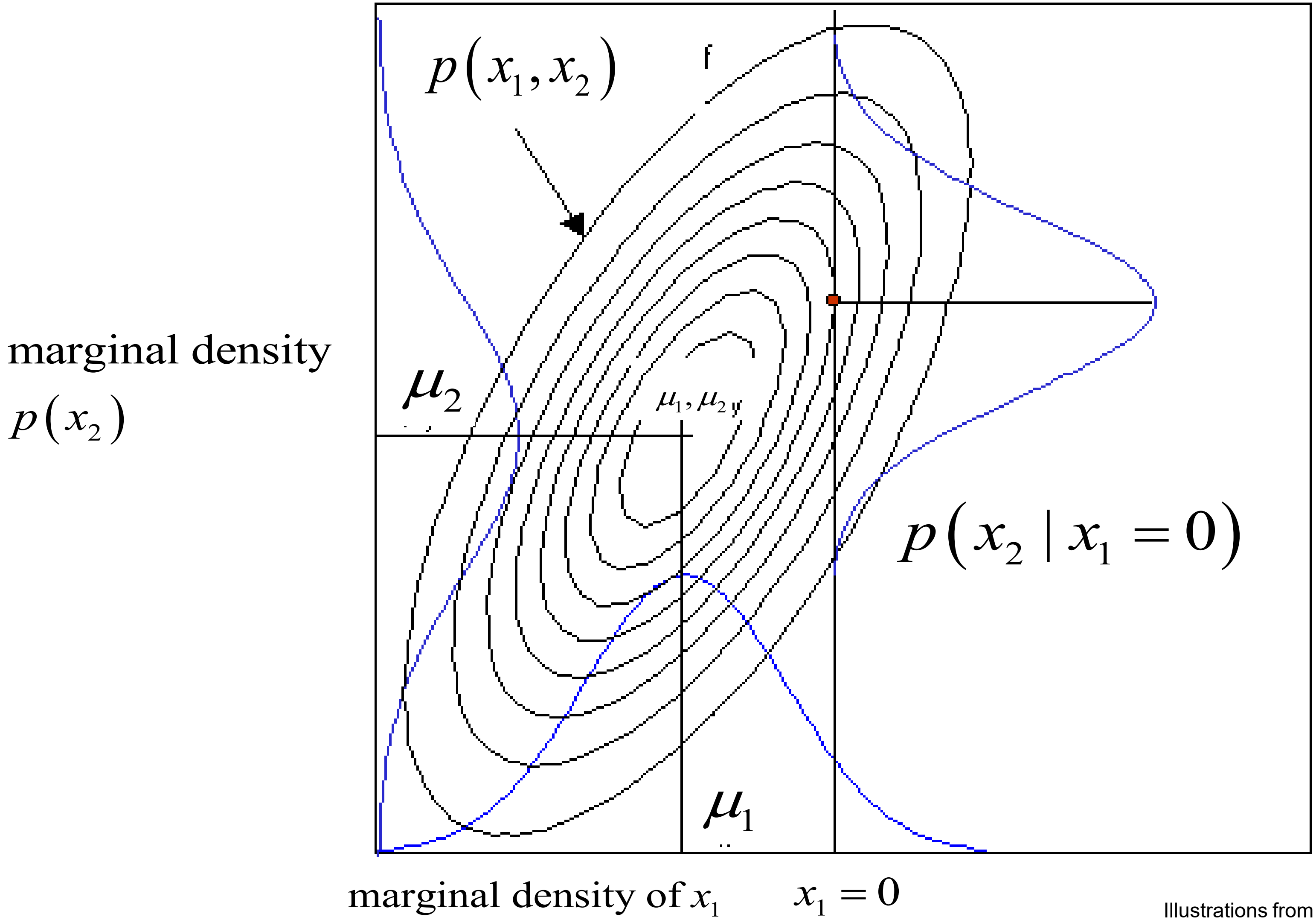
Full

$$\Sigma=\begin{bmatrix}\sigma_1 & \sigma_{12}\\ \sigma_{21} & \sigma_2\end{bmatrix}$$

Need a method to derive optimal parameters (mean and covariance matrix)

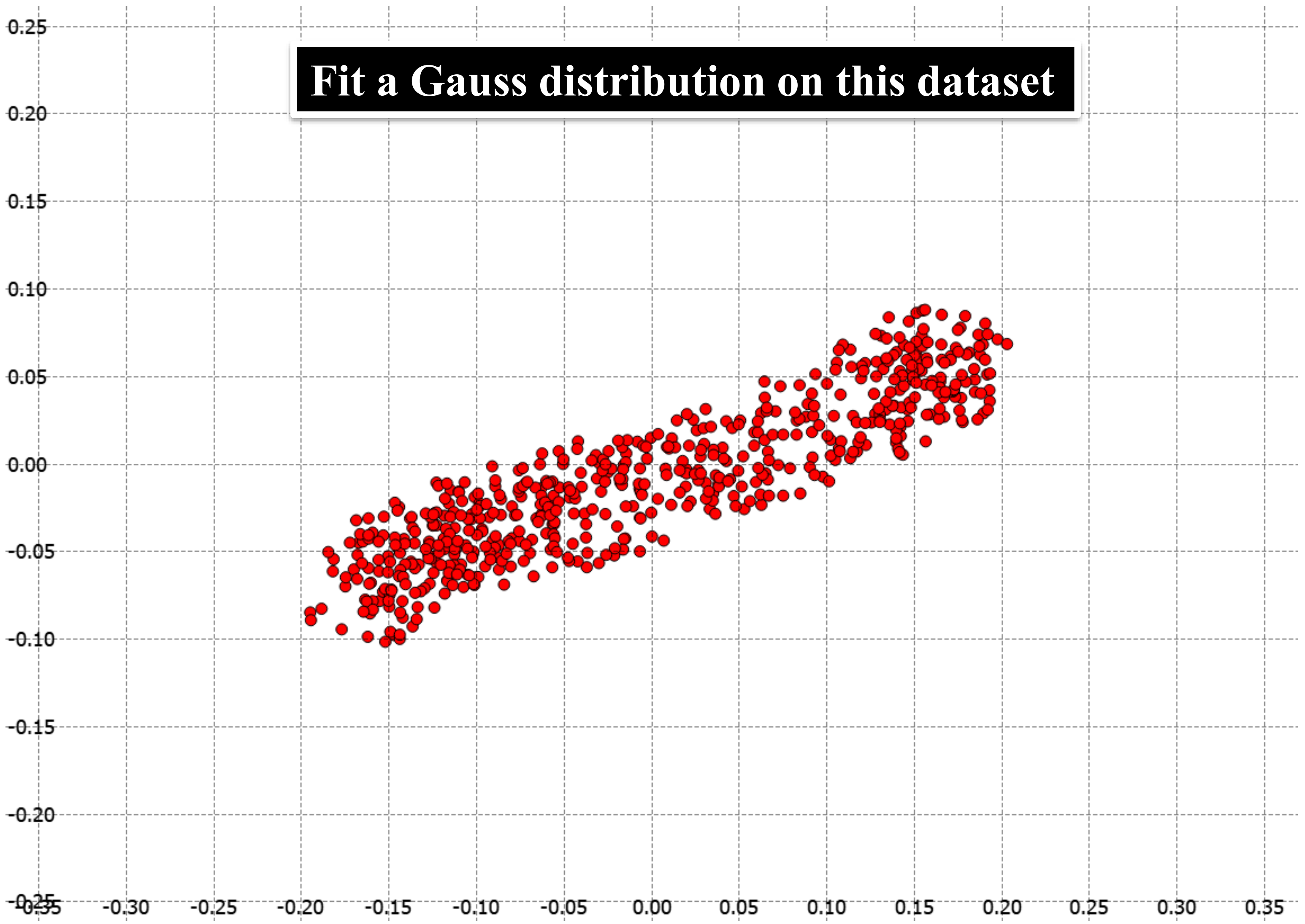
Marginal, Conditional Pdf of Gauss Functions

The conditional and marginal pdf of a multi-dimensional Gauss function are all Gauss functions!

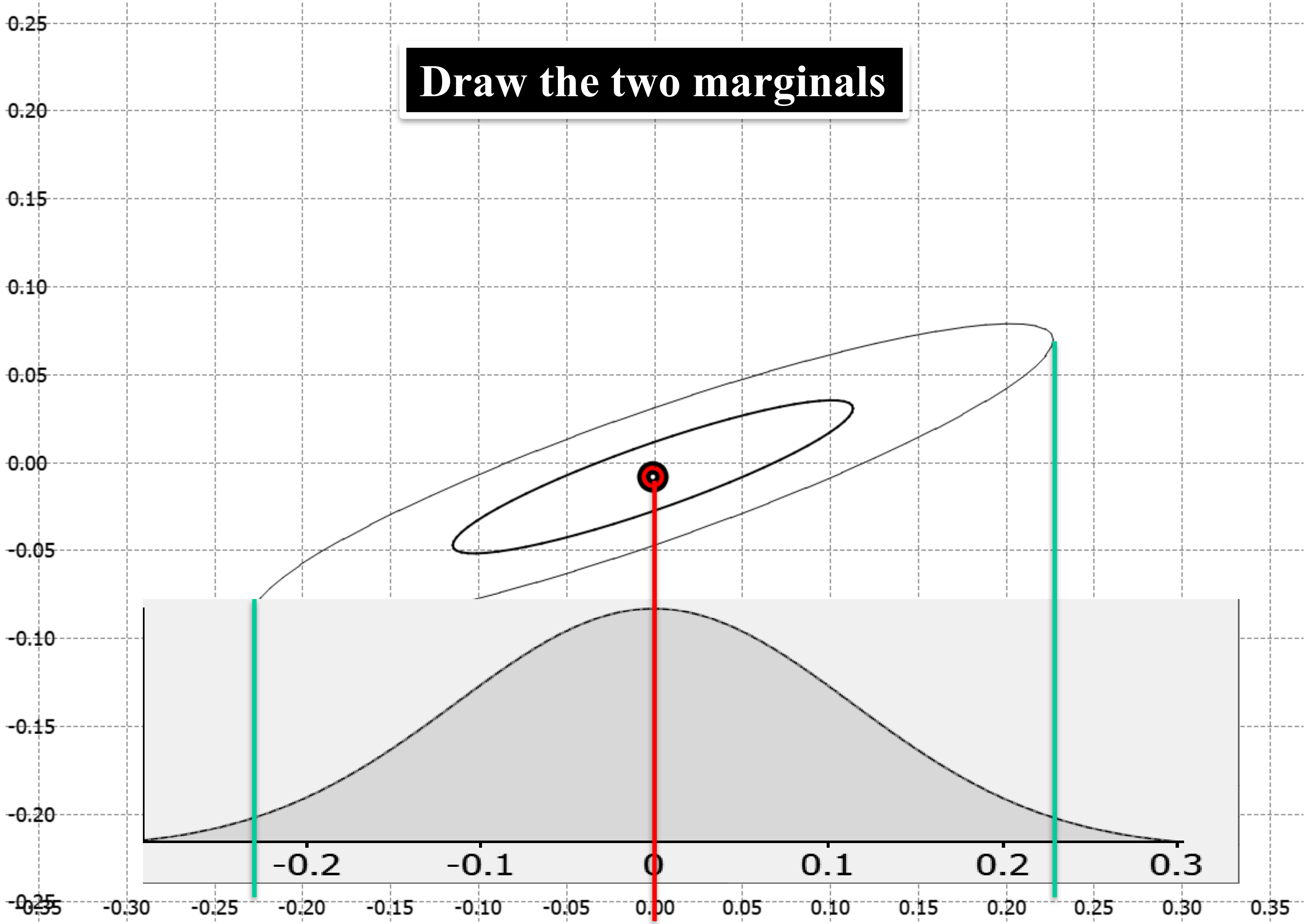


Illustrations from Wikipedia

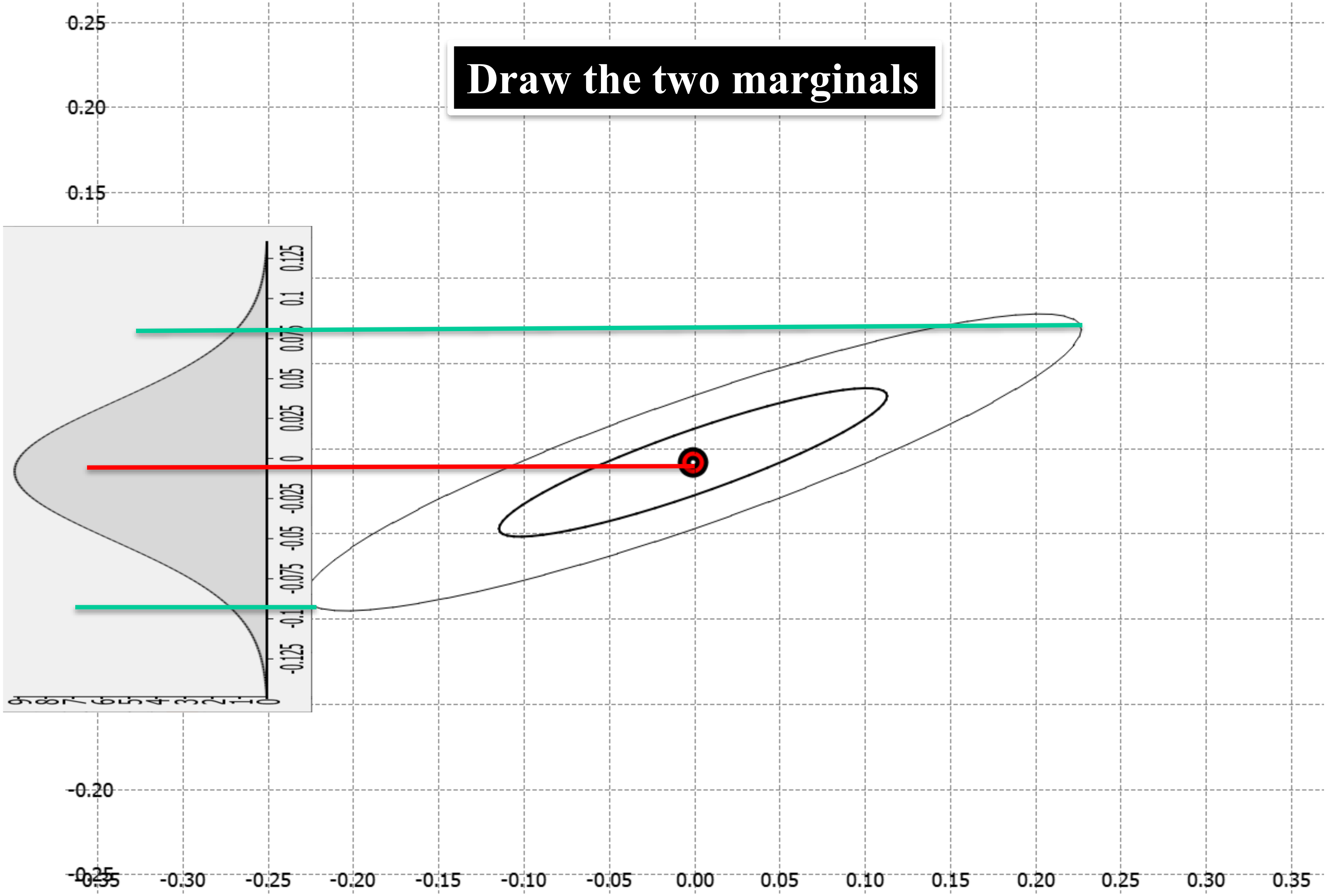
Marginal, Conditional Pdf of Gauss Functions



