UNIVERSITÀ DEGLI STUDI DI PADOVA

Mean shift

Stefano Ghidoni

Agenda

- Introduction to mean shift
- An efficient approach to density estimation
  - Density gradient estimation
- Visualization of the mean shift procedure

Mean shift – key idea

- Mean shift is a tool for finding stationary points (i.e. peaks)
- Key idea:
  - Find peaks in high-dimensional data distribution without computing the distribution function explicitly
- It implicitly models the distribution using a smooth continuous non-parametric model

Density estimation

- Assumption: the data points are sampled from an underlying density function (PDF)
- Direct PDF estimation can be made by means of functions like:

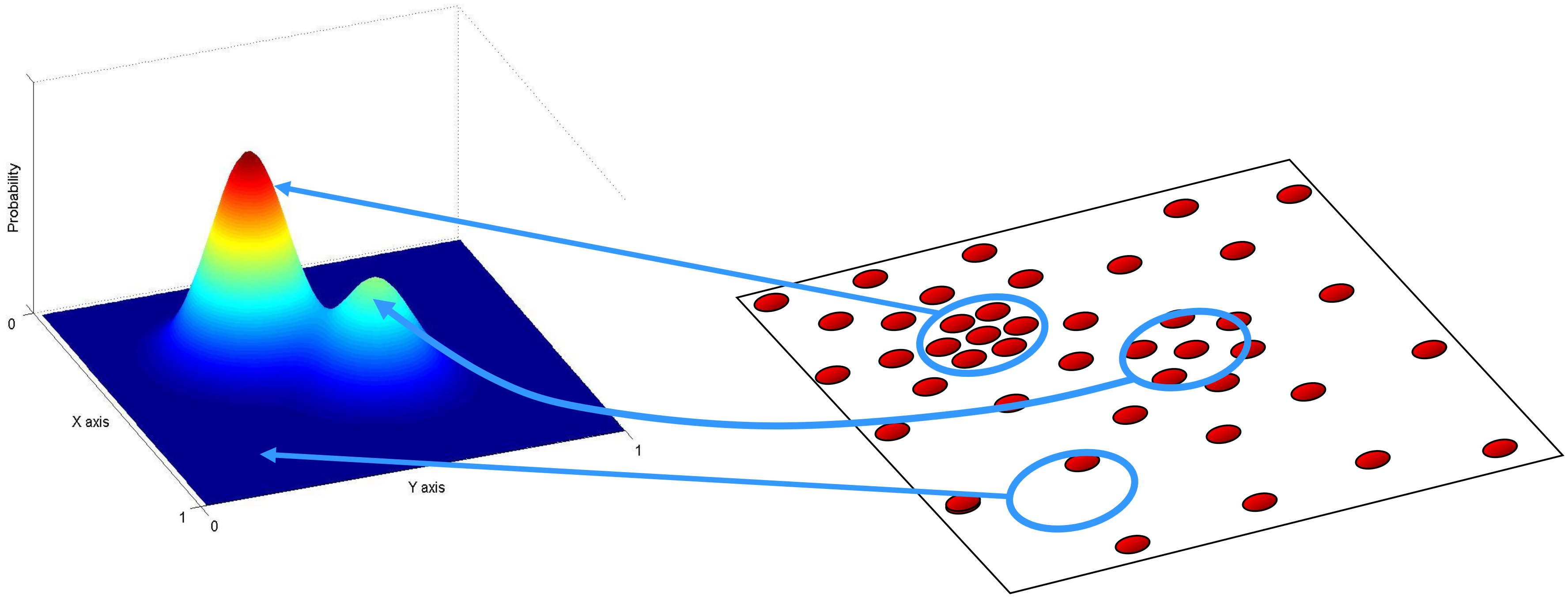
$$f(\boldsymbol{x}) = \sum_i c_i e^{-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_i)^2}{2\sigma_i^2}}$$

- Needs estimates!

