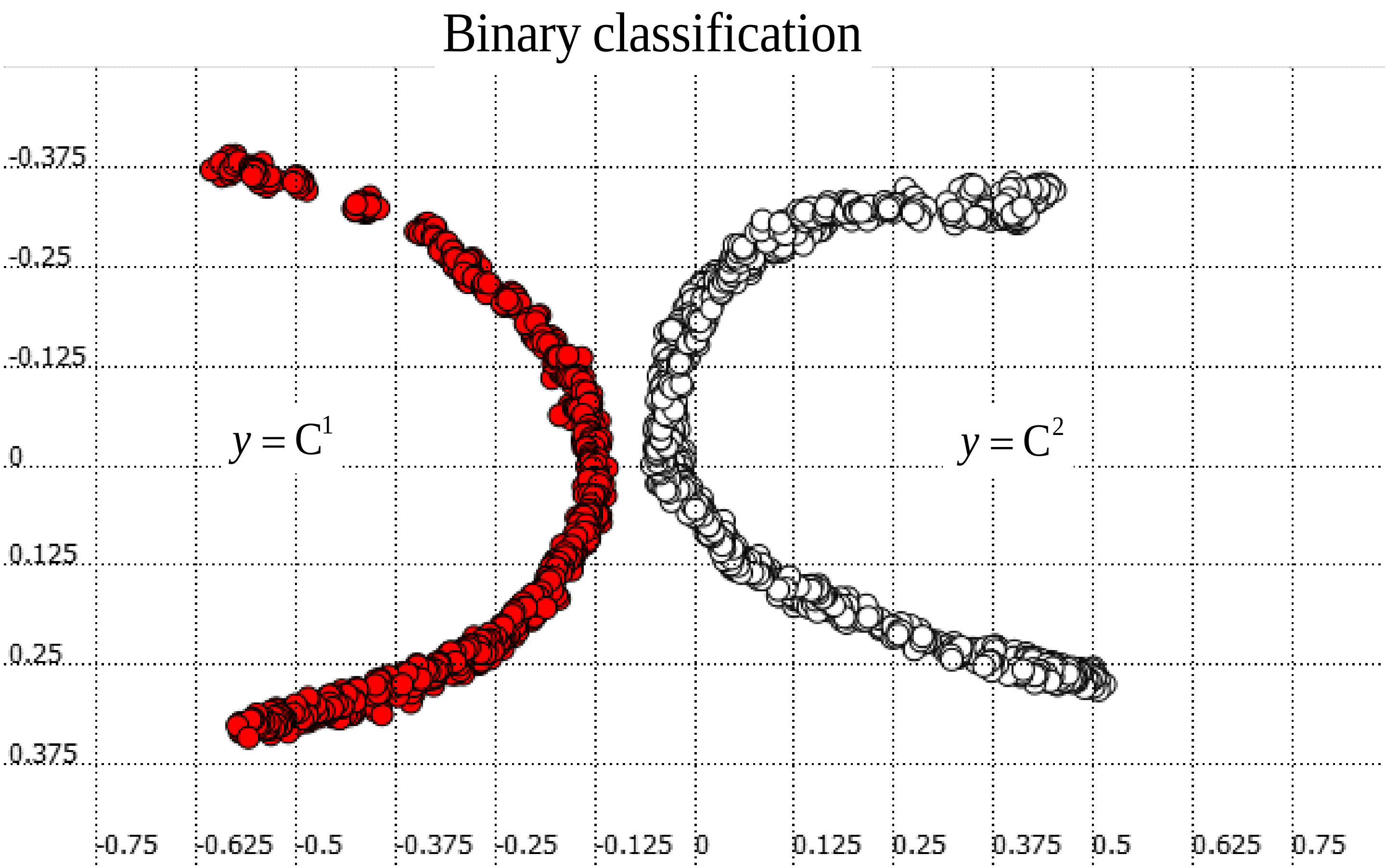
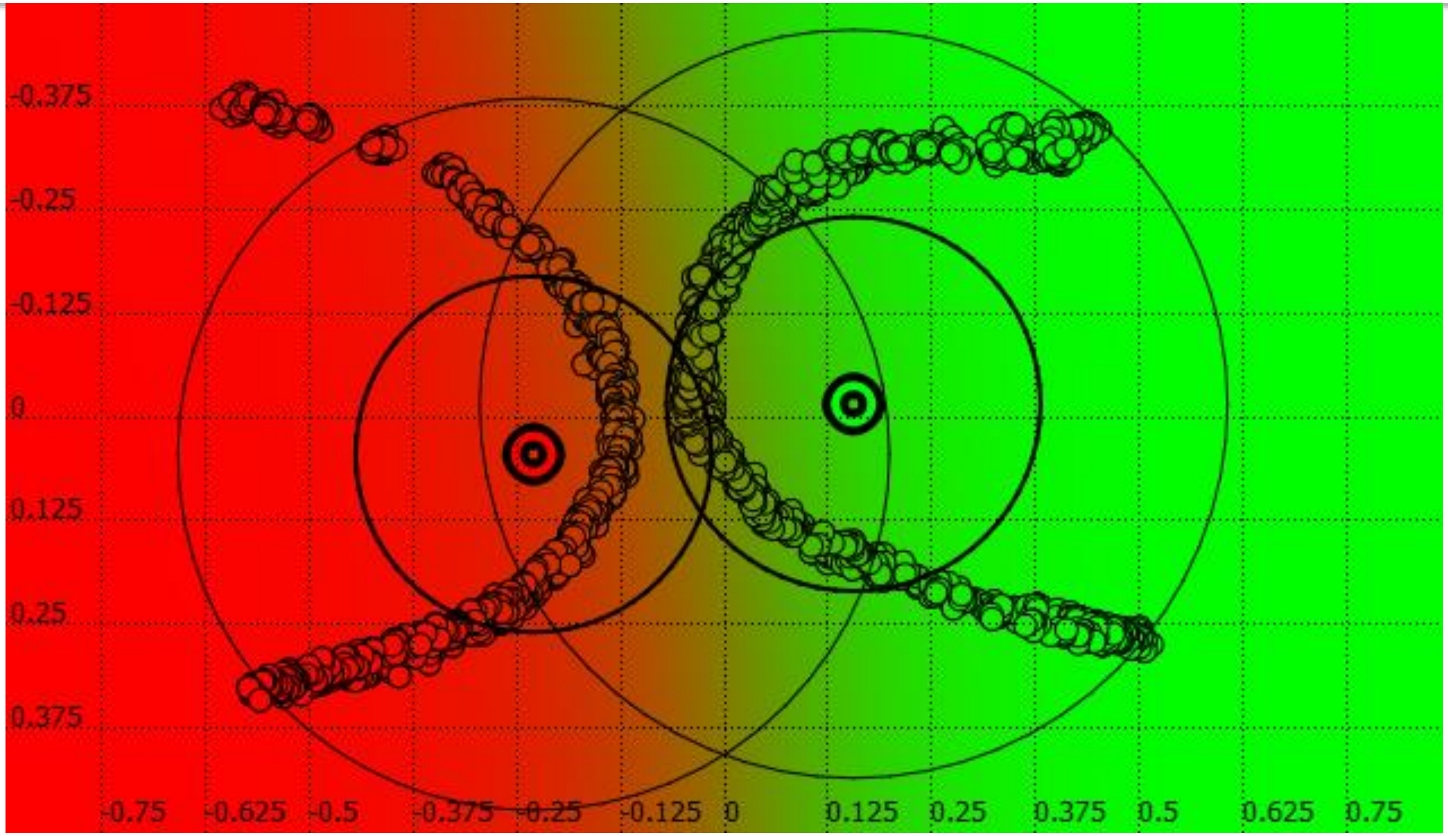
Classification with Gaussian Mixture Models