Marginal, Conditional Pdf of Gauss Functions

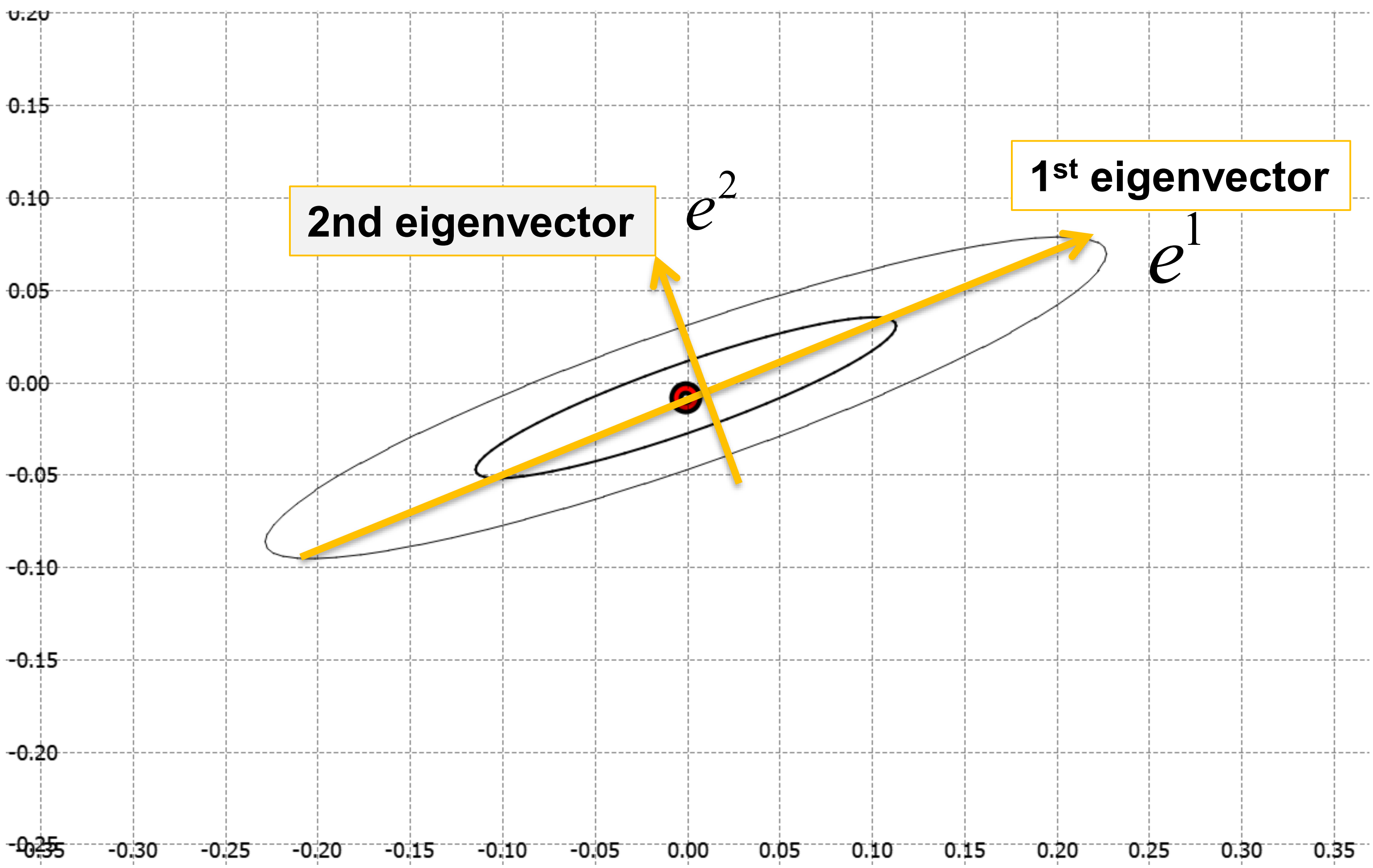


Marginal, Conditional Pdf of Gauss Functions



Marginal, Conditional Pdf of Gauss Functions

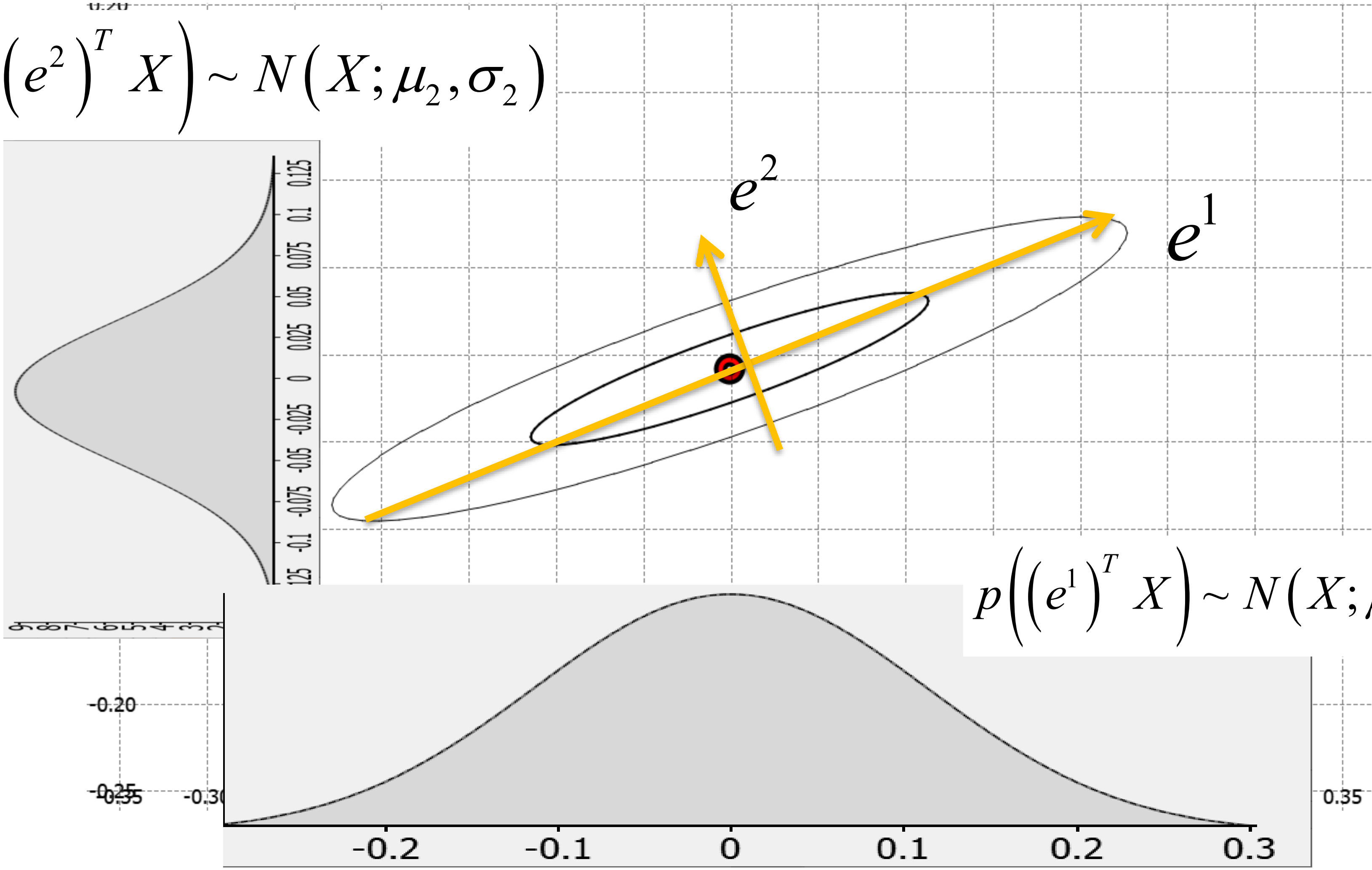
$$\Sigma \text{ eigenvalue decomposition: } \Sigma = V \Lambda V^T, V = \left[e^1 \; e^2 \right]$$



Marginal, Conditional Pdf of Gauss Functions

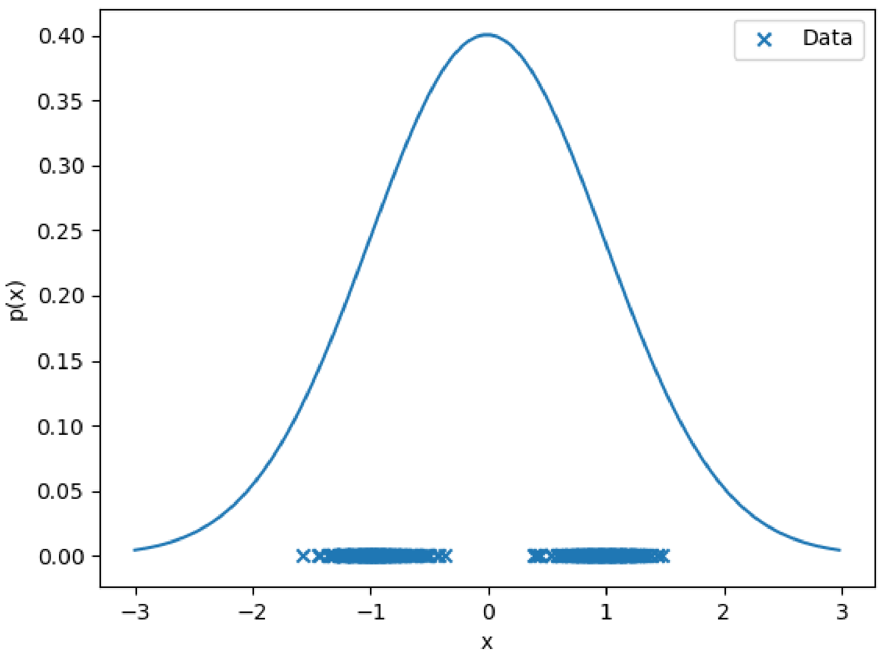
$$\Sigma \text{ eigenvalue decomposition: } \Sigma = V \Lambda V^T, V = \begin{bmatrix} e^1 & e^2 \end{bmatrix}$$

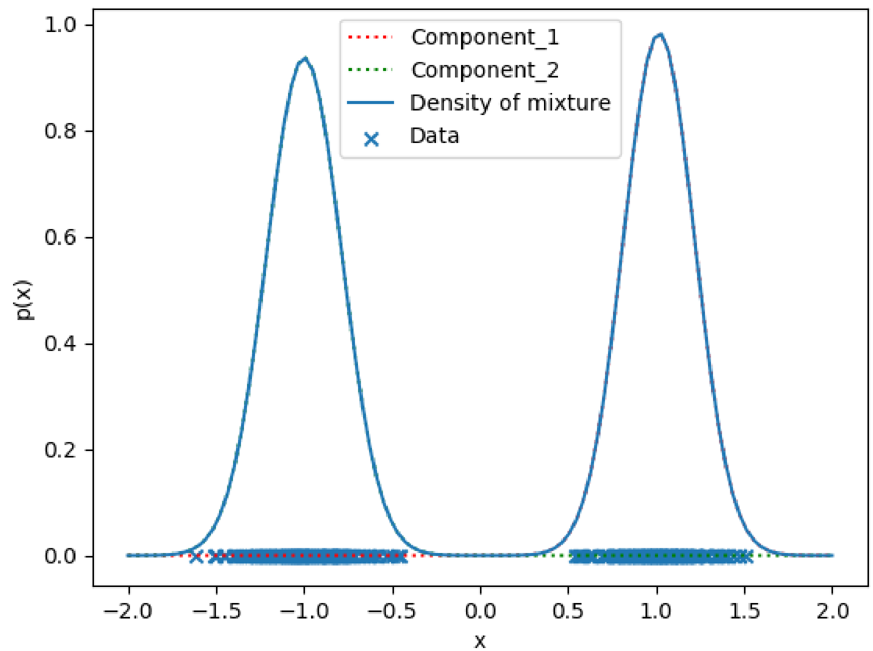
$$p\left(\left(e^2\right)^T X\right) \sim N\left(X; \mu_2, \sigma_2\right)$$



$$p\left(\left(e^1\right)^T X\right) \sim N\left(X; \mu_1, \sigma_1\right)$$

Mixture of Gauss Functions





Combination of K Gauss functions

$$p(x;\Theta) = \sum_{k=1}^{K} \alpha_k p_k\left(x;\mu^k,\Sigma^k\right) \text{ with } \Theta = \{\mu^1,\Sigma^1,\ldots\mu^K,\Sigma^K\},\ \alpha_k \in [0,1].$$

Linear weighted combination

K Gaussian Functions

EPFL

Training of GMM uses Expectation-Maximization

E-M searches for optimum of the likelihood of the model given the data, i.e.:

$$\max_{\Theta} \; L(\Theta \mid X) = \max_{\Theta} \; p(X \mid \Theta)$$

The parameters of a GMM are the means, covariance matrices and priors:

$$\Theta = \left\{ \mu^1, \ldots .. \mu^K, \Sigma^1, \ldots .. \Sigma^K, \alpha_1, \ldots .. \alpha_K \right\}$$

Computing the log of the likelihood yields the same optimum:

$$\max_{\Theta} \; p(X \mid \Theta) = \max_{\Theta} \; \log p(X \mid \Theta)$$

$$\max_{\Theta} \; \log \prod_{i=1}^{M} \sum_{k=1}^{K} \alpha_k \cdot p\left(x^i; \mu^k, \Sigma^k\right) = \max_{\Theta} \; \sum_{i=1}^{M} \log\left(\sum_{k=1}^{K} \alpha_k \cdot p\left(x^i; \mu^k, \Sigma^k\right) \right)$$

No closed-form solution → Solve through Expectation-Maximization (E-M)
E-M is an *iterative* procedure to estimate the best set of parameters
It converges to **a local optimum** → Sensitive to initialization!
See derivation of E-M for GMM in the annexes posted on the website

Likelihood

Likelihood for a single unnormalized Gauss function, given a set of M datapoints $\{x^i\}_{i=1}^{M}$

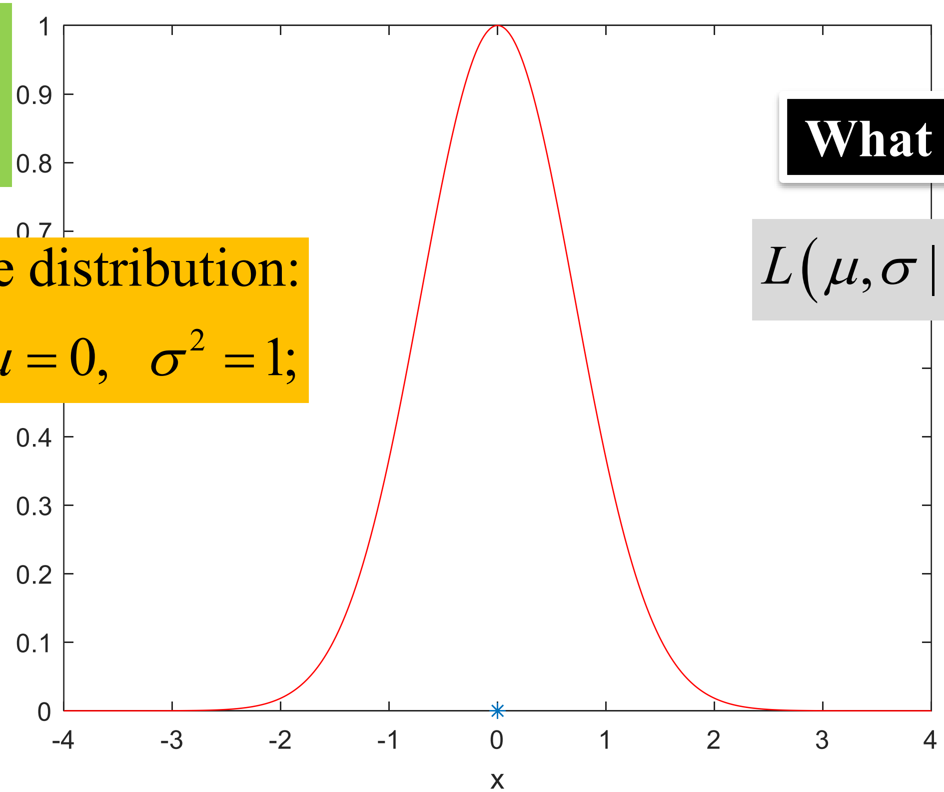
$$L(\mu,\sigma|X) = \prod_{i=1}^{M} p(x^i;\mu,\sigma) = e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

Dataset:

$X = [x^1],\ x^1 = 0;$

Sampled from the distribution:

$x \sim p(x;\mu,\sigma);\ \mu = 0,\ \ \sigma^2 = 1;$



What is the likelihood?

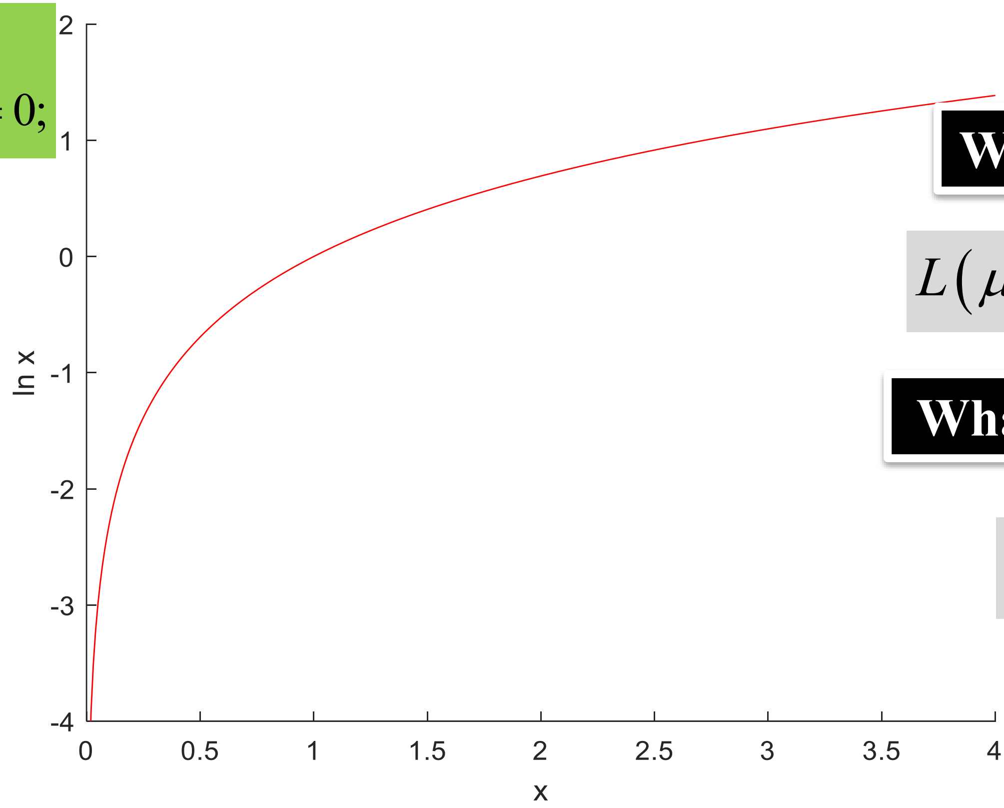
$L(\mu,\sigma \mid X) = p(x^1;\mu,\sigma) = 1$

Here: we consider an un-normalized Gauss function, also called Radial Basis Functions (RBF).