To be estimated

Density estimation

- Data point density is an implicit description of PDF value
  - Non-parametric estimation

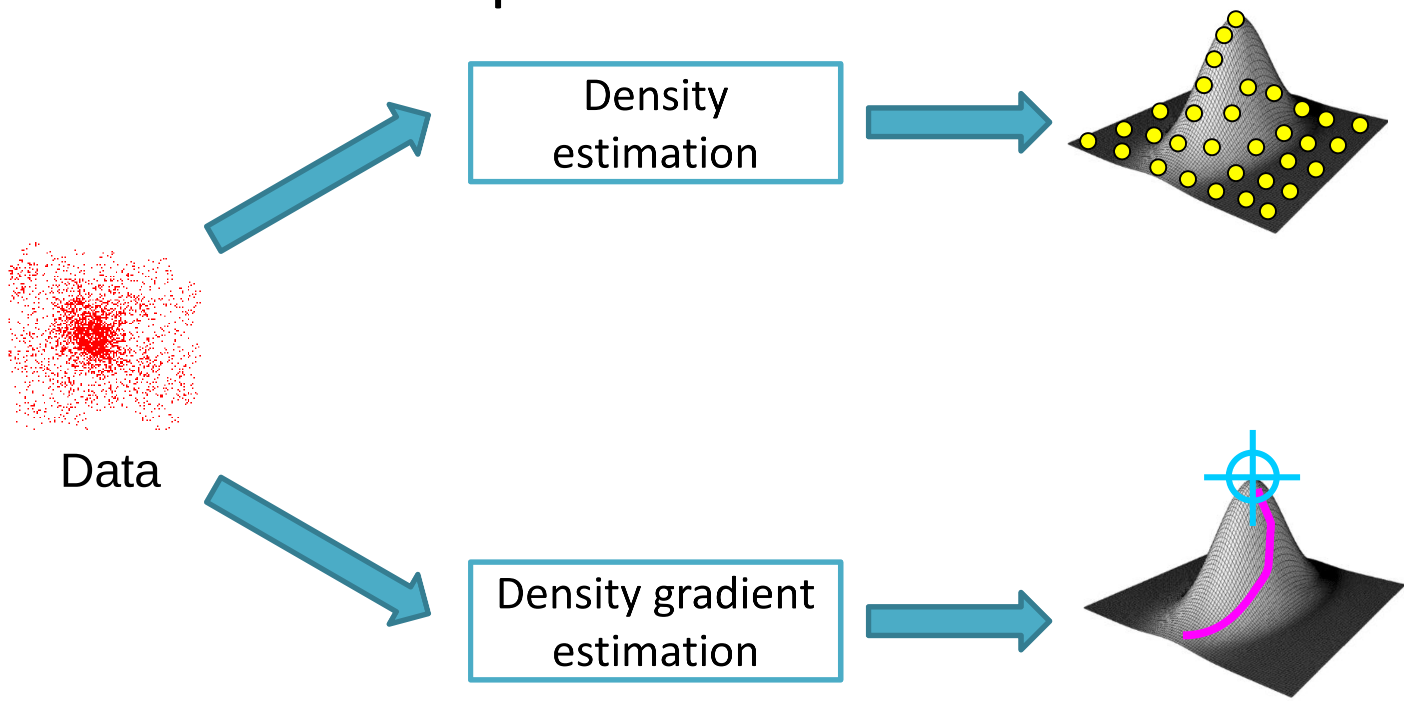


Assumed underlying PDF

Data samples

Density estimation

- How to avoid evaluating the whole density function?
- Do not estimate the PDF
  - Estimate its gradient instead!
- Mean shift is a steepest-ascend method



Density gradient estimation

Kernel density estimation → Kernel density gradient estimation

- Kernel density gradient estimation:

$$\nabla f(\boldsymbol{x}) = \frac{1}{Nr^n} \sum_{i=1}^{N} \nabla K(\boldsymbol{x} - \boldsymbol{x}_i)$$

- Recall:

$$K(\boldsymbol{x} - \boldsymbol{x}_i) = c_k k\left(\frac{\|\boldsymbol{x} - \boldsymbol{x}_i\|^2}{r^2}\right)$$