From Clustering to Classification with GMM



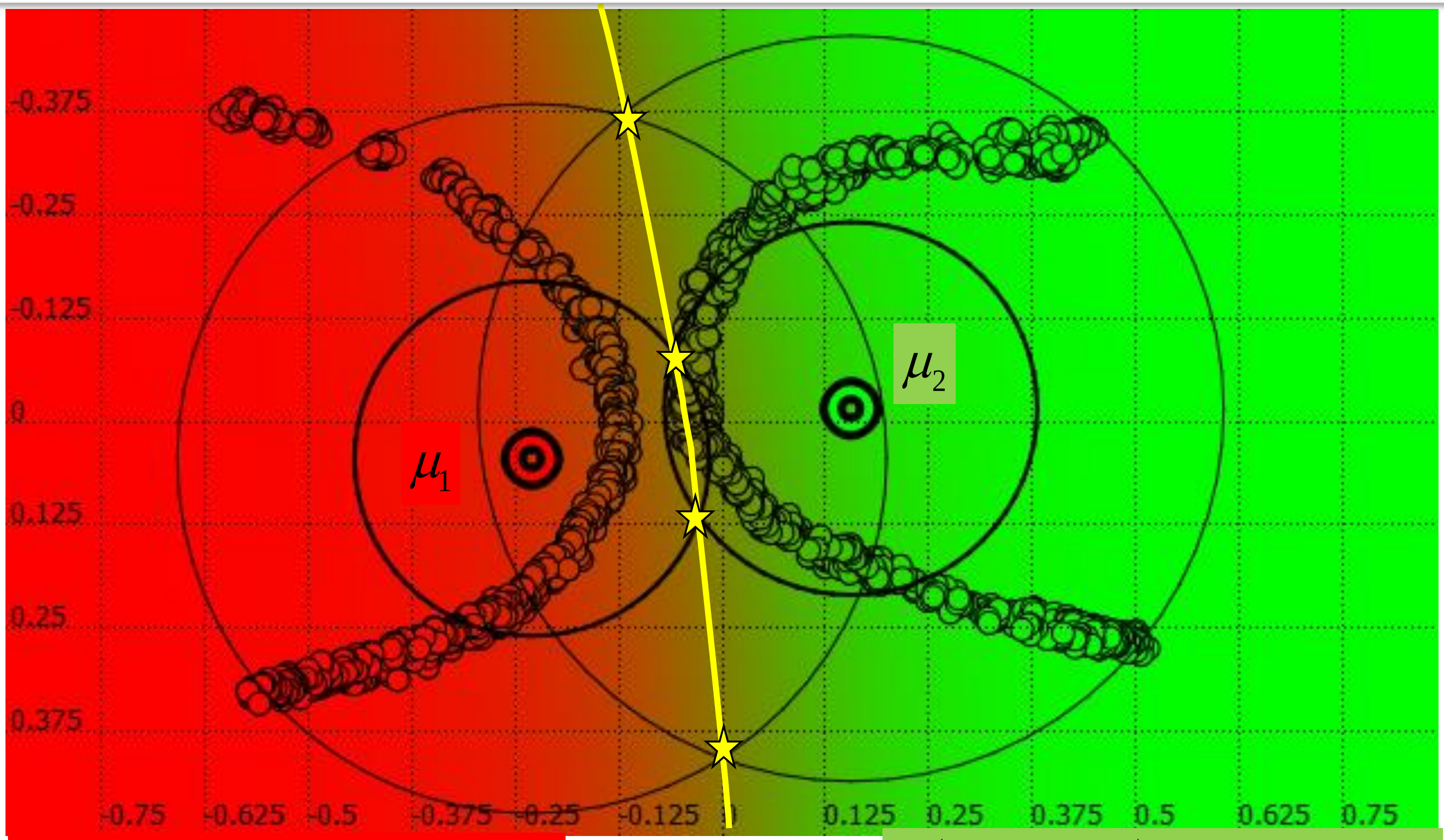
From Clustering to Classification with GMM

GMM clustering with two Gaussian functions with isotropic/spherical covariance



From Clustering to Classification with GMM

Need to determine the boundary across the clusters (classes)