Likelihood

Dataset:

$X = \left[x^1 \right],\ x^1 = 0;$



What is the likelihood?

$$L\left(\mu,\sigma \mid X\right) = p\left(x^1;\mu,\sigma\right) = 1$$

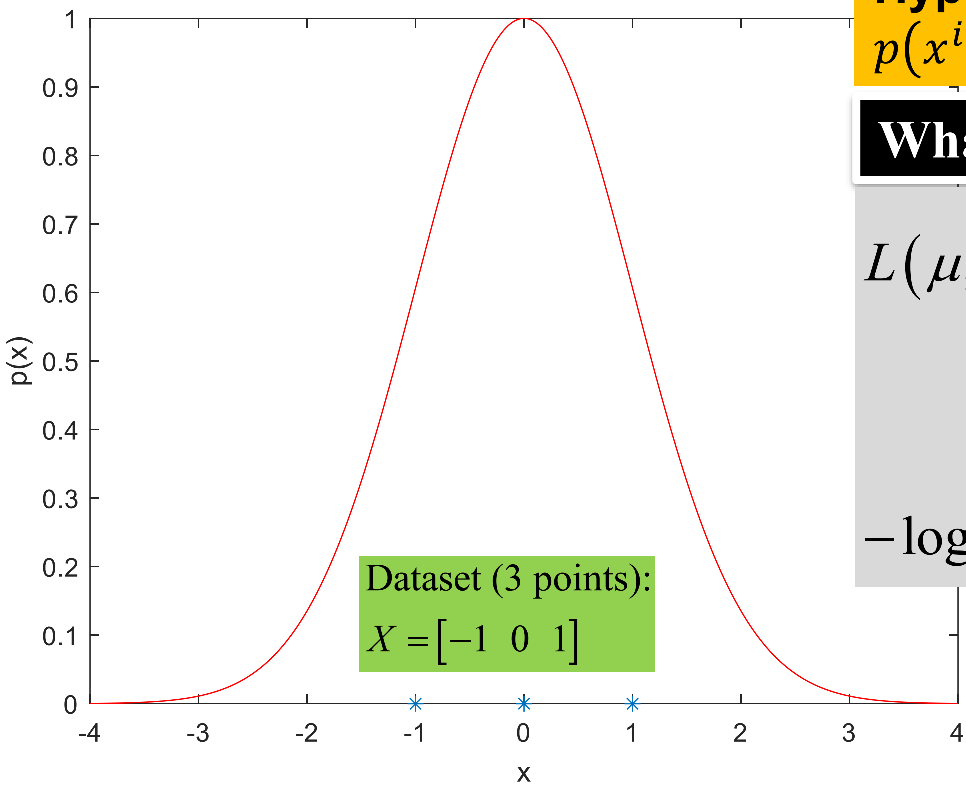
What is the log-likelihood?

$$\log p\left(x^1;\mu,\sigma\right) = 0$$

Likelihood

Likelihood for a single unnormalized Gauss function, given a set of M datapoints $\{x^i\}_{i=1}^{M}$

$$L(\mu,\sigma|X) = \prod_{i=1}^{M} p(x^i;\mu,\sigma) = e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

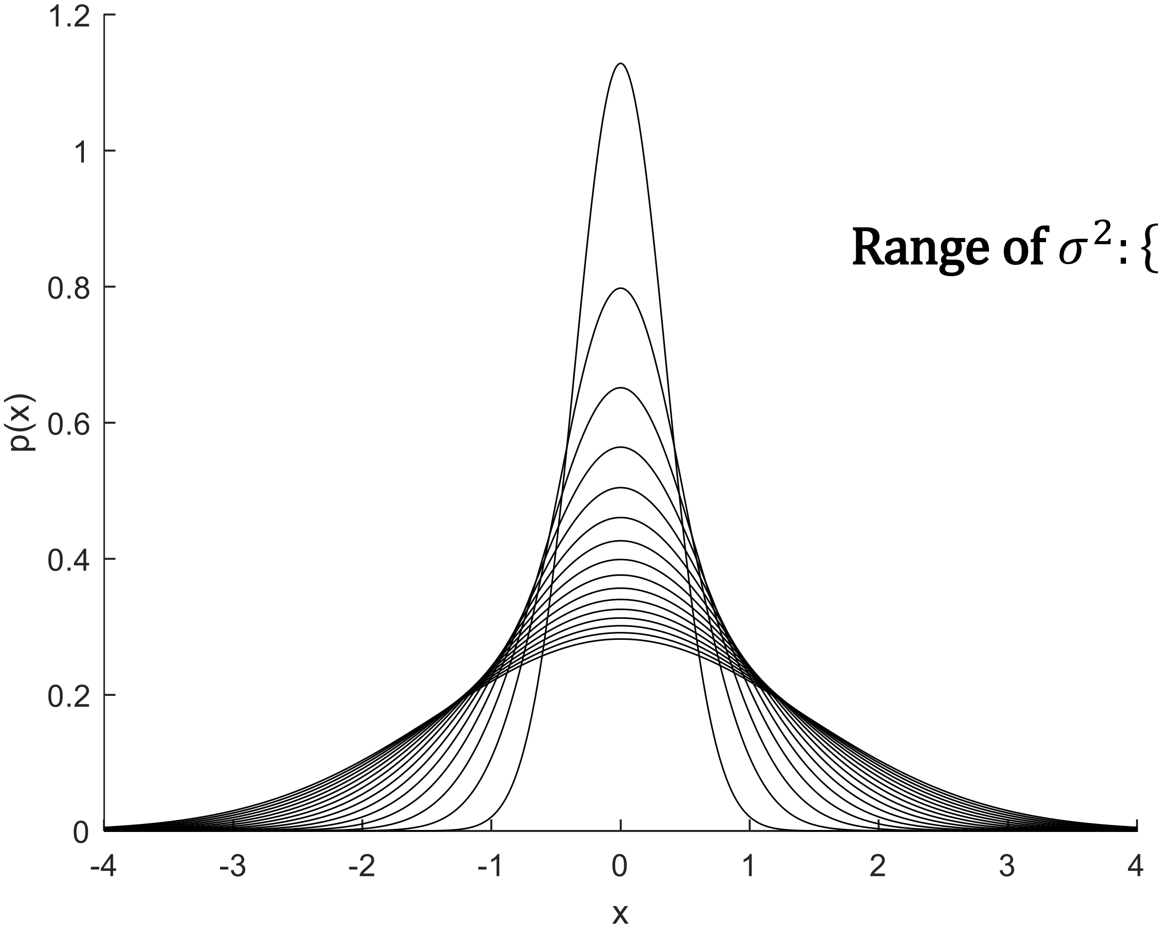


Hypothetical Model:
$p(x^i;\mu,\sigma)$; $\mu = 0$, $\sigma^2 = 2$;

What is the likelihood?

$$L(\mu,\sigma|X) = \prod_{i=1}^{3} p(x^i;\mu,\sigma)$$
$$= 0.8*1*0.8$$
$$= 0.6$$
$$-\log L(\mu,\sigma|X) = 0.5$$

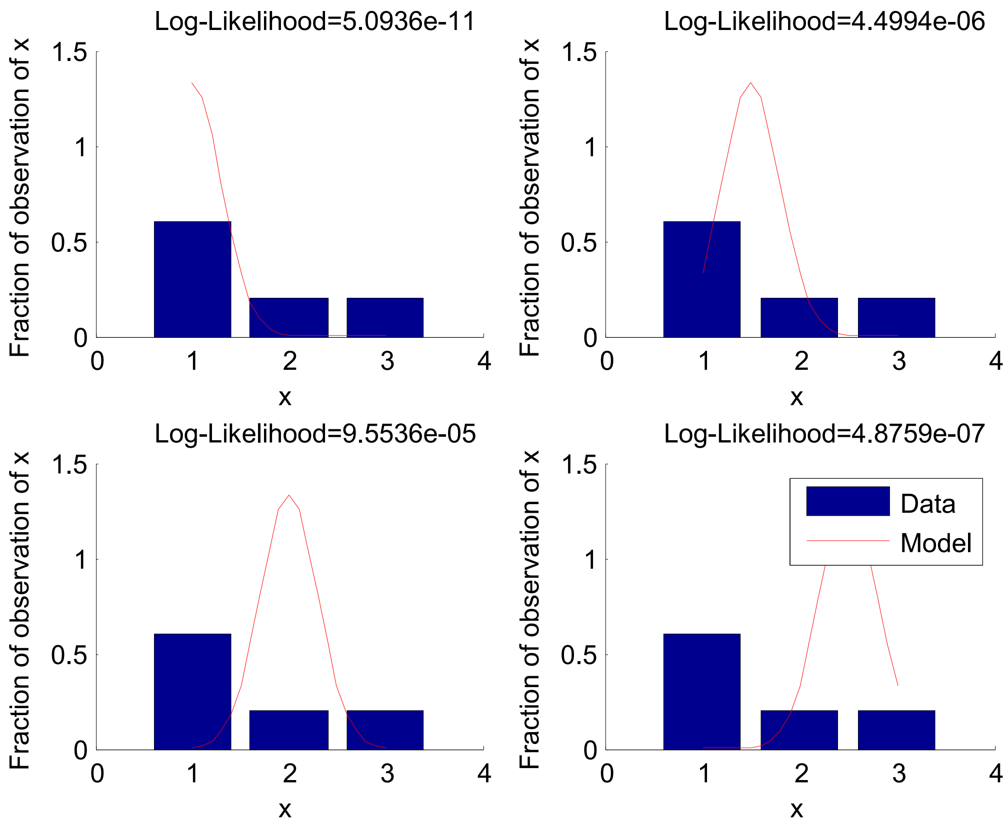
Gauss pdfs



Range of σ^2: $\{0.25, \ldots\ldots, 2\}$

Pdf for the normalized distributions

Likelihood Function for data not gauss distributed

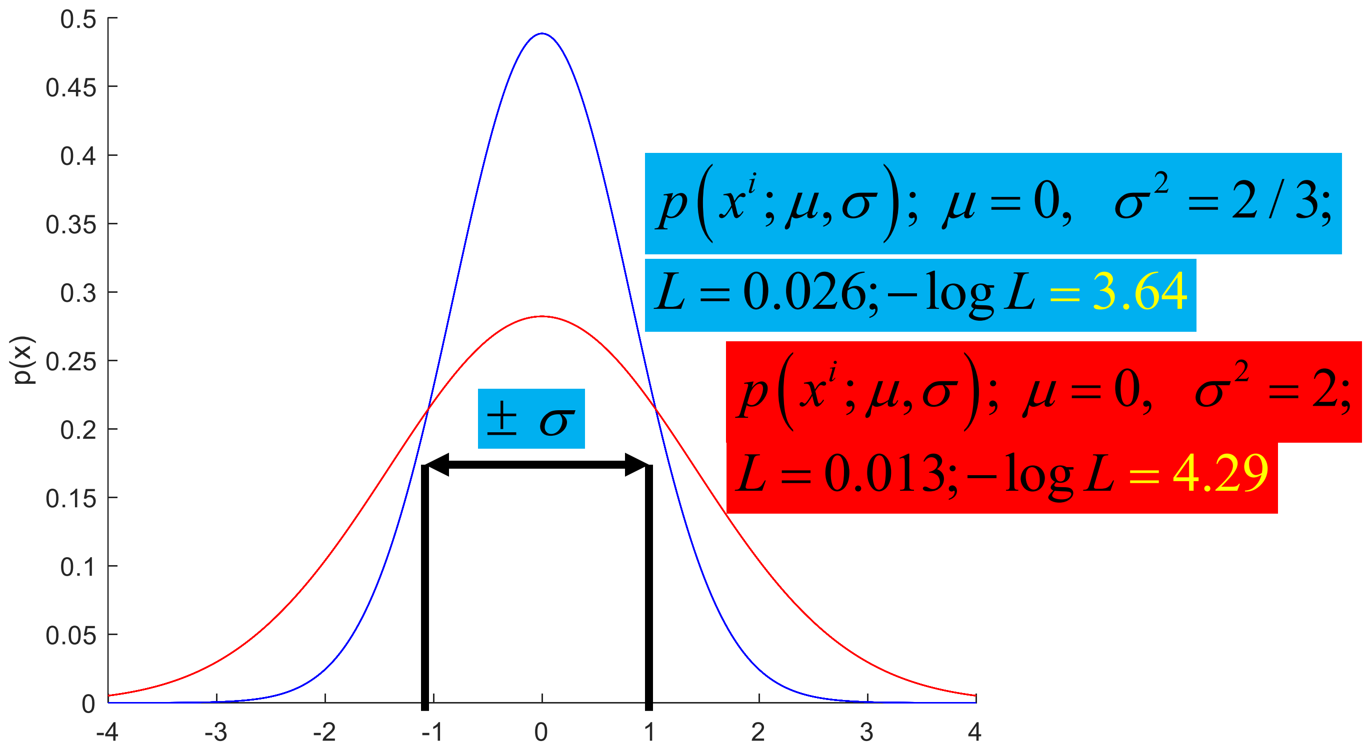


Log-Likelihood for a series of Gauss functions applied to datasets with pdfs that do not follow a Gauss distribution. The Likelihood increases as the fit is closer to the real mean of the data, even if this may appear as a poorer fit.

Maximum Likelihood

Likelihood for a single Gauss function, given a set of M datapoints $\{x^i\}_{i=1}^{M}$

$$L(\mu,\sigma \mid X) = \prod_{i=1}^{M} p\left(x^i;\mu,\sigma\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$



Pdf for the normalized distributions

Maximum Likelihood

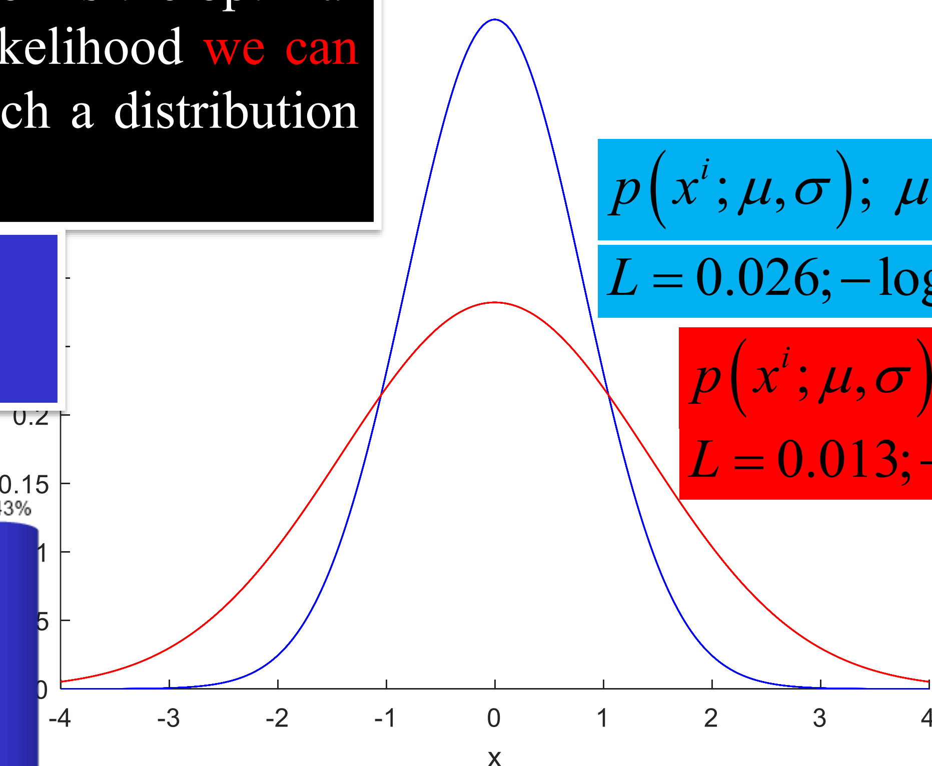
Likelihood for a single Gauss function, given a set of M datapoints $\left\{x^i\right\}_{i=1}^{M}$

$$L(\mu,\sigma \mid X)=\prod_{i=1}^{M} p\left(x^i;\mu,\sigma\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

The blue solution is the optimal value on the likelihood we can ever get for such a distribution of points.

A. True

B. False

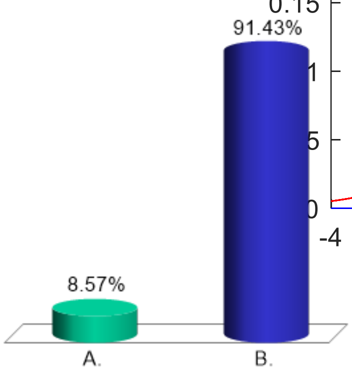


$p\left(x^i;\mu,\sigma\right);\ \mu=0,\ \ \sigma^2=2/3;$

$L=0.026; -\log L=3.64$

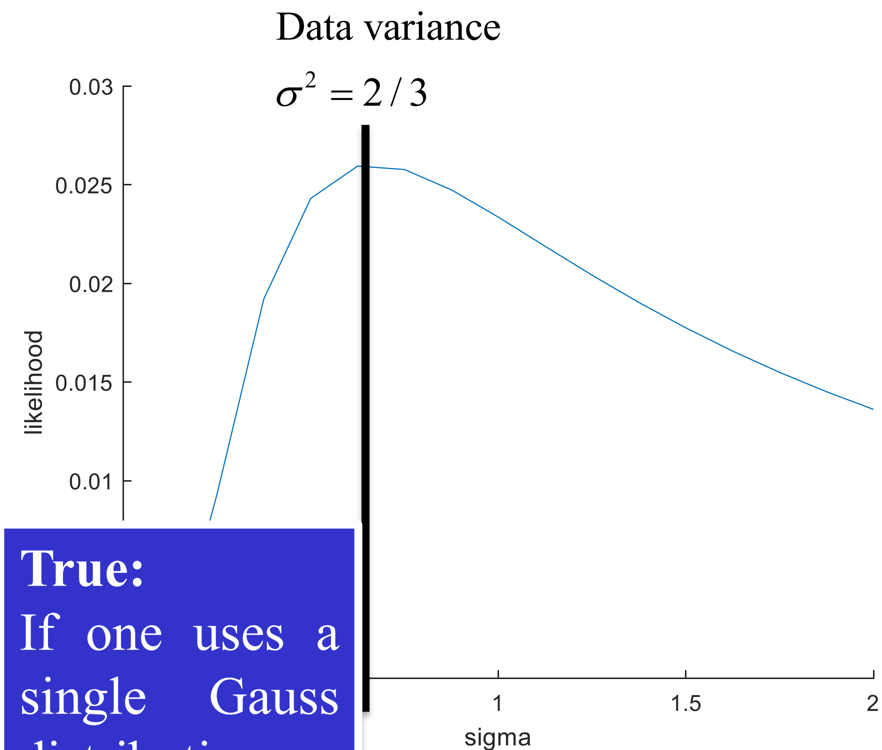
$p\left(x^i;\mu,\sigma\right);\ \mu=0,\ \ \sigma^2=2;$

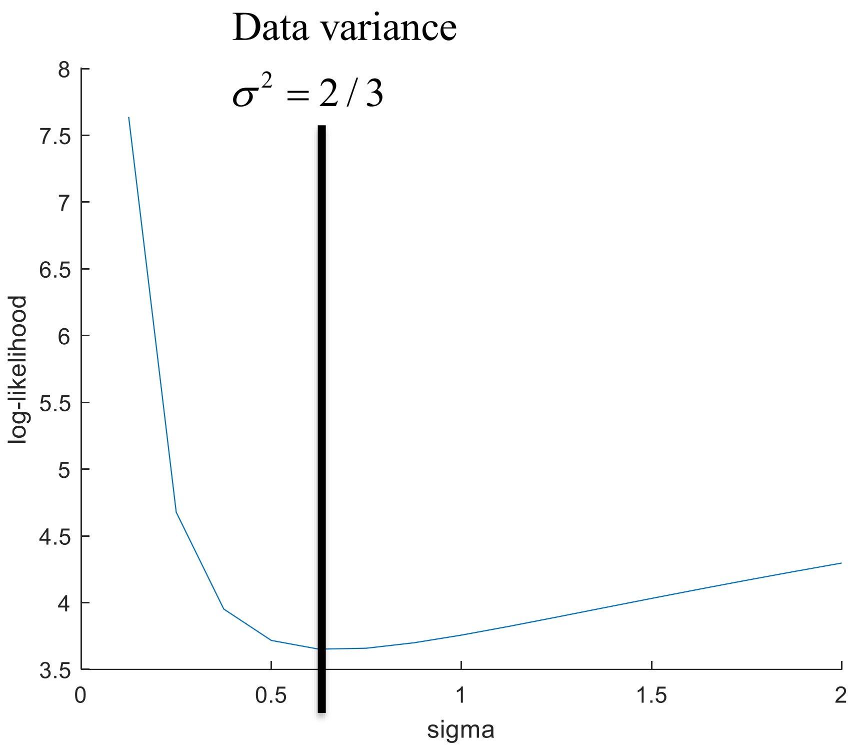
$L=0.013; -\log L=4.29$



Maximum Likelihood

The maximum of the likelihood, minimum of -log-likelihood, is obtained for a distribution with same variance as that of the data.





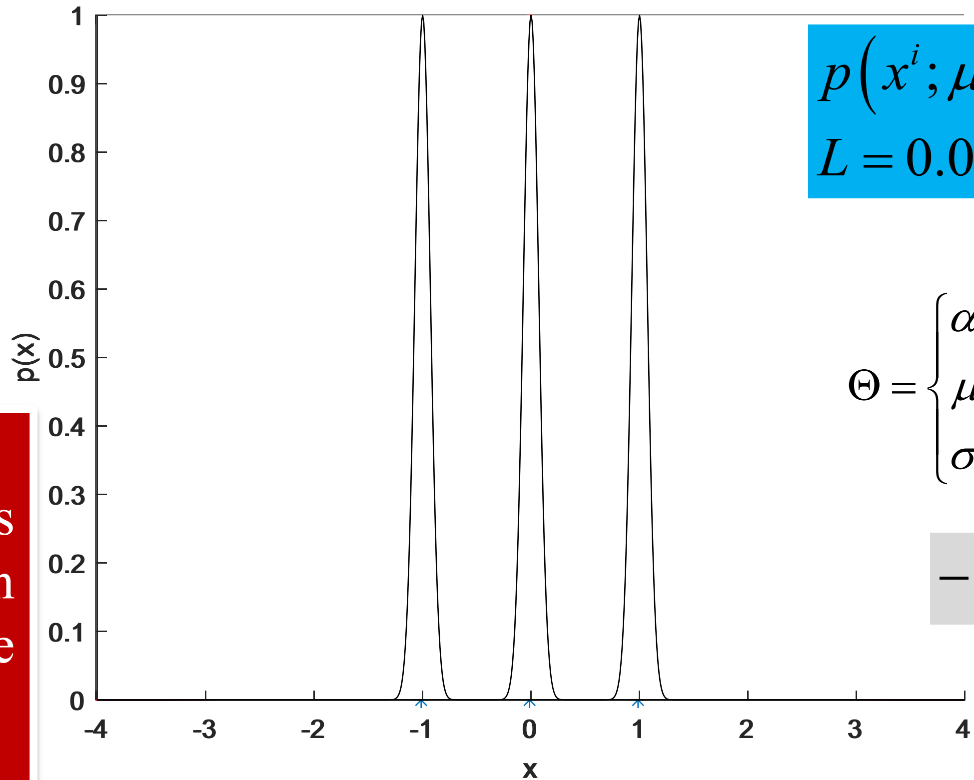
True:
If one uses a single Gauss distribution as model fit.

Real values for the normalized distributions

Maximum Likelihood

Likelihood for a single Gauss function, given a set of M datapoints $\left\{x^i\right\}_{i=1}^{M}$

$$L(\Theta \mid X) = \prod_{i=1}^{M} \sum_{k=1}^{3} \alpha_k \, p\left(x^i; \mu_k, \sigma_k\right)$$



$p\left(x^i; \mu, \sigma\right); \; \mu = 0, \quad \sigma^2 = 2/3;$

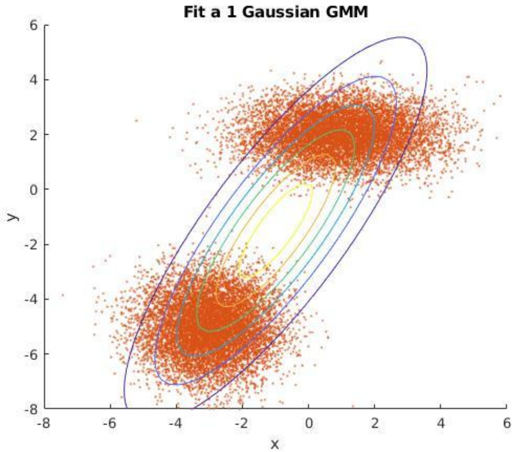
$L = 0.026; -\log L = 3.64$

$$\Theta = \begin{Bmatrix} \alpha_1 = 1/3, \alpha_2 = 1/3, \alpha_3 = 1/3, \\ \mu_1 = -1, \mu_3 = 0, \mu_3 = 1, \\ \sigma_1 = 0.01, \sigma_2 = 0.01, \sigma_3 = 0.01 \end{Bmatrix}$$

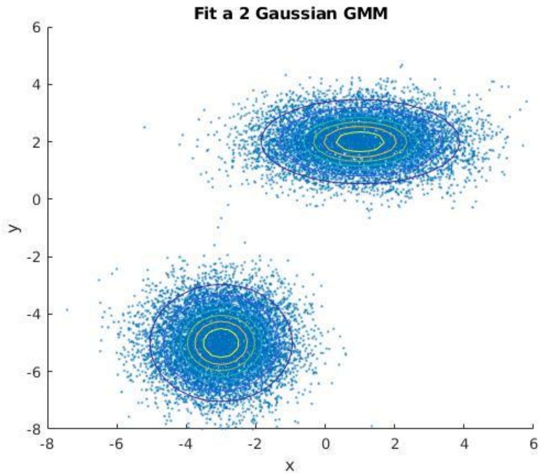
$-\log L(\Theta \mid X) = 3.29$

False:
A better fit is obtained when using more complex distributions.

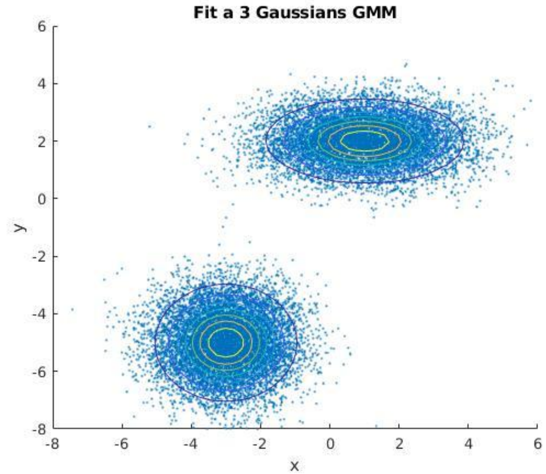
Maximum likelihood with Mixture of Gaussians



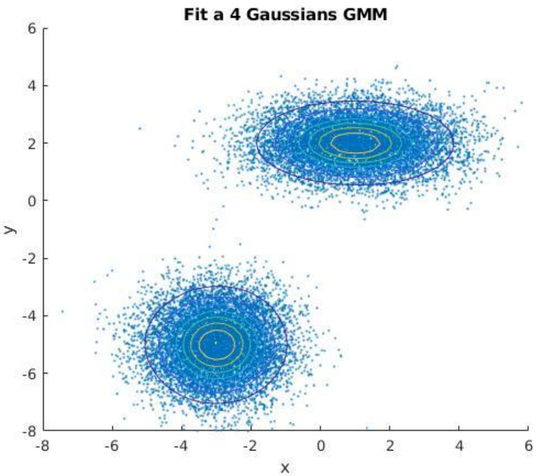
Log-Likelihood = -87825



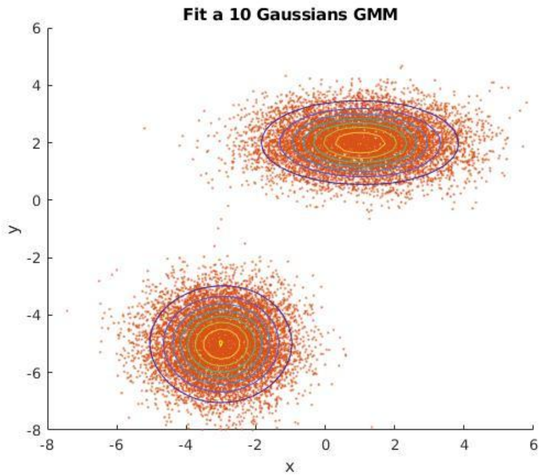
Log-Likelihood = - 70610



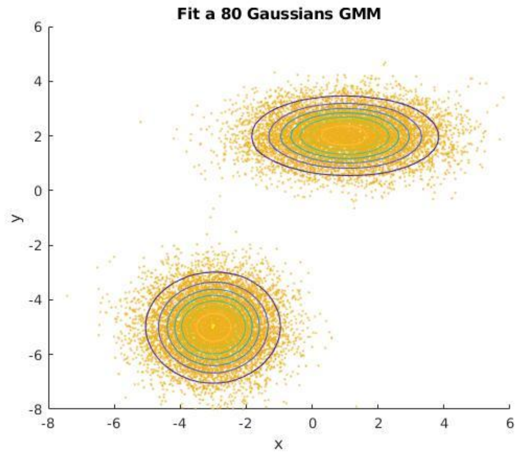
Log-Likelihood = -70610



Log-Likelihood = -70604



Log-Likelihood = -70601



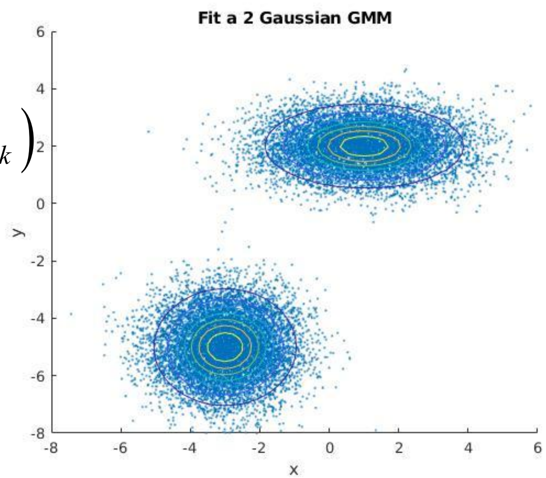
Log-Likelihood = -70581

Maximum likelihood with Mixture of Gaussians

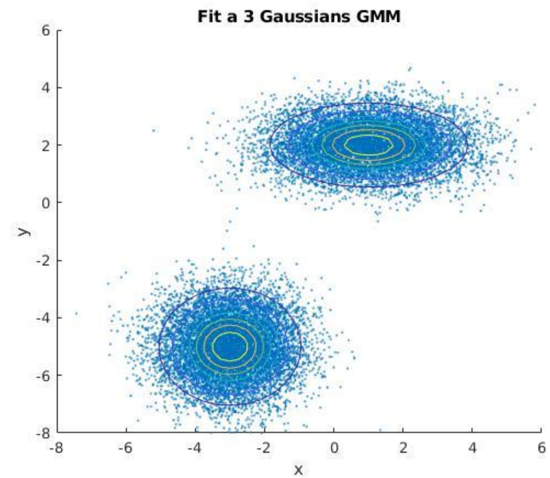
Likelihood identical:

$$L(\Theta \mid X) = \prod_{i=1}^{M} \sum_{k=1}^{K} \alpha_k \, p\left(x^i ; \mu_k, \sigma_k\right)$$