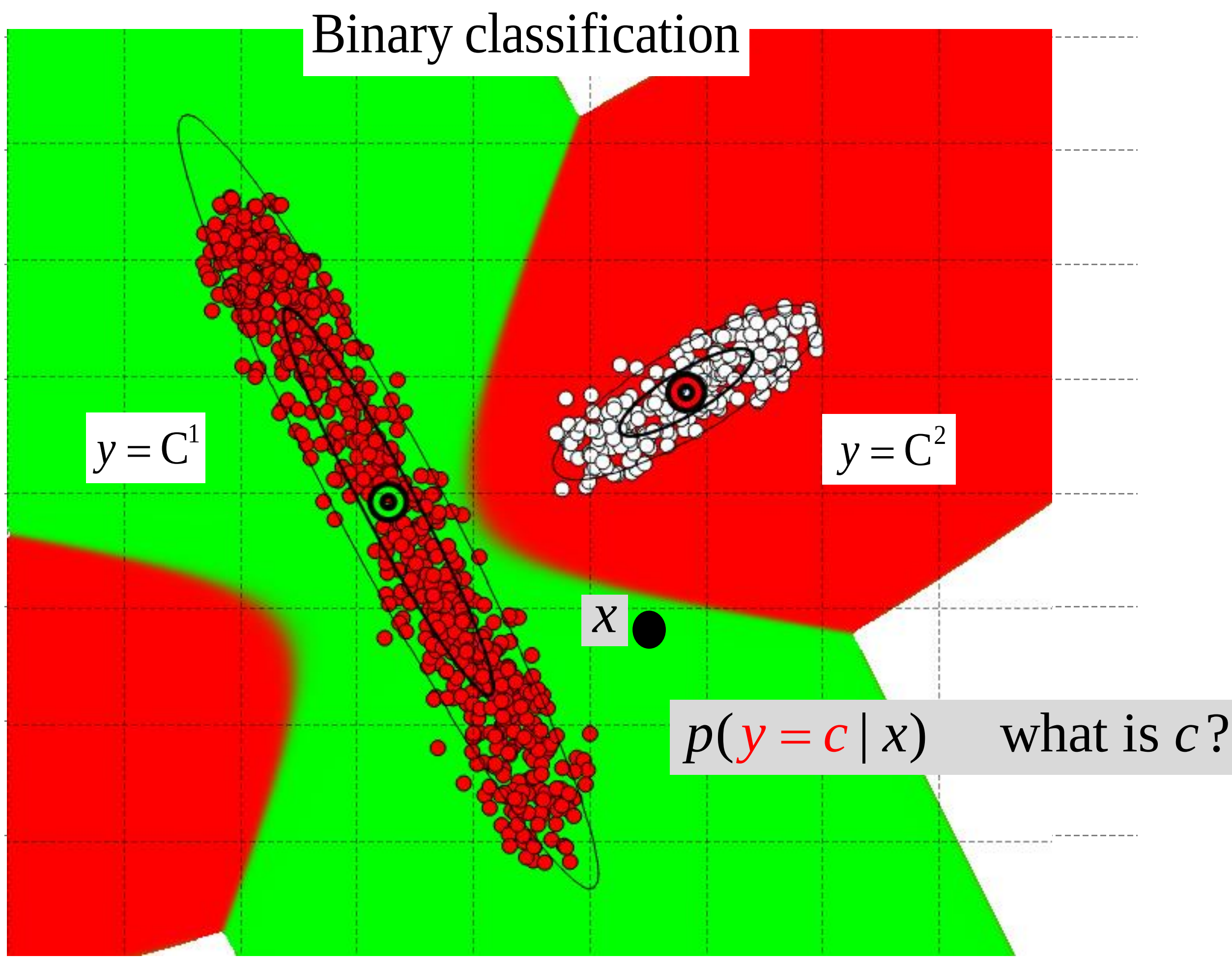


$p\left(x \mid y=C^1\right) \sim p\left(x \mid \mu_1, \sigma\right)$

$p\left(x \mid y=C^2\right) \sim p\left(x \mid \mu_2, \sigma\right)$

Boundary: all points x such that $p\left(y=C^1 \mid x\right) = p\left(y=C^2 \mid x\right)$

From Clustering to Classification with GMM



Determining the boundary across two pdf-s

We must determine the class with class label c that is most likely to have generated the datapoint x: $p(y=c \mid x)$

Bayes's rule: $$p(y=c \mid x) = \frac{p(y=c)\,p(x \mid y=c)}{p(x)}$$

$p\left(y=c\right)$: Probability of class c

$p(x \mid y=c)$: class conditional distribution of x
~ how the samples are distributed within class c.

$p\left(x\right)$: Marginal on x

Parameters estimation

We can estimate $p(y=c)$ and $p(x\mid y=c)$ from data.

$$p(y=c)=\frac{N_c}{N} \quad \text{where:} \quad \begin{array}{l} N_c : \text{num. of samples in class c} \\ N : \text{total number of samples} \end{array}$$

This can be omitted if classes are balanced (equally likely).

Fit a single Gaussian by maximum likelihood:

$$p(x\mid y=c) \sim p\left(x\mid \mu_c, \Sigma_c\right)$$

μ_c, Σ_c : mean and covariance matrix

$$\mu_c = \frac{1}{N_c}\sum_{i=1}^{N_c} x_i \qquad \Sigma_c = \frac{1}{N_c}\sum_{i=1}^{N_c}\left(x_i-\mu_c\right)\left(x_i-\mu_c\right)^T$$

Gaussian Discriminant Rule

2-class problem, conditional densities to belong to classes C^1 and C^2:

$$p\left(x \mid y=C^1\right) \sim p\left(x \mid \mu_1, \Sigma_1\right)=\frac{1}{(2\pi)^{N/2}\left|\Sigma_1\right|^{1/2}} e^{-(x-\mu_1)(\Sigma_1)^{-1}(x-\mu_1)^T}$$

$$p\left(x \mid y=C^2\right) \sim p\left(x \mid \mu_2, \Sigma_2\right)=\frac{1}{(2\pi)^{N/2}\left|\Sigma_2\right|^{1/2}} e^{-(x-\mu_2)(\Sigma_2)^{-1}(x-\mu_2)^T}$$

To determine the class label, compute optimal *Bayes classifier.*

$$\boxed{\text{A point } x \text{ belongs to class } C^1 \text{ if } p\left(y=C^1 \mid x\right)>p\left(y=C^2 \mid x\right)}$$

Optimal Bayes's classifier

$$p\left(y=C^{1} \mid x\right)>p\left(y=C^{2} \mid x\right) \qquad (1)$$

By Bayes: $p\left(y=C^{i} \mid x\right)=\dfrac{p\left(x \mid y=C^{i}\right) p\left(y=C^{i}\right)}{p(x)}, \quad i=1,2.$

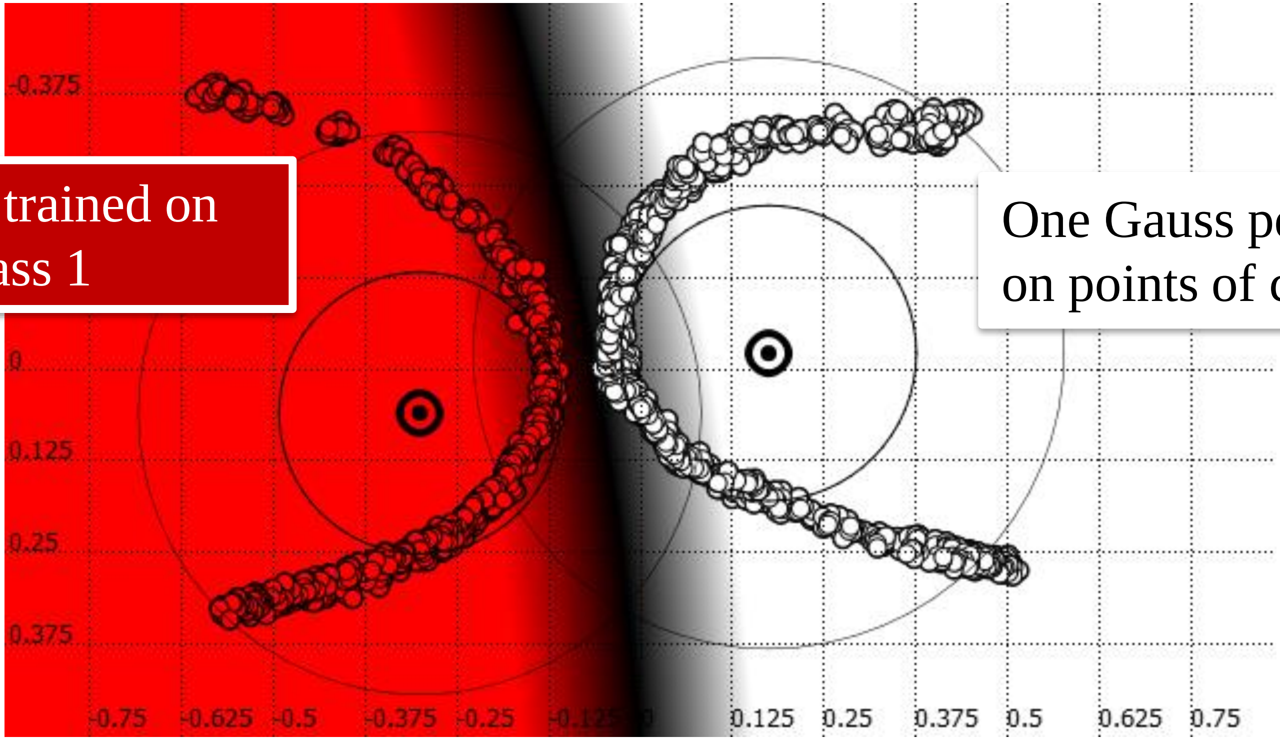
$$\frac{p\left(x \mid y=C^{1}\right) p\left(y=C^{1}\right)}{p(x)}>\frac{p\left(x \mid y=C^{2}\right) p\left(y=C^{2}\right)}{p(x)} \qquad (2)$$

Assuming equal class distribution, $p\left(y=C^{1}\right)=p\left(y=C^{2}\right)$ and replacing in Equation (2)

$$\Rightarrow \frac{p\left(x \mid y=C^{1}\right)}{p\left(x \mid y=C^{2}\right)}>1 \quad \Rightarrow \ln \left(\frac{p\left(x \mid y=C^{1}\right)}{p\left(x \mid y=C^{2}\right)}\right)>0$$

$$\Leftrightarrow \left(x-\mu^{1}\right)^{T}\left(\Sigma^{1}\right)^{-1}\left(x-\mu^{1}\right)+\log \left|\Sigma^{1}\right|<\left(x-\mu^{2}\right)^{T}\left(\Sigma^{2}\right)^{-1}\left(x-\mu^{2}\right)+\log \left|\Sigma^{2}\right|$$

Classification with two Gaussians



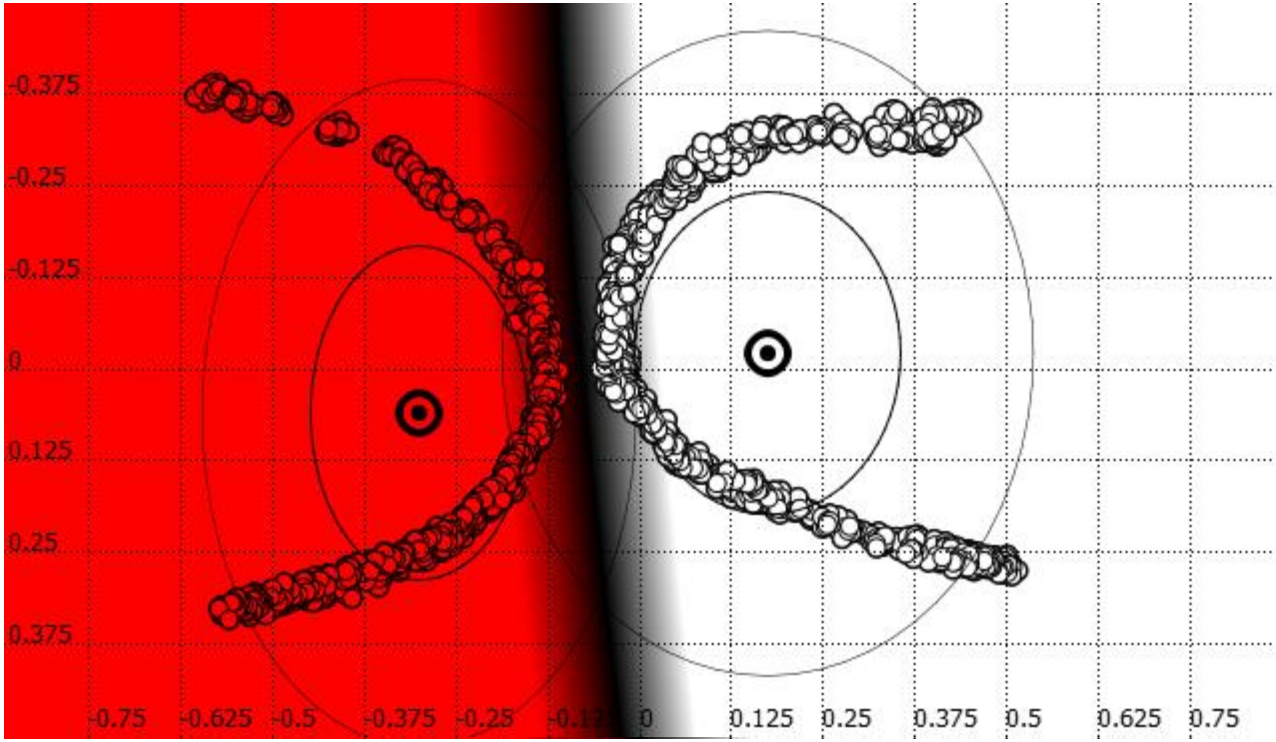
1 Gauss pdf trained on points of Class 1