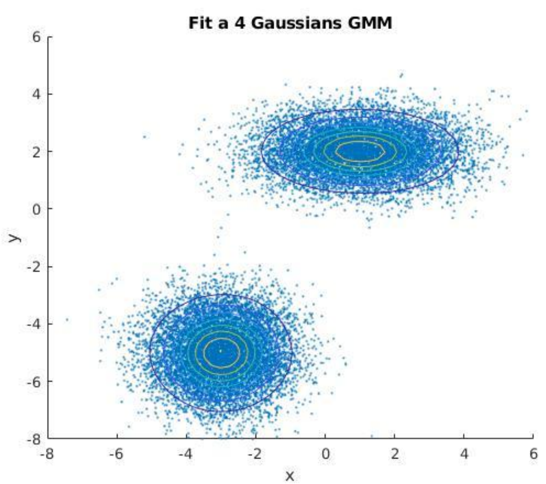
$$\sum_{k=1}^{K} \alpha_k = 1$$



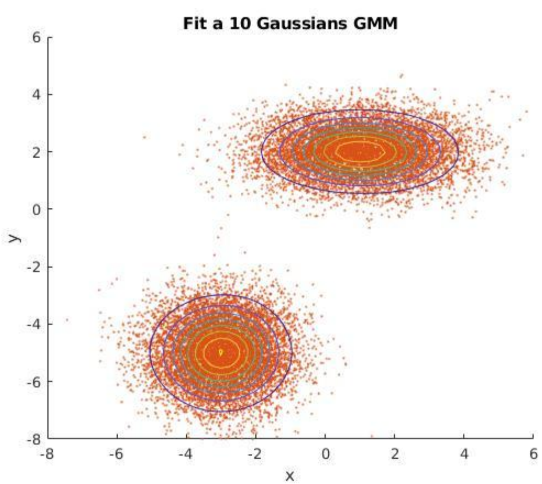
Log-Likelihood = - 70610



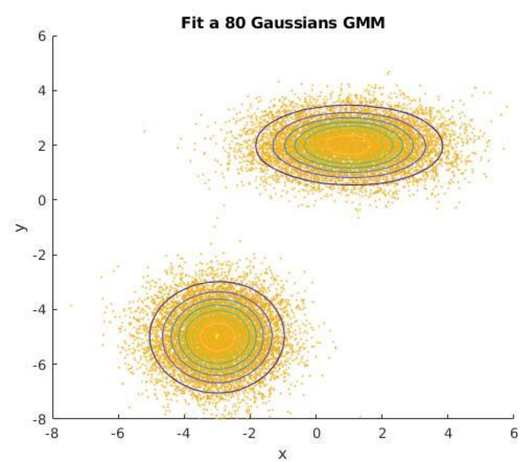
Log-Likelihood = -70610



Log-Likelihood = -70604

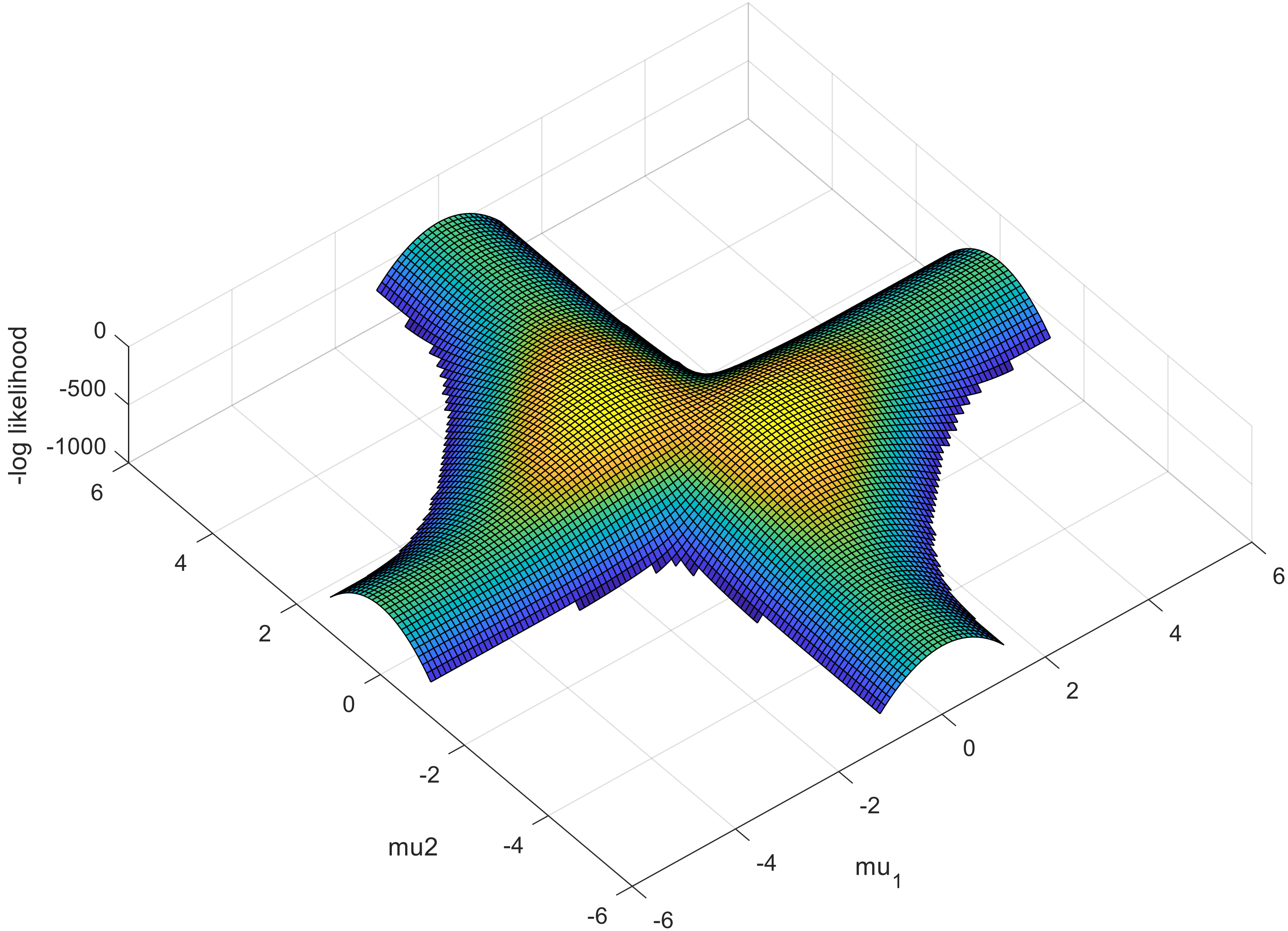


Log-Likelihood = -70601



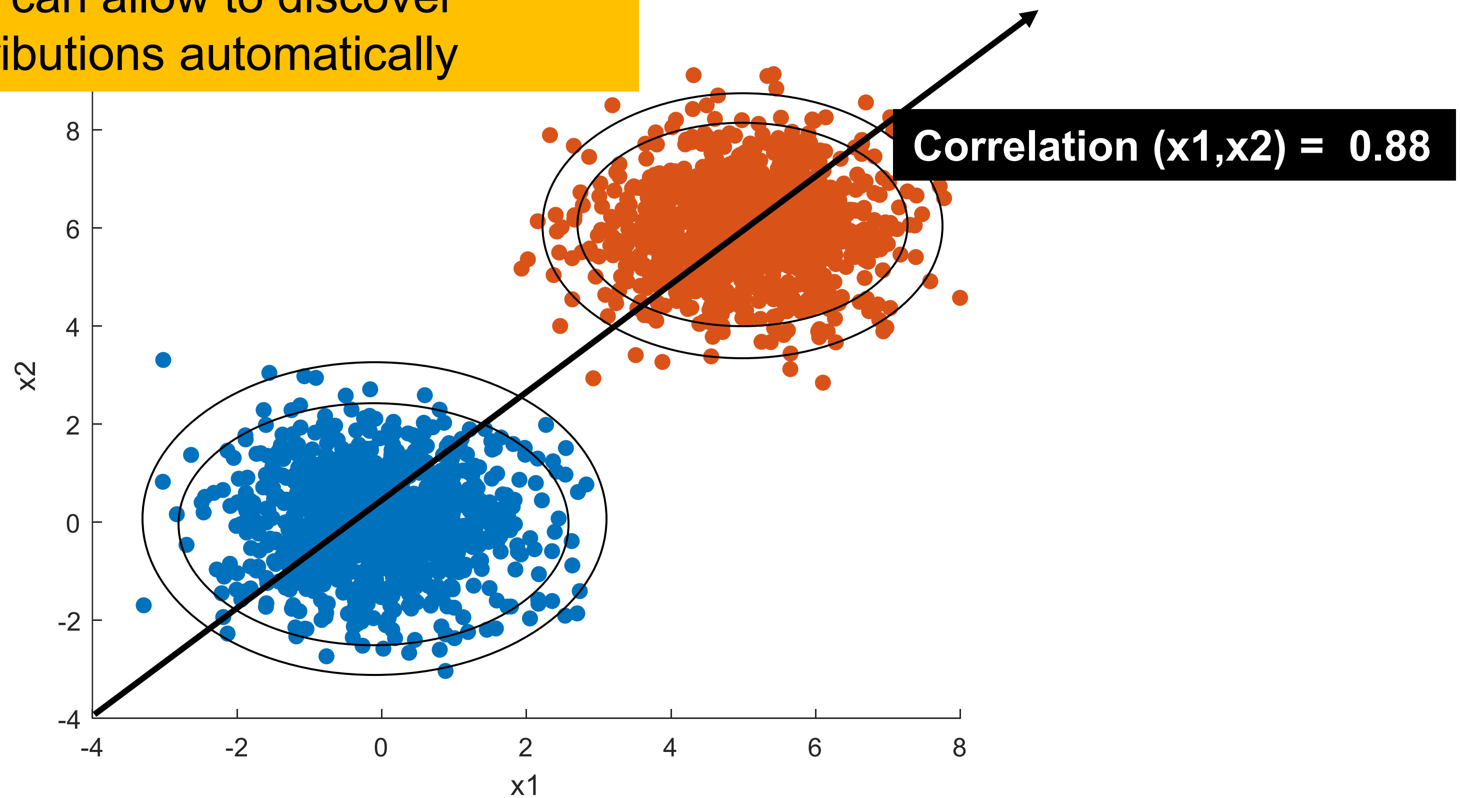
Log-Likelihood = -70581

Non-convexity of the likelihood



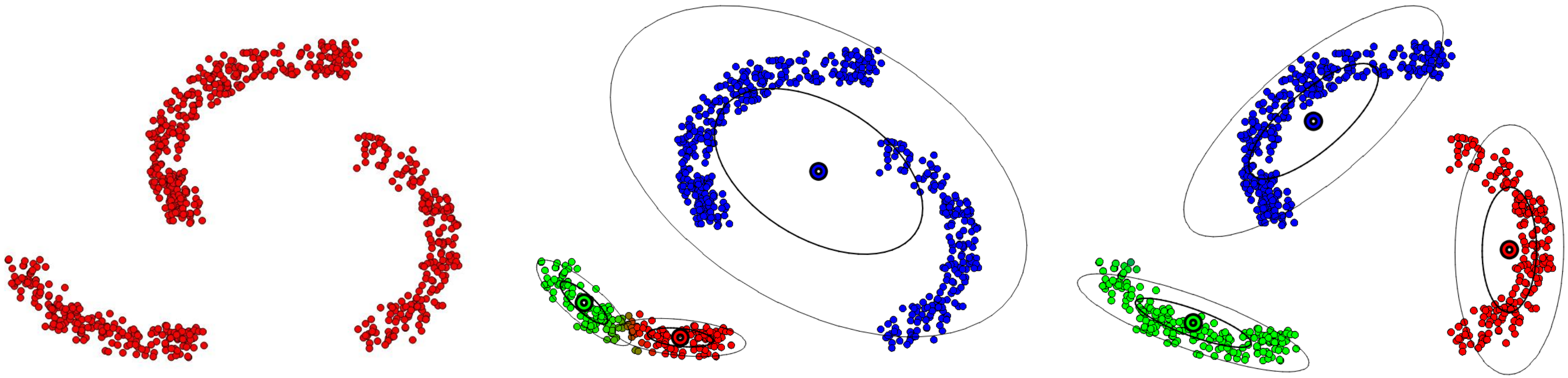
Correlations

Clustering can allow to discover these distributions automatically



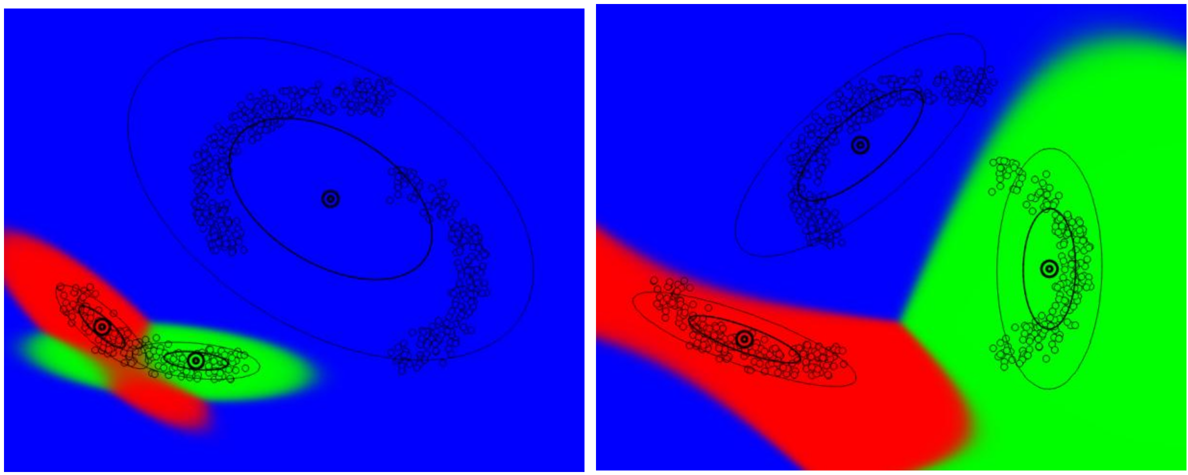
Spurious correlations as we compare two groups of data that come from two different distributions

Clustering with Gaussian Mixture Models



Original data **Solution 1** **Solution 2**

Compute boundary across clusters by comparing likelihood of each cluster, i.e. of each Gauss function.

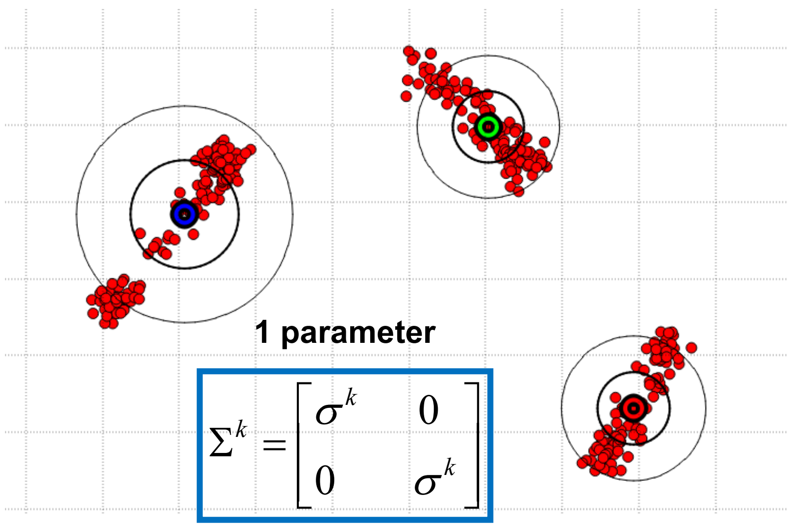


Boundaries

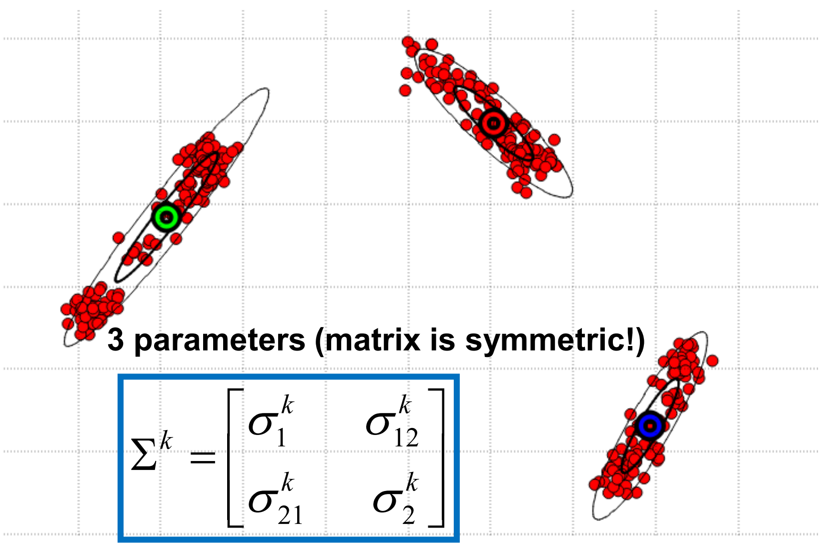
Tradeoff between computation costs and better fit

- A GMM can fit very closely the local distribution of datapoints.
- But this comes at the cost of an increase in the number of parameters

Full covariance matrices require N(N+1)/2 parameters against N for diagonal matrices and 1 for spherical matrices.*



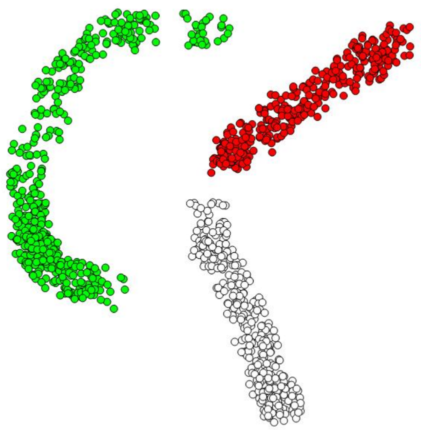
$$\Sigma^k = \begin{bmatrix} \sigma^k & 0 \\ 0 & \sigma^k \end{bmatrix}$$



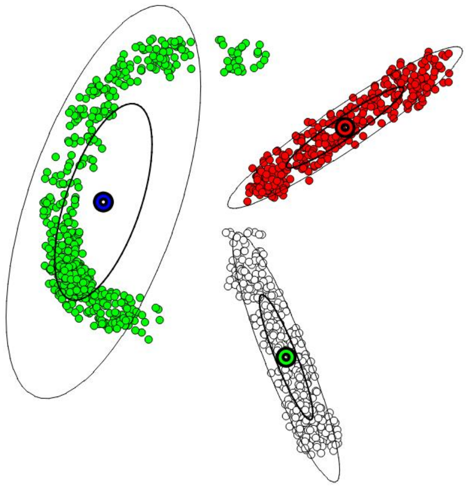
$$\Sigma^k = \begin{bmatrix} \sigma_1^k & \sigma_{12}^k \\ \sigma_{21}^k & \sigma_2^k \end{bmatrix}$$

How to determine the best mixtures of Gaussians?

Which Model?



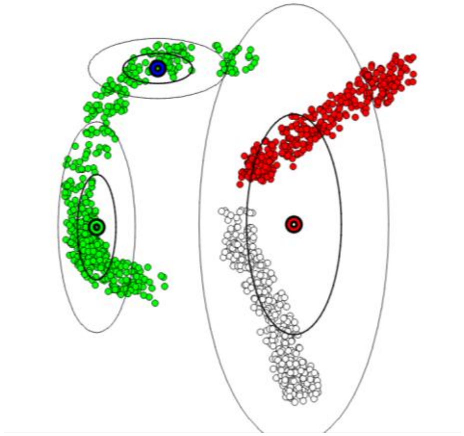
Original data



Solution with full matrices

Full matrices:

$$\Sigma^k = \begin{bmatrix} \sigma_1^k & \sigma_{12}^k \\ \sigma_{21}^k & \sigma_2^k \end{bmatrix}$$



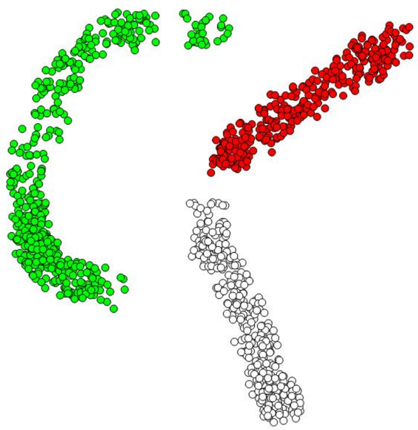
Diagonal matrices:

$$\Sigma^k = \begin{bmatrix} \sigma_1^k & 0 \\ 0 & \sigma_2^k \end{bmatrix}$$

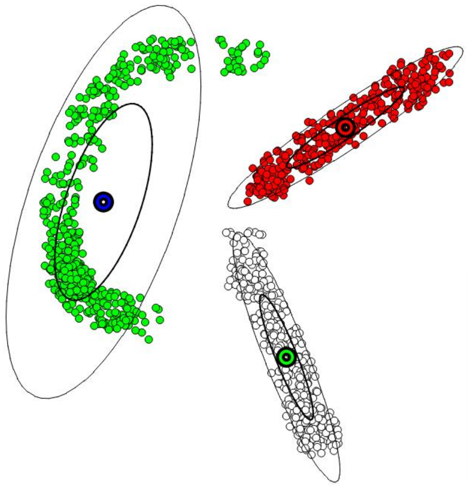
Solution with diagonal matrices

Metrics to choose model

1180 sample points



Original data



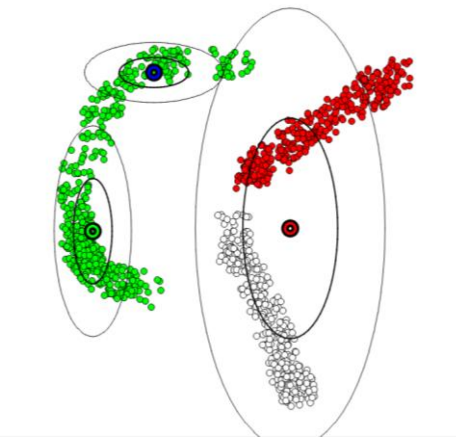
Solution with full matrices

Parameters: 3*(2+3)=15

Likelihood: 2719,
AIC: - 5419
BIC: -5373

Even if it requires more parameters, the gain on likelihood is important
→ Optimal solution

Which of the two solutions would get the best values on AIC or BIC?