One Gauss pdf trained on points of class 2

Example of binary classification using one Gauss pdf per class and Bayes rule (isotropic Gaussian functions)

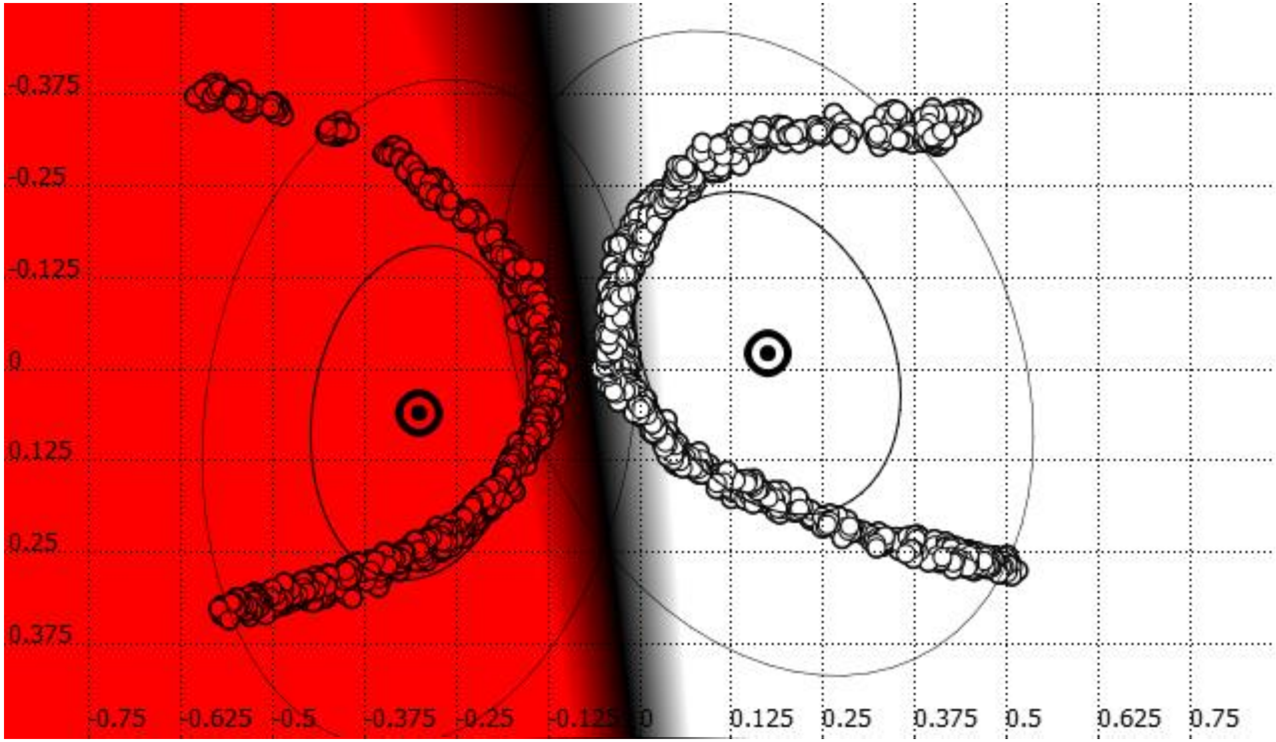
Train each Gaussian separately, using dataset of Class 1 for Gaussian 1 and dataset of class 2 for Gaussian 2