Solution with diagonal matrices

Parameters: 3*(2+2)=12

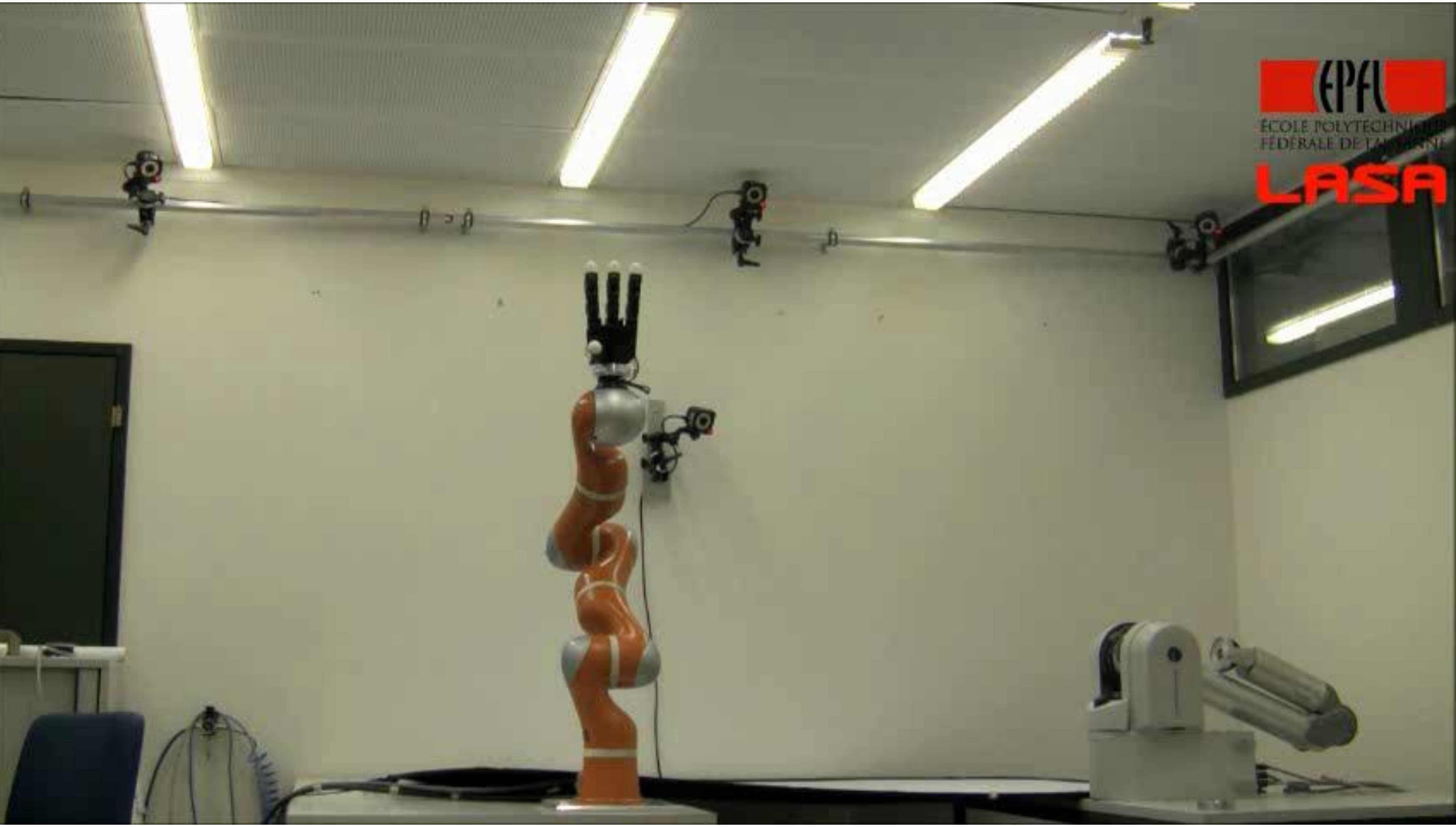
Likelihood: 1999,
AIC: - 2912
BIC: -2972