Classification with two Gaussians



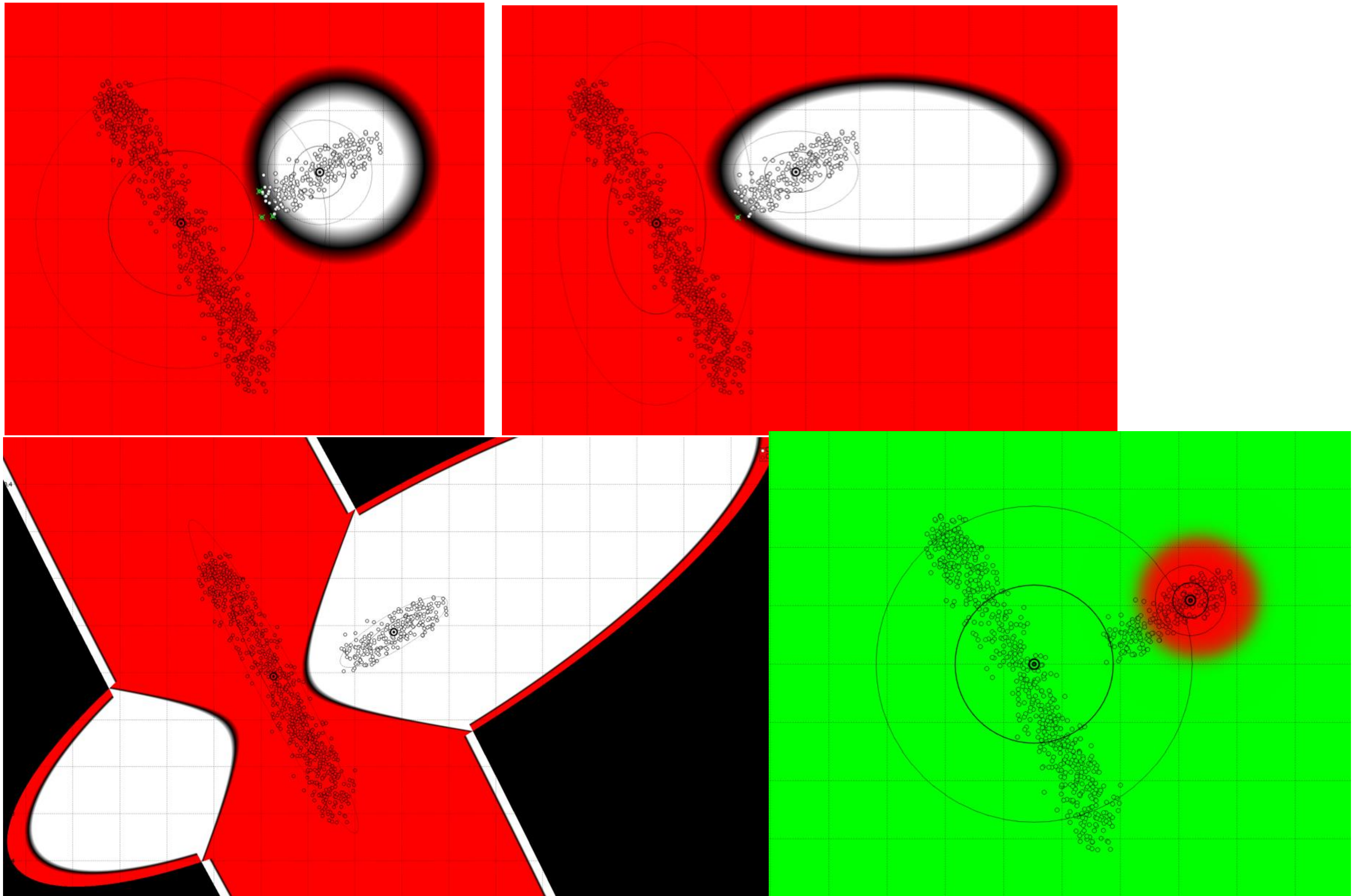
Example of binary classification using one Gauss pdf per class and Bayes rule (diagonal Gaussian functions)

Classification with two Gaussians



Example of binary classification using one Gauss pdf per class and Bayes rule (full covariance Gaussian functions)

Classification with two Gaussians



Maximum Likelihood Discriminant Rule

A maximum likelihood classifier chooses the class label that is the most likely.

Conditional density that a data point *x* has associated class label *y*=c is:

$$p_c(x) = p(x \mid y = c)$$

The maximum likelihood (ML) discriminant rule predicts the class of an observation *x* using:

$$c(x) = \arg\max_c p_c(x)$$

Maximum Likelihood Discriminant Rule for Multi-Classes

Muticlass problem with k=1…K classes, conditional densities for each class is a multivariate Gaussian:

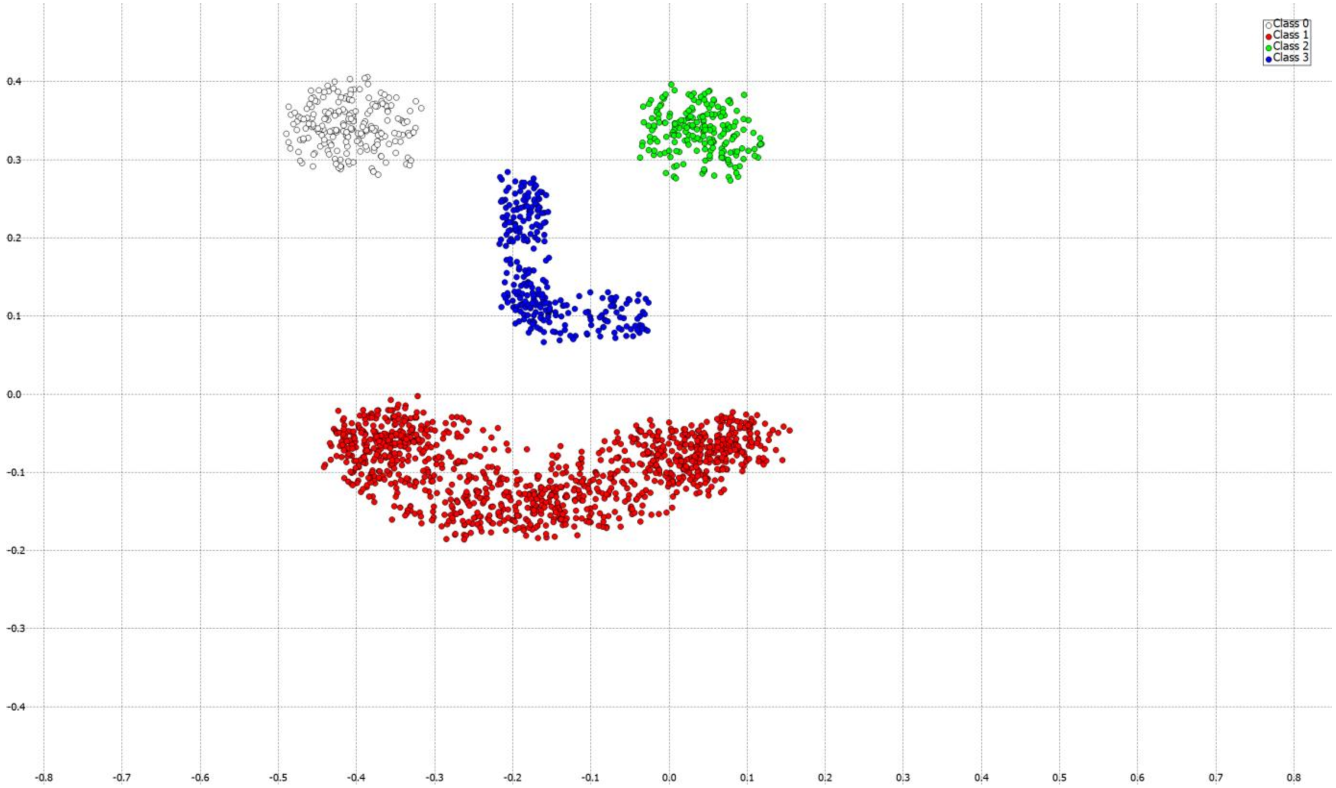
$$p(x \mid y=k) \sim p\left(x \mid \mu^k, \Sigma^k\right)$$