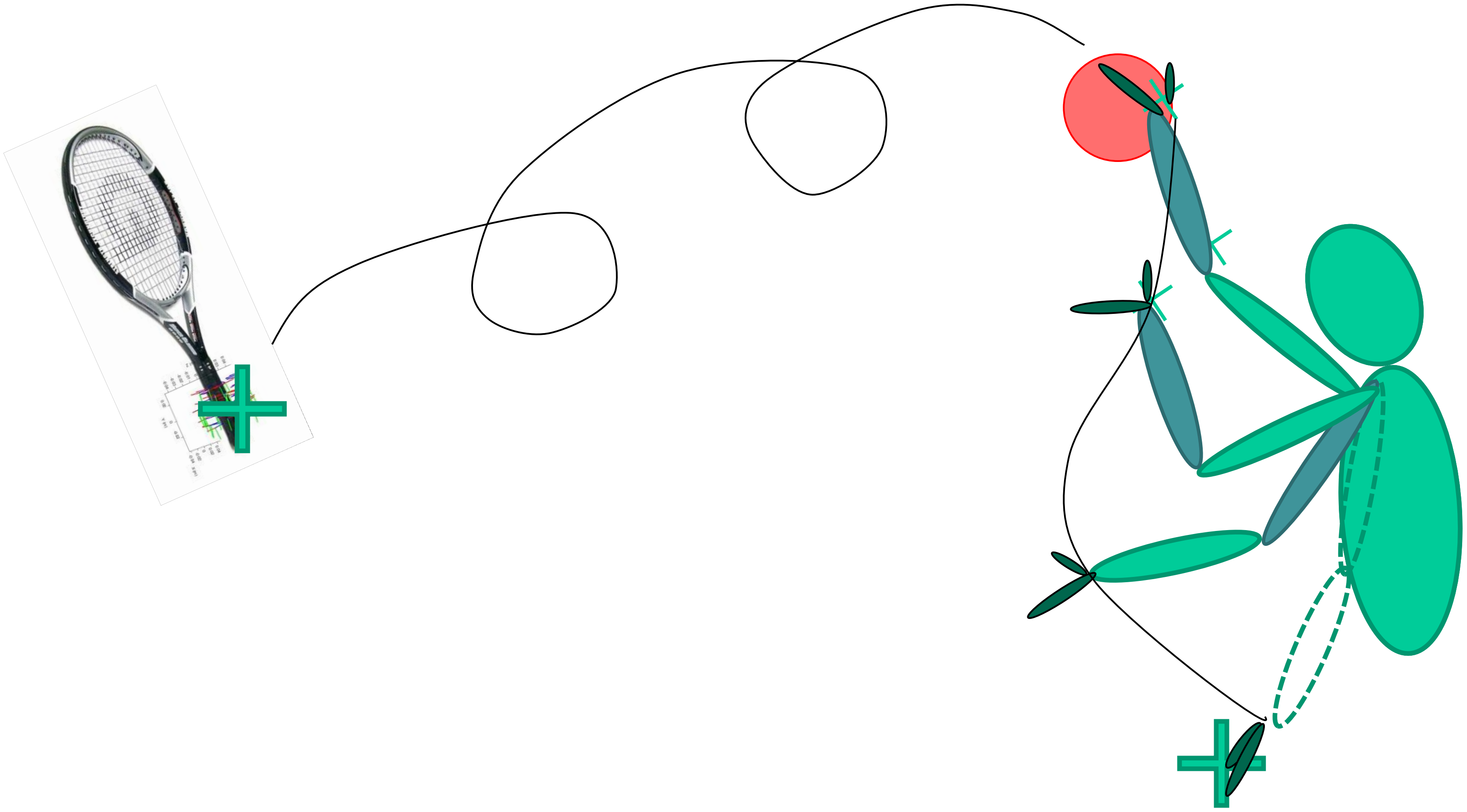
$$AIC = -2\ln(L) + 2B$$

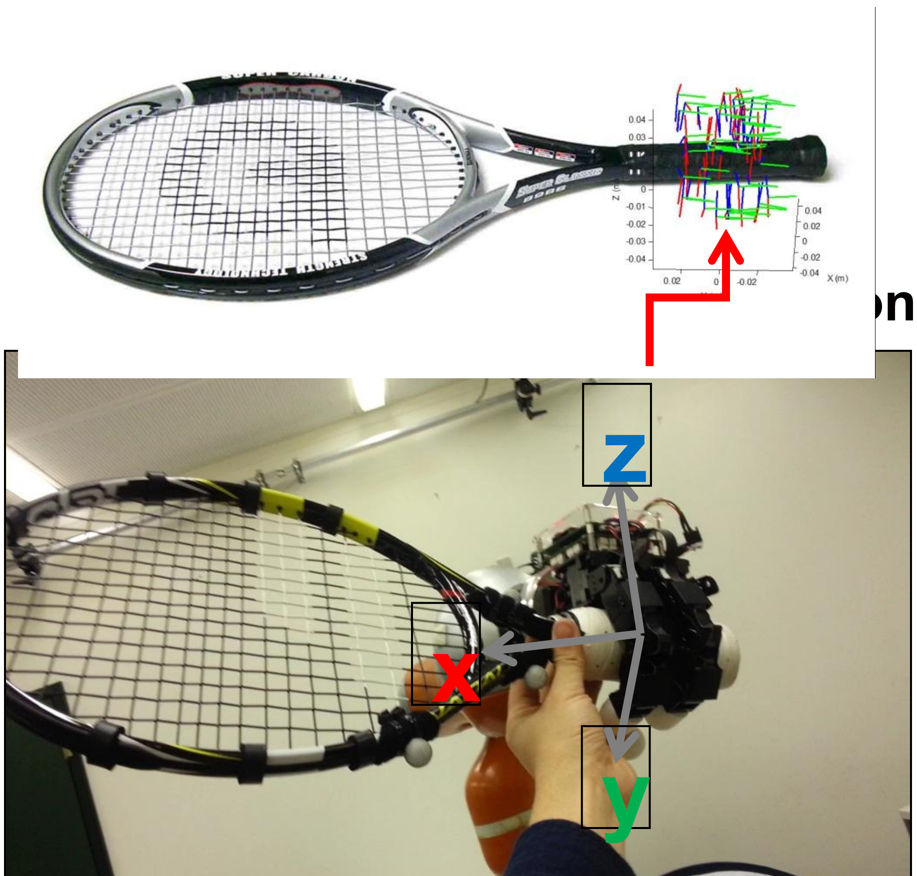
$$BIC = -2\ln(L) + \ln(M)B$$

B=# Parameters

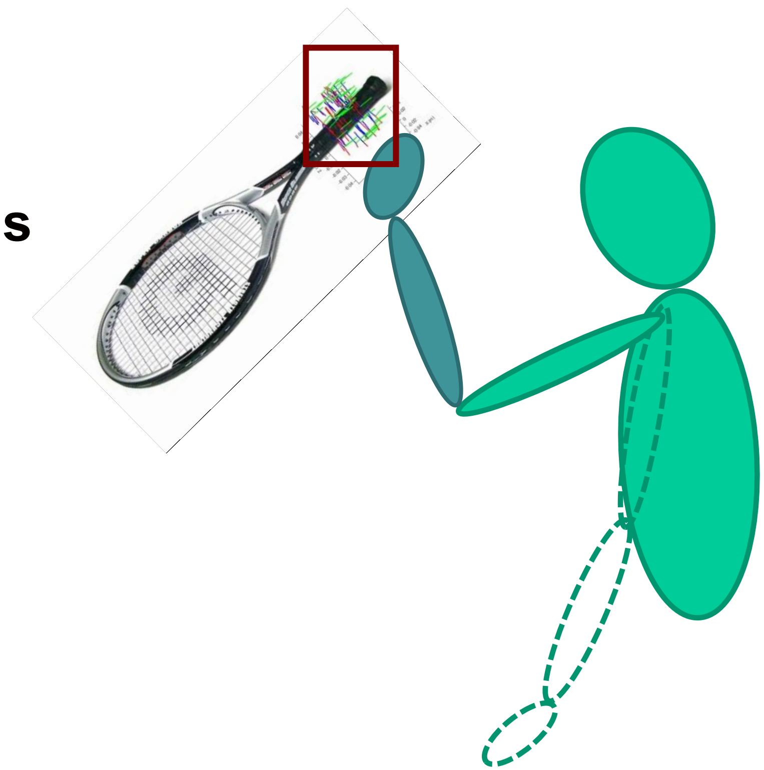
Examples of application of GMM



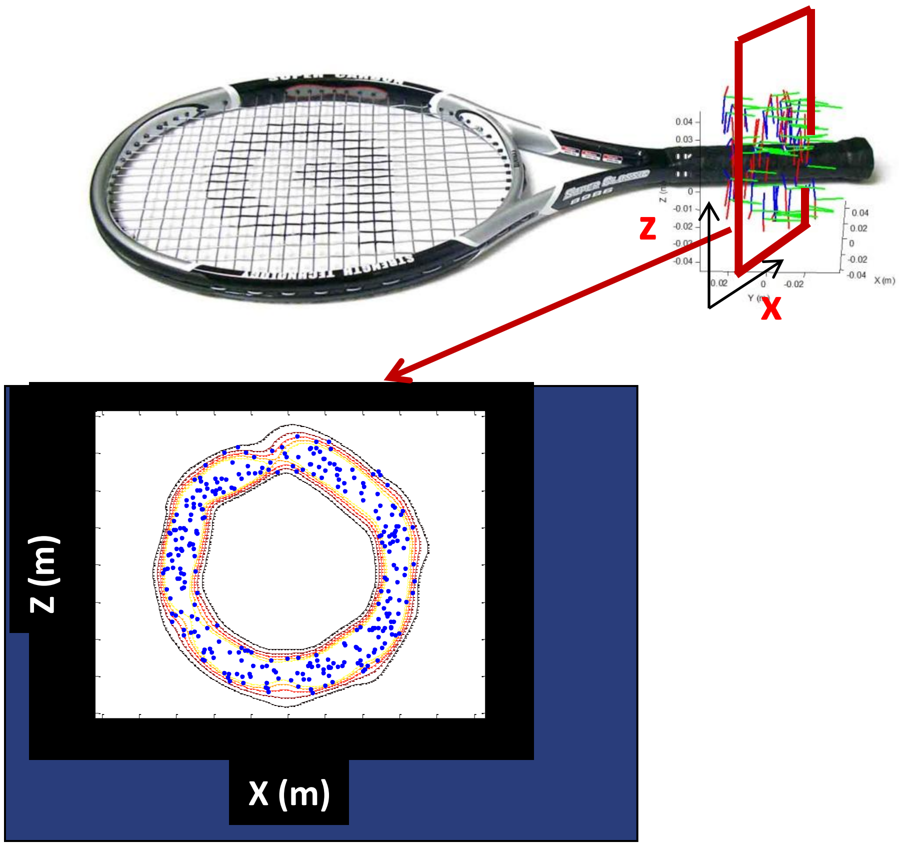


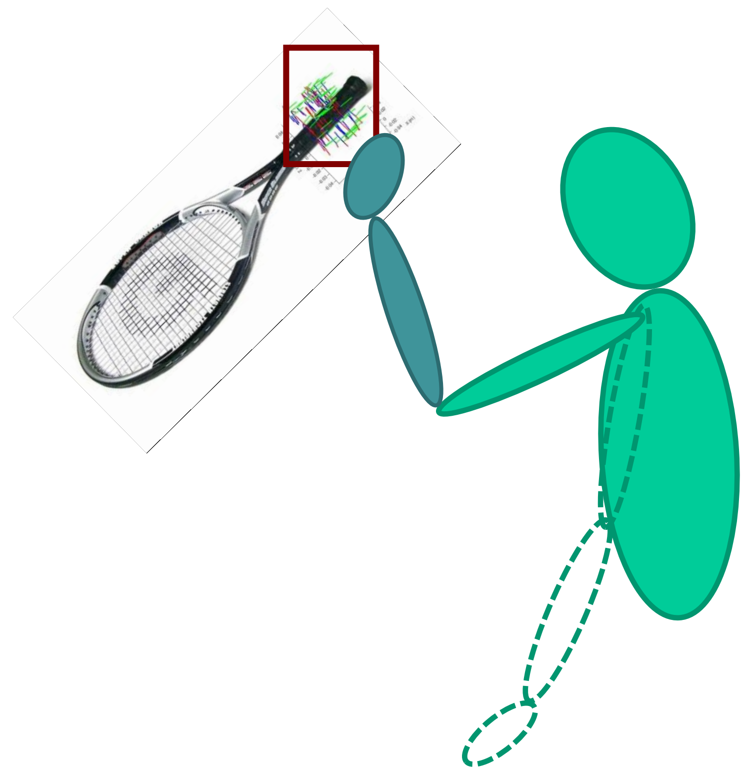


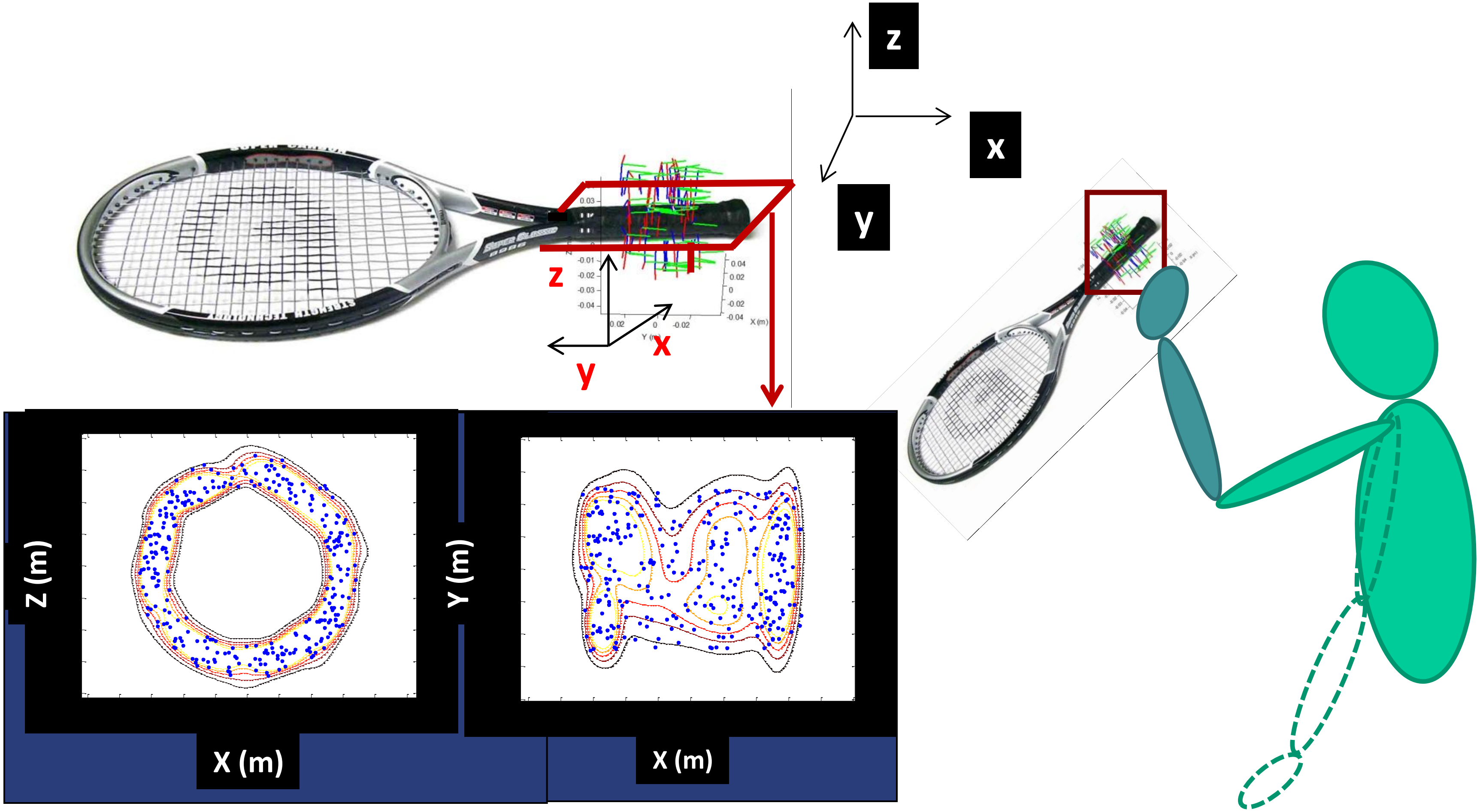
ons

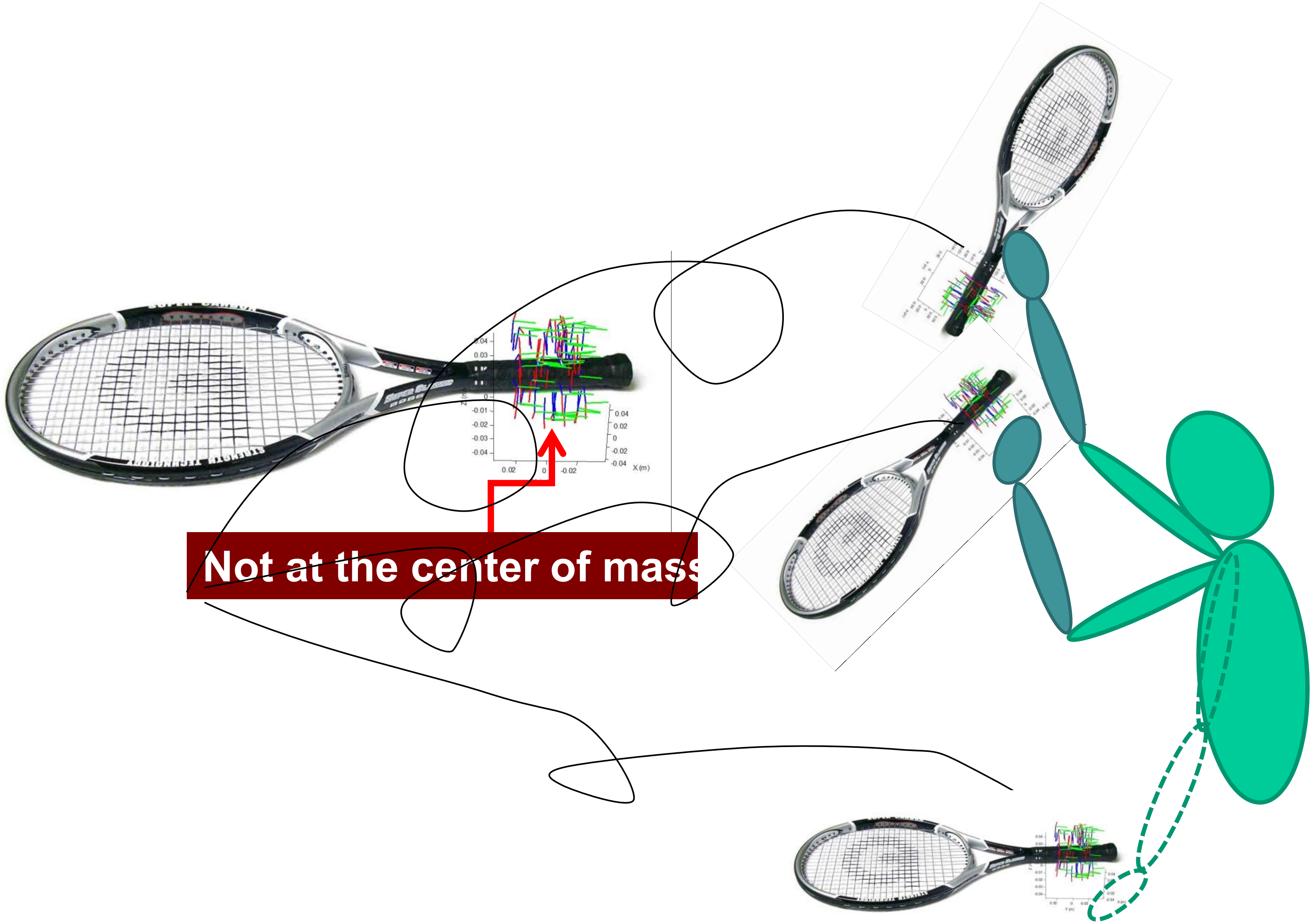


Learn catchable region on the object:
Train a GMM to model the distribution of grasps on the object from human demonstrations

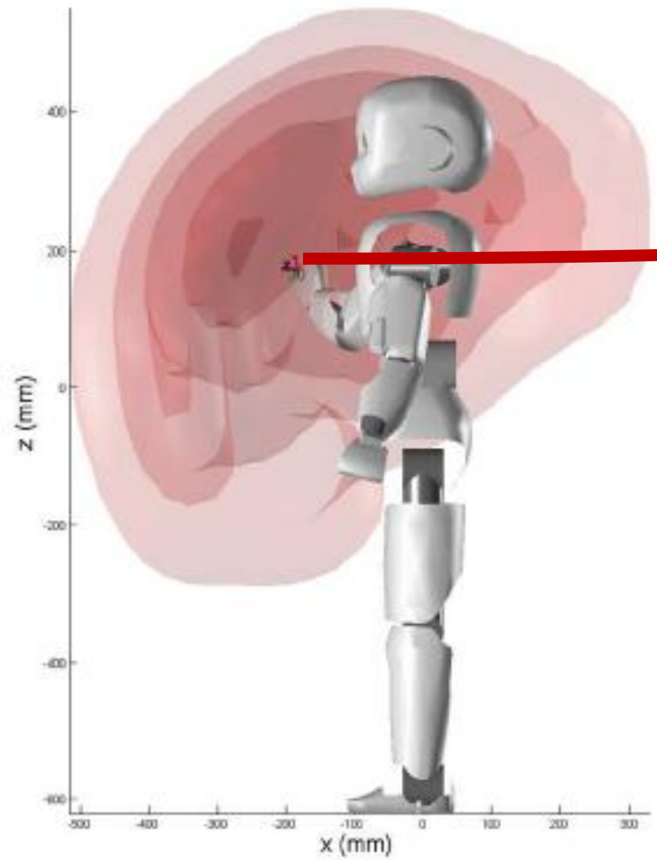


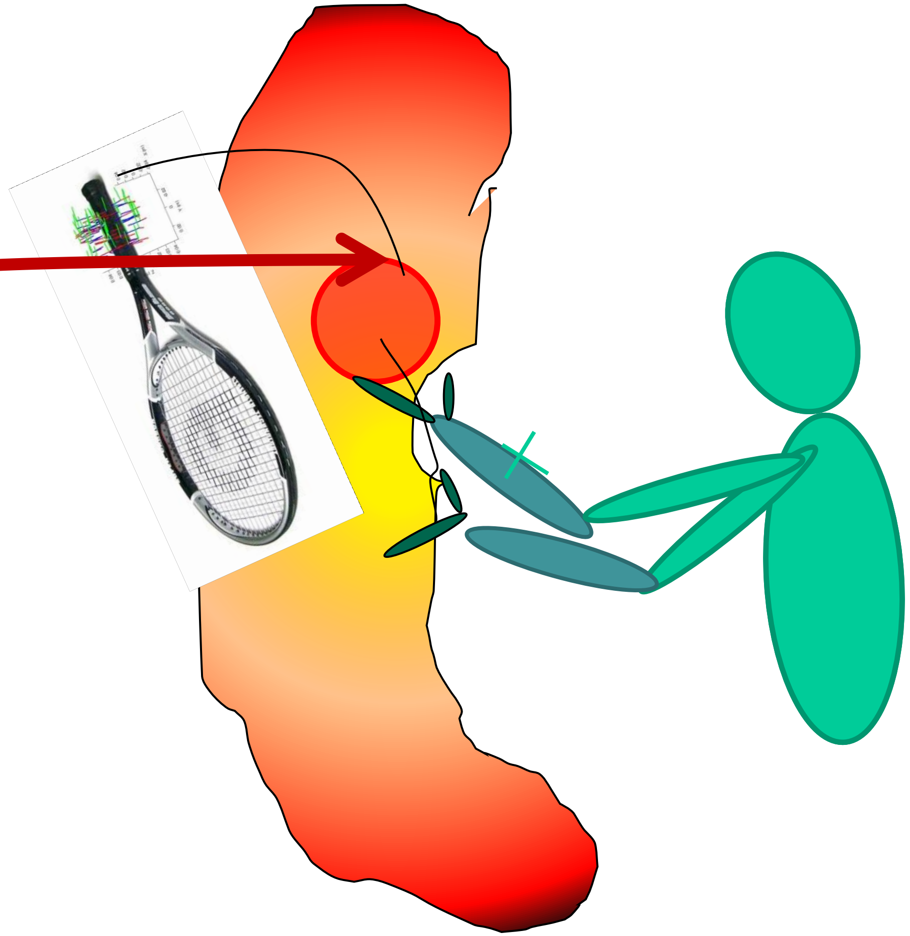



z
x
y
z
y
x
Z (m)
X (m)
Y (m)
X (m)

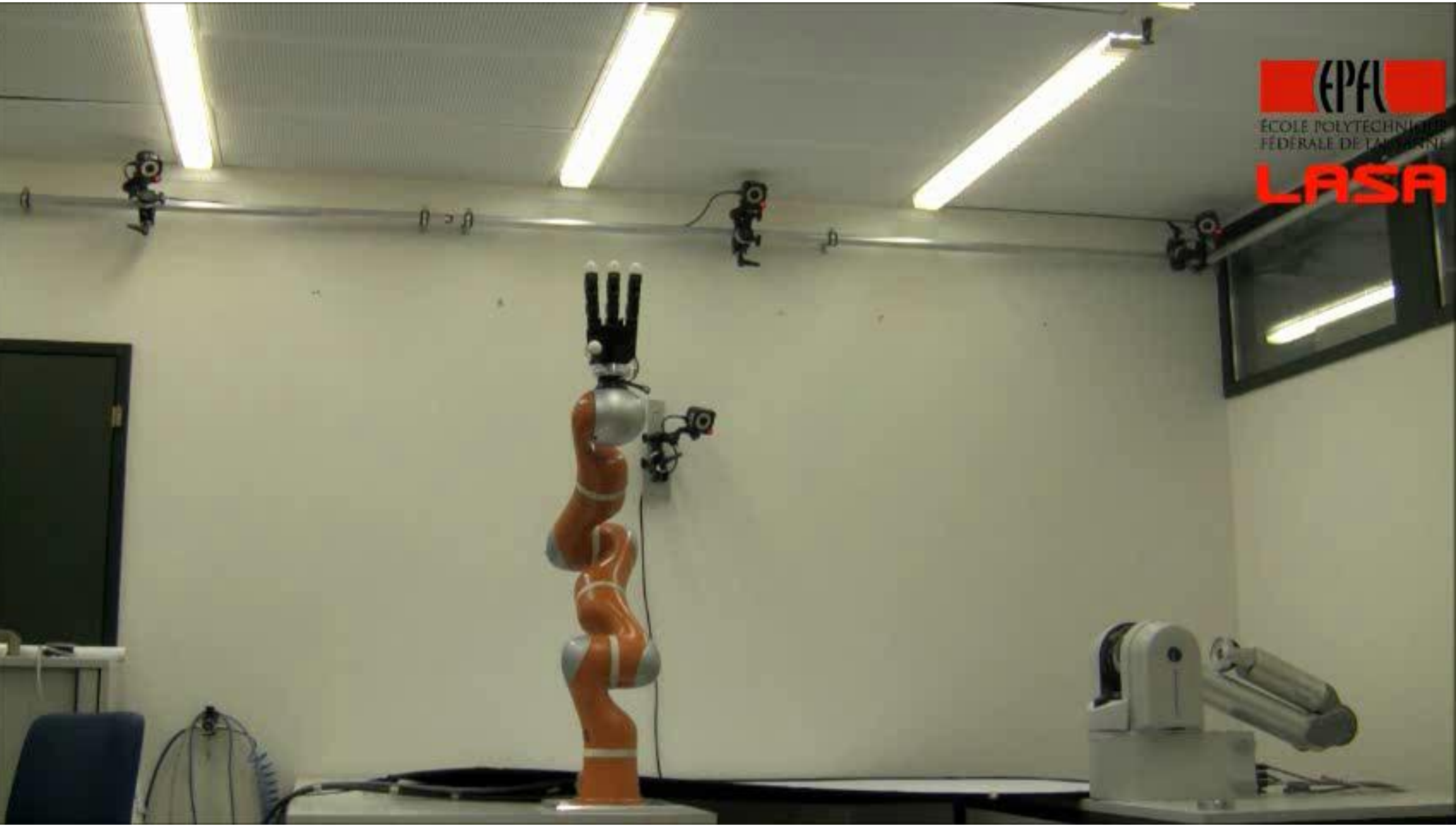

Not at the center of mass

Learn most likely region to catch object:
Train a GMM to model workspace of the 7 degrees of freedom robot arm





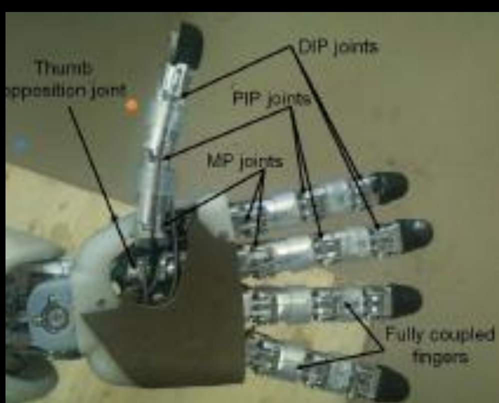
Examples of application of GMM



Examples of application of GMM

Reproduce variety of human grasps

Generate a variety of grasps for robots



Find 3 contact points that
minimize torques at fingers' contact position (θ):

$$\arg\min_{\theta} \sum_{i} \tau_i(\theta);$$

and ensure:

1) Stability (force closure)

2) Kinematic feasibility

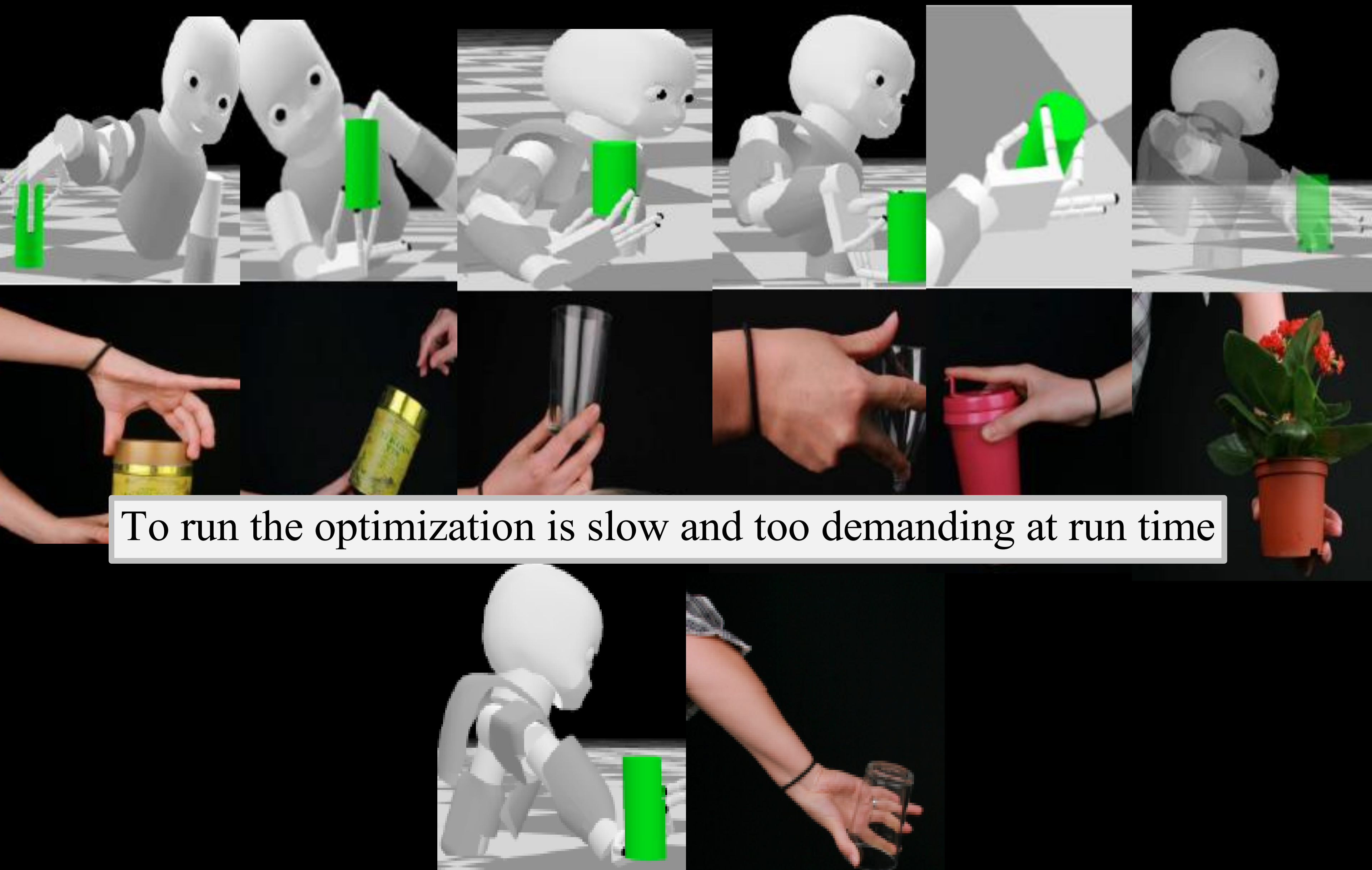
3) Collision avoidance

→ Nonlinear constraint-based optimization

Exploit the non-convex optimization formulation!
→ Search for numerous feasible solutions