ML discriminant rule is minimum of minus the log-likelihood (equiv. to maximizing the likelihood):

$$C^k(x) = \arg\min_k \left\{ \left(x-\mu^k\right)\left(\Sigma^k\right)^{-1}\left(x-\mu^k\right)^T + \log\left|\Sigma^k\right| \right\}$$

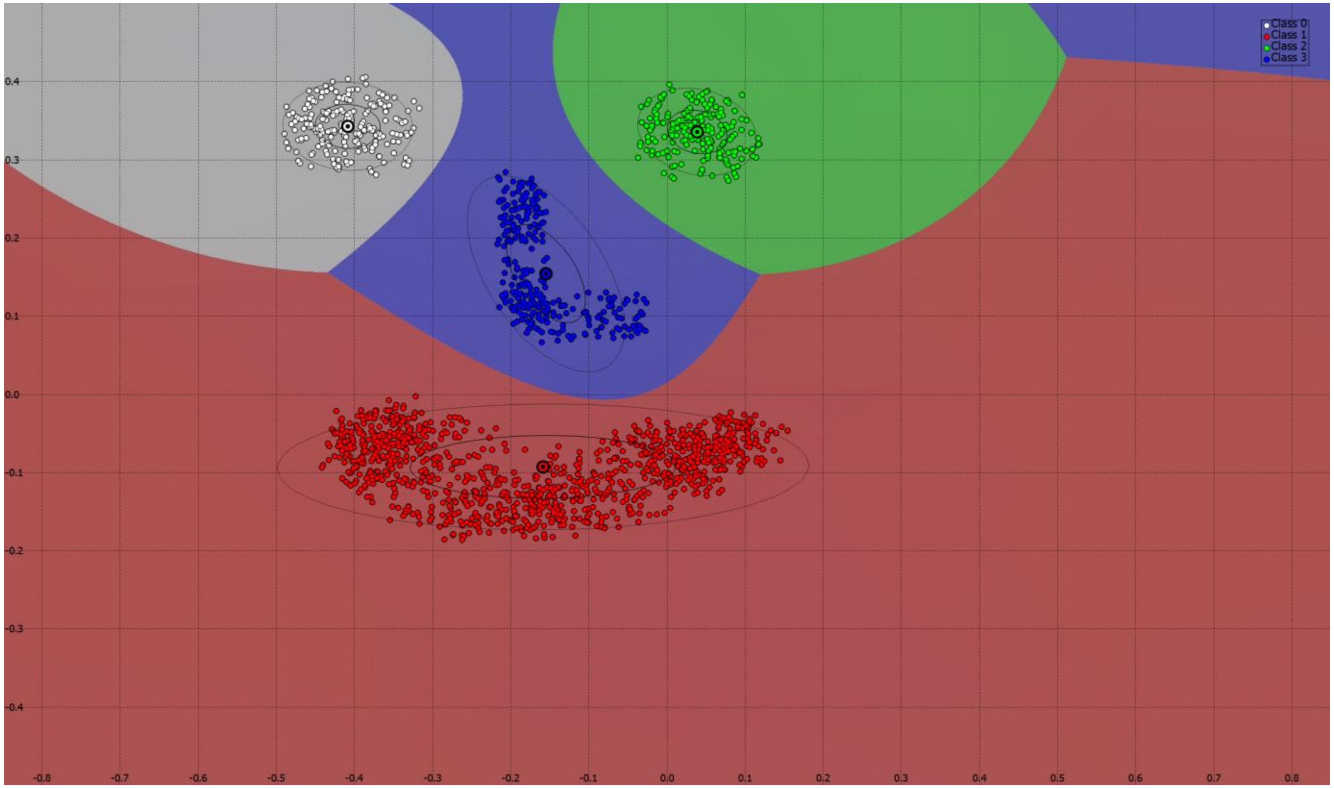
For multiple class classification: 1) Train one Gaussian per class; 2) choose the class that minimizes the above argument.

Maximum Likelihood Discriminant Rule for Multi-Classes



Example of 4-classes classification using
4 Gaussian distributions

Maximum Likelihood Discriminant Rule for Multi-Classes



Example of 4-classes classification using
4 Gaussian distributions

Classification with GMM-s

Muti-class problem with $y = C^1 \dots C^K$ classes, and each class is modeled with

a GMM composed of L multivariate Gaussian functions:

$$p\left(x \mid y = C^k\right) \sim \sum_{l=1}^{L} \alpha_l N\left(\mu_l^k, \Sigma_l^k\right)$$

ML discriminant rule is the minimum of -log-likelihood :

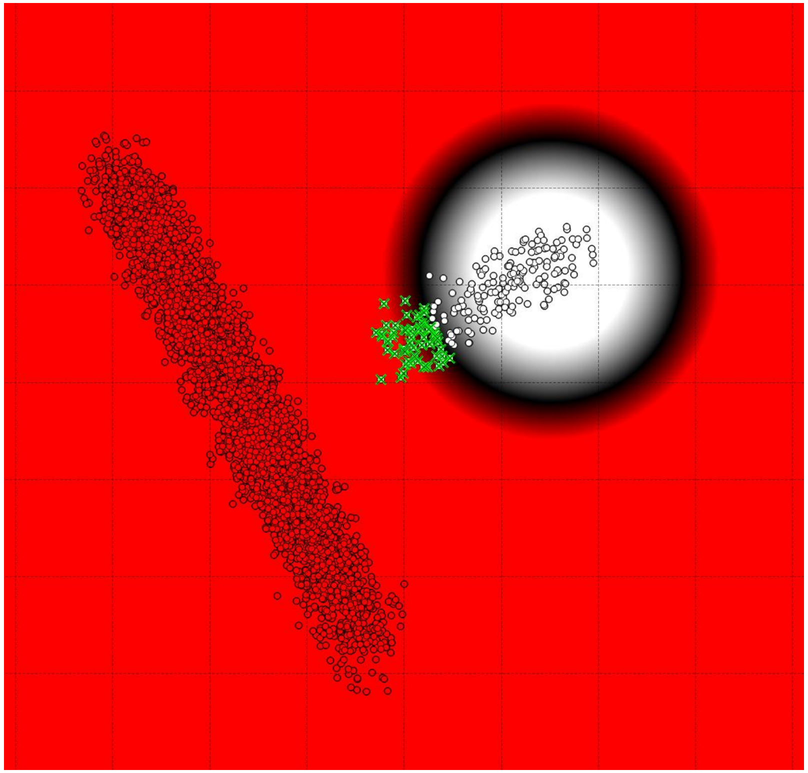
$$y = \arg\min_{i=1\dots K} \left\{ -\log\left(p\left(x \mid y = C^i\right)\right)\right\}$$

Each of the K GMM fits each class separately
→ No information on the class probability *p(y=i)*.
When comparing the likelihood *p(x|y)* directly, as above, this is equivalent to assuming equal class probability.