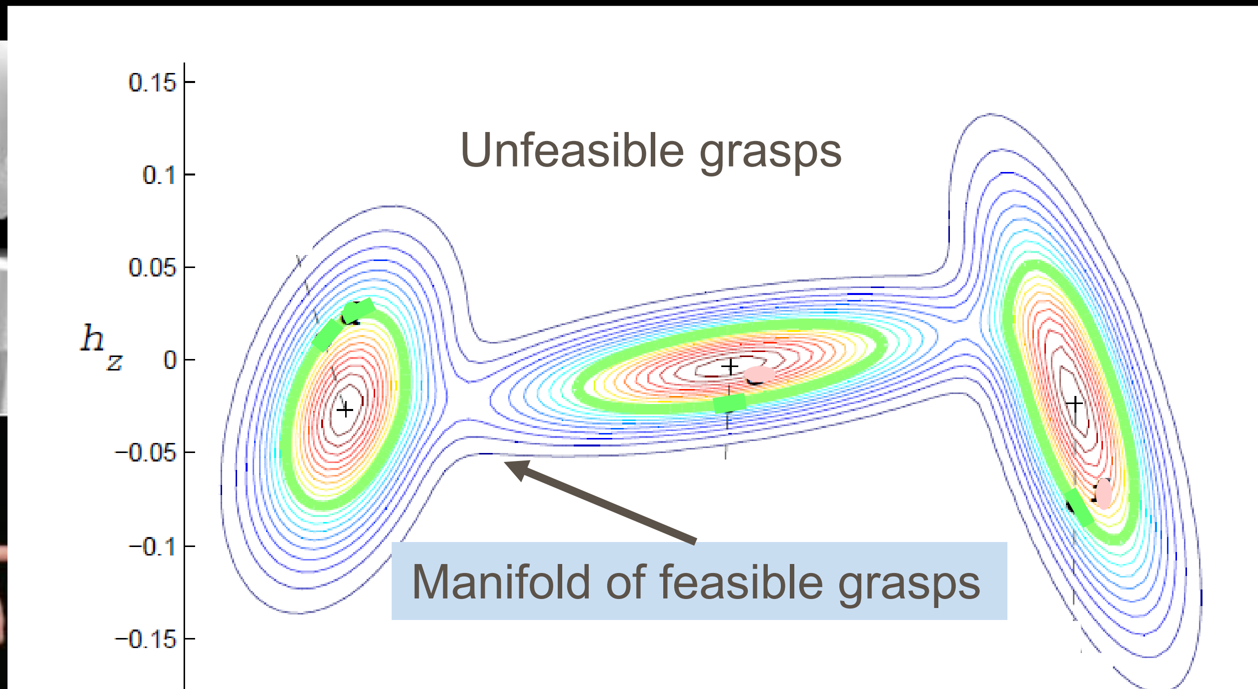
Found ~1500 solutions; clustered into 14 classes

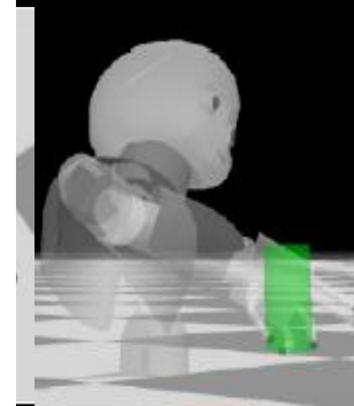
Generate a variety of grasps for robots



To run the optimization is slow and too demanding at run time

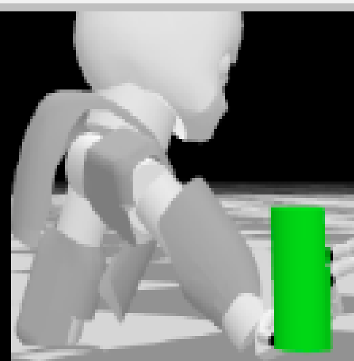
Generate a variety of grasps for robots

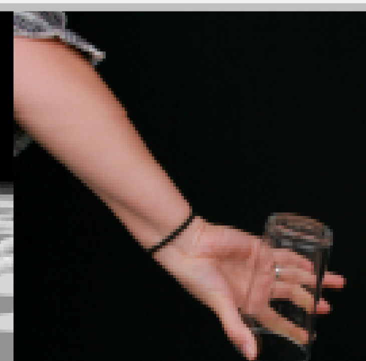




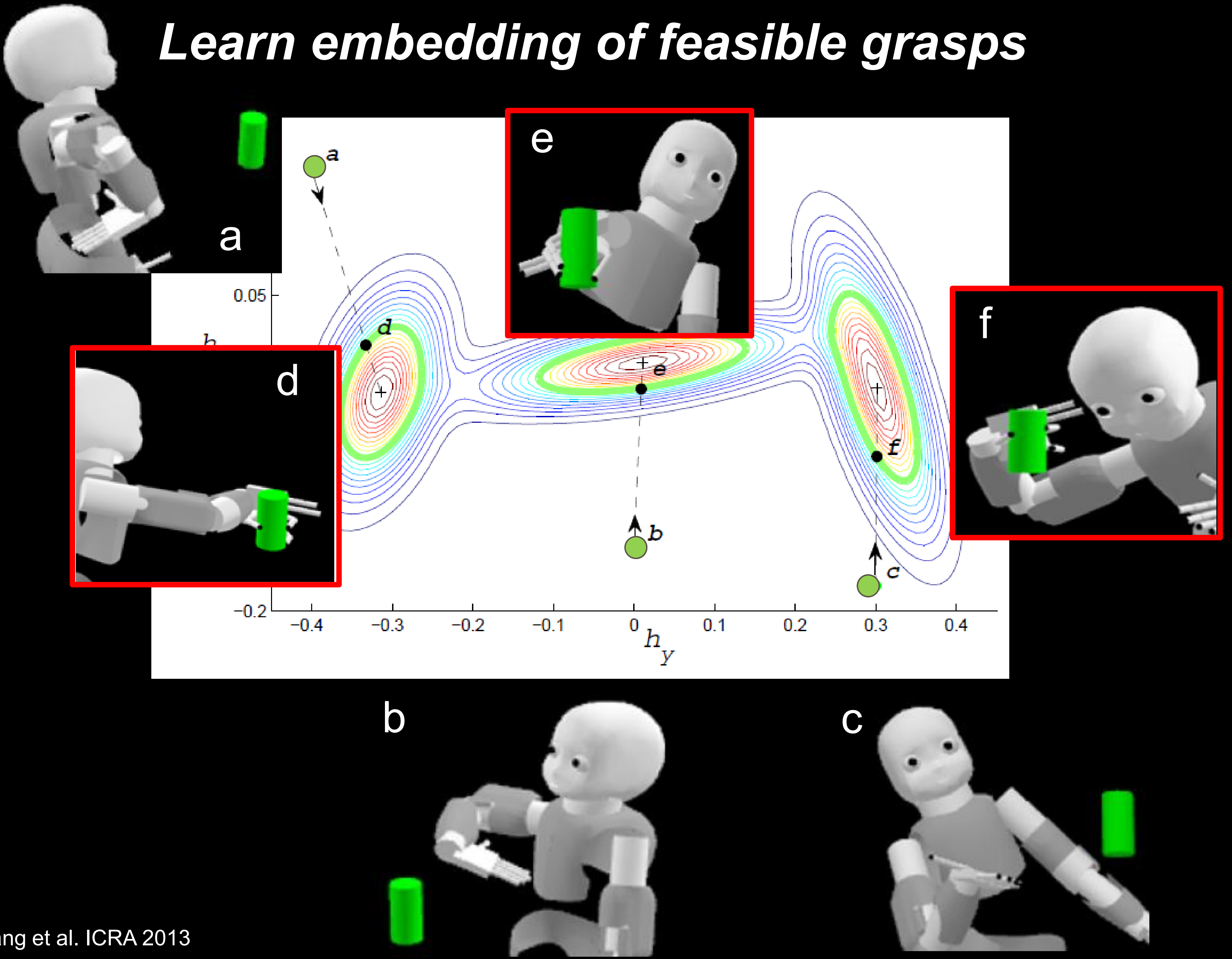
Learn distribution of feasible postures on cylinder $P(\theta)$

$\theta \in \mathbb{R}^{14}$: Hand position and orientation + 8 finger joints





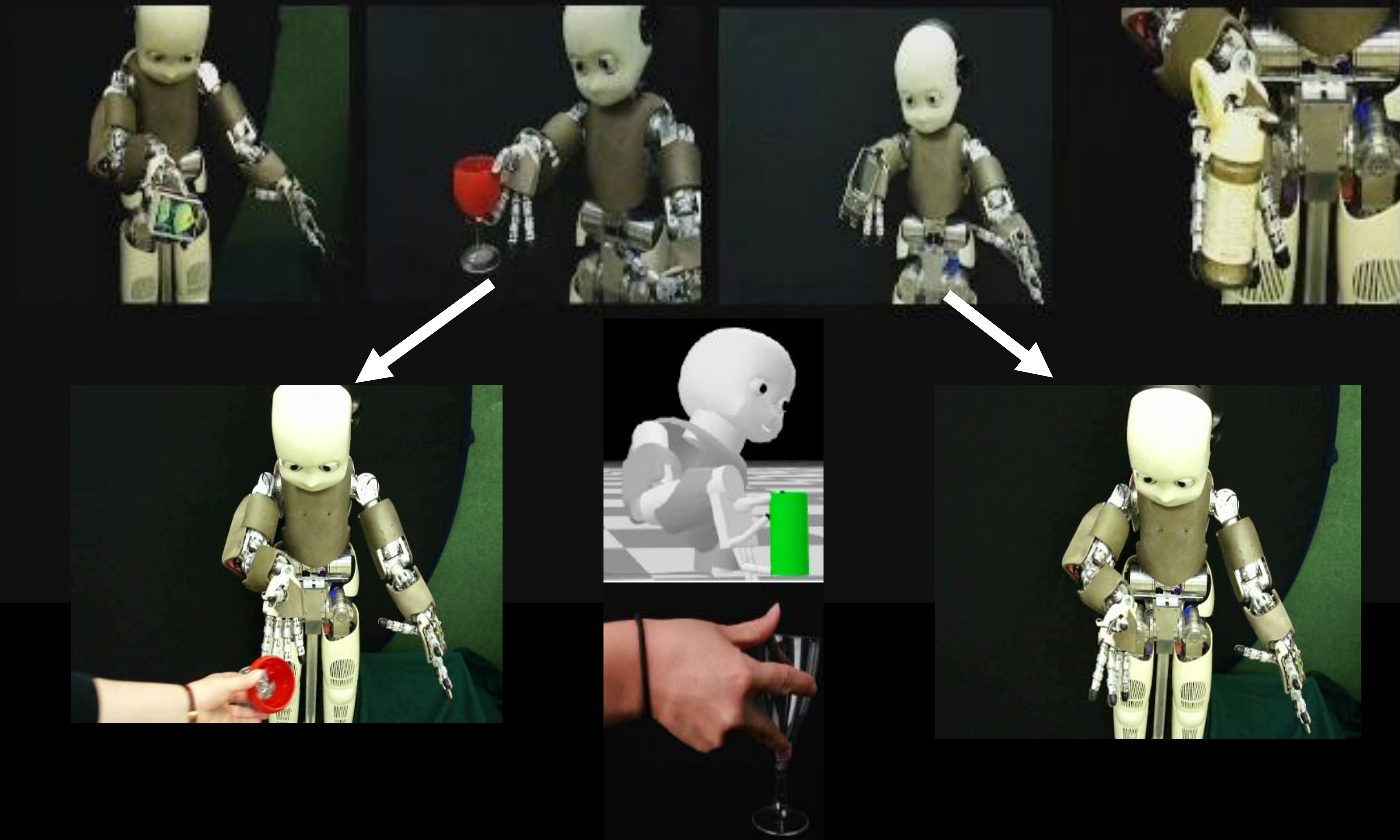
Learn embedding of feasible grasps



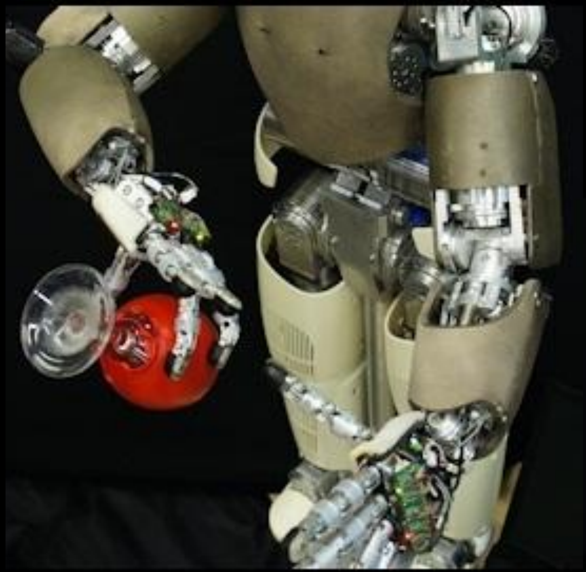
Huang et al. ICRA 2013

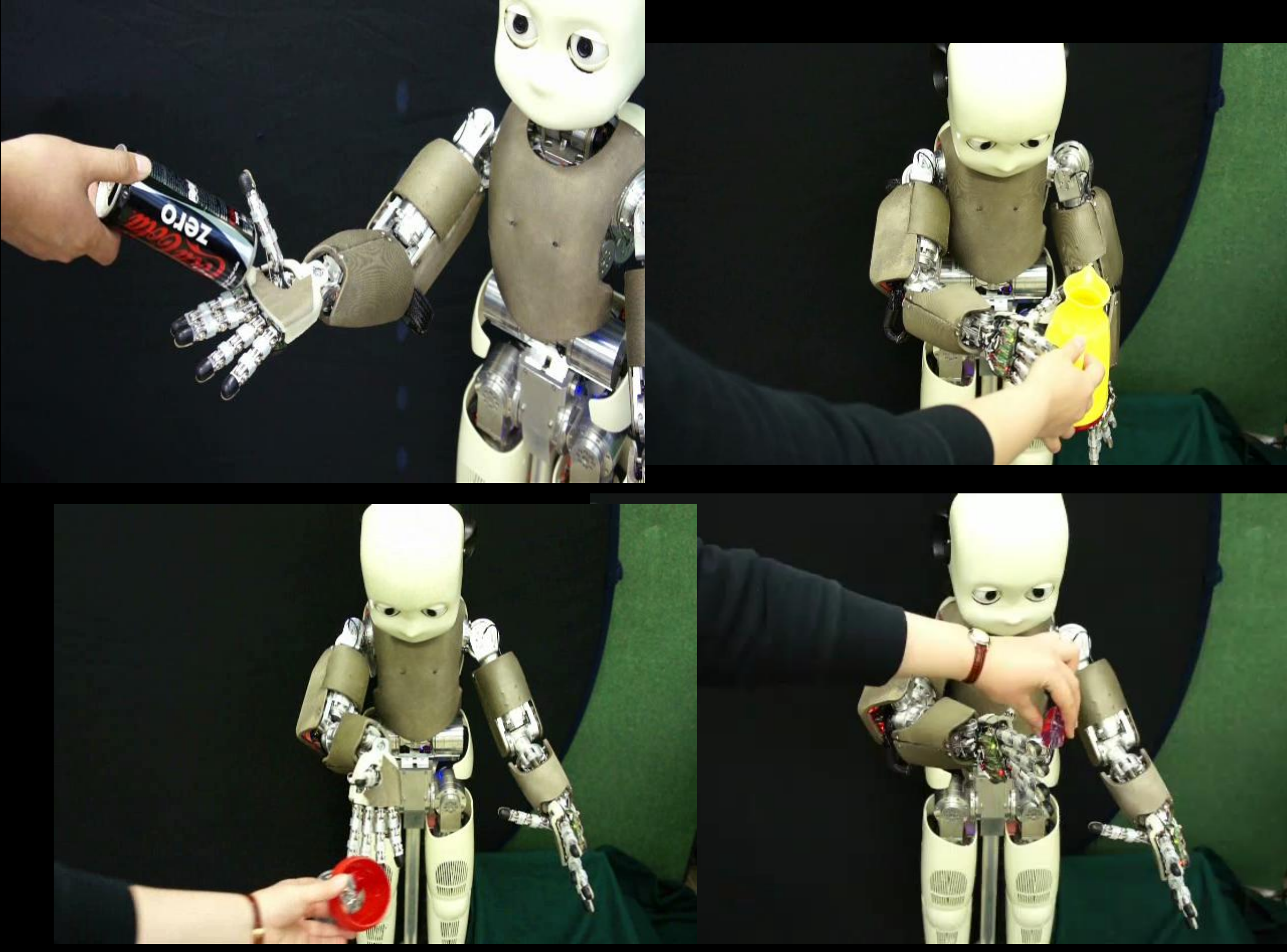
Generalization to variety of objects

Generalization to variety of objects

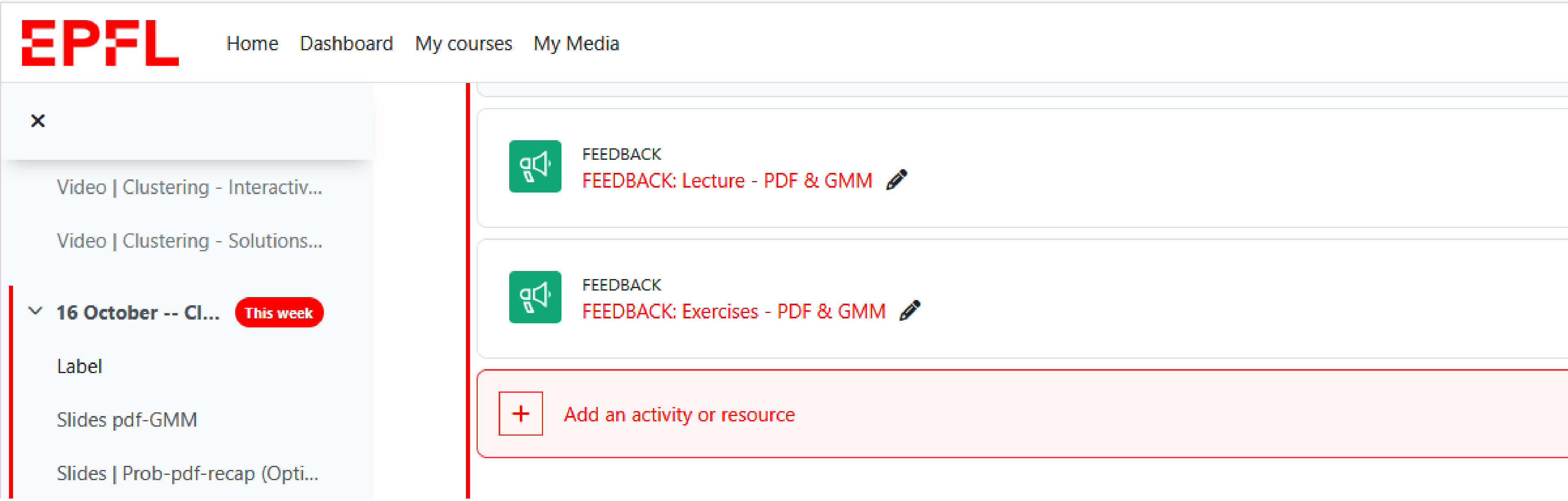


EPFL




zero

Give Feedback through Moodle



You can give feedback on a weekly basis separately for each lecture, exercise and practice session.

Add comments or suggestions. A yes-no feedback is not very useful.