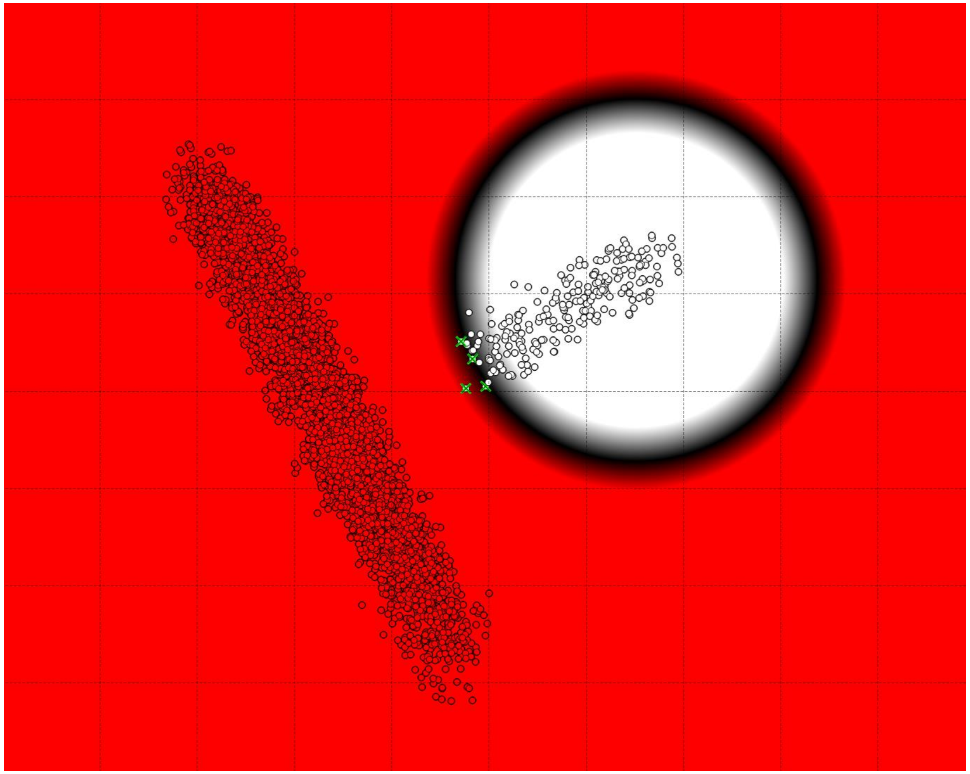
Classification with GMM-s

Unbalanced dataset:

3950 Samples
3720 Positives
230 Negatives



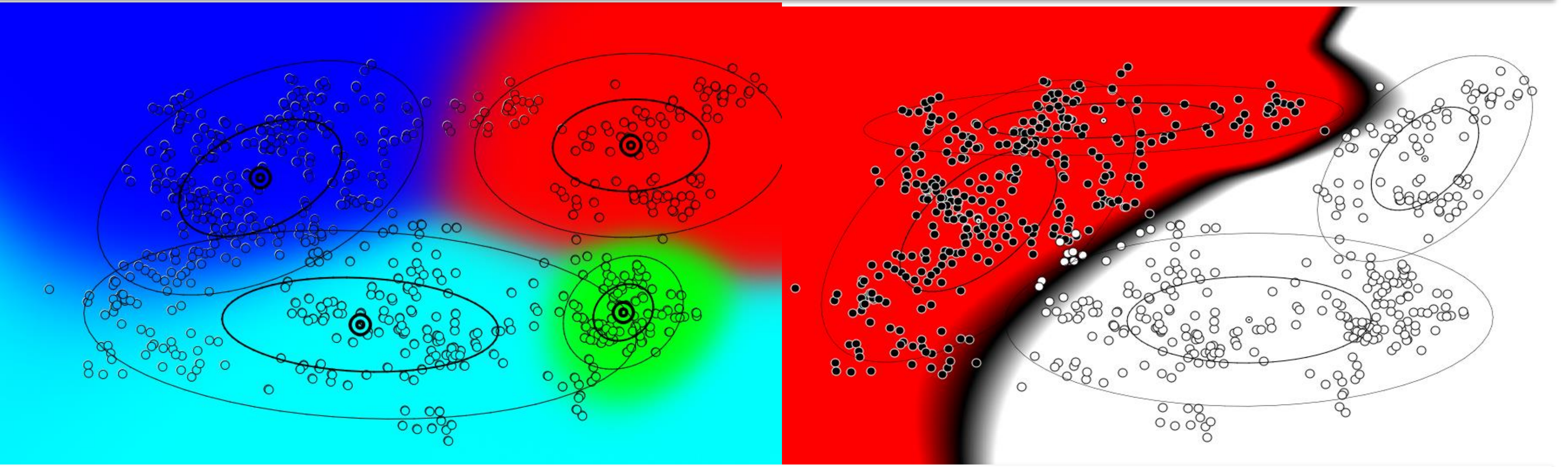
Unbalanced class distribution



Force equal class distribution

Classification with GMM-s: summary

- ❑ Clustering with GMM does not have the class labels. It ends up merging the classes when these are too tight.
- ❑ Classification has class labels and can hence determine boundary between two tight groups, but this comes at the price of labeling the data.



- ❑ Classification in GMM can be expressed as a Maximum Likelihood problem.
- ❑ It has a closed form solution for each datapoint.
- ❑ The boundary across classes can be very complex and depends on the complexity of the GMM.
- ❑ Unbalanced class distribution can be compensated for if we can estimate it.