Density gradient estimation

- $\nabla f(\boldsymbol{x})$ is the multiplication of

$$c_k k'\left(\frac{\|\boldsymbol{x}-\boldsymbol{x}_i\|^2}{r^2}\right)$$

- And the derivative of the inner function

$$\frac{2}{r^2}\|\boldsymbol{x}-\boldsymbol{x}_i\|$$

Density gradient estimation

- Now define

$$\boldsymbol{y}_i = \frac{1}{r^2}\|\boldsymbol{x} - \boldsymbol{x}_i\|^2$$

- And

$$g(a) = -k'(a)$$

- Where the minus sign is used to express in terms of $(\boldsymbol{x}_i - \boldsymbol{x})$

Density gradient estimation

- The derivative $\nabla f(\boldsymbol{x})$ can then be written as

$$\nabla f(\boldsymbol{x}) = \frac{2c_k}{Nr^{n+2}} \sum_{i=1}^{N} [(\boldsymbol{x}_i - \boldsymbol{x}) \cdot g(\boldsymbol{y}_i)]$$

$$= \frac{2c_k}{Nr^{n+2}} \left(\sum_{i=1}^{N} [\boldsymbol{x}_i g(\boldsymbol{y}_i)] - \boldsymbol{x} \cdot \sum_{i=1}^{N} g(\boldsymbol{y}_i) \right)$$

$$= \frac{2c_k}{r^2 c_g} \left[\frac{c_g}{Nr^n} \sum_{i=1}^{N} g(\boldsymbol{y}_i) \right] \left[\frac{\sum_{i=1}^{N} \boldsymbol{x}_i g(\boldsymbol{y}_i)}{\sum_{i=1}^{N} g(\boldsymbol{y}_i)} - \boldsymbol{x} \right]$$

- Where the constant c_g normalizes the integral in the feature space when $g(\cdot)$ is used as a kernel

Density gradient estimation

- Considering the last formulation:

$$\frac{2c_k}{r^2 c_g}\left[\frac{c_g}{Nr^n}\sum_{i=1}^{N} g(\boldsymbol{y}_i)\right]\left[\frac{\sum_{i=1}^{N}\boldsymbol{x}_i g(\boldsymbol{y}_i)}{\sum_{i=1}^{N} g(\boldsymbol{y}_i)}-\boldsymbol{x}\right]$$

Constant term

Kernel density estimator using a kernel function $g(\cdot)$

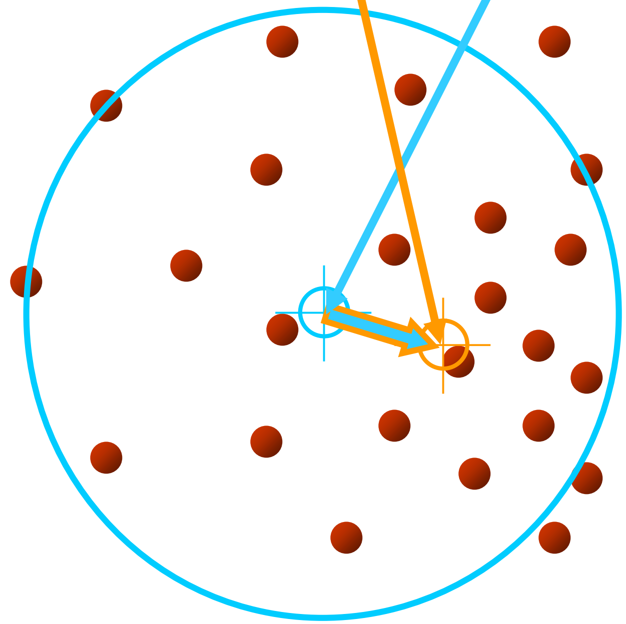
Mean shift vector

The mean shift vector starts at vector $\boldsymbol{x}$

- We can rewrite the previous equation as:
$$\nabla f_k(\boldsymbol{x}) = \mathrm{A} \cdot f_g(\boldsymbol{x}) \cdot m_g(\boldsymbol{x})$$
- Which can be rearranged as:
$$m_g(\boldsymbol{x}) = \frac{\nabla f_k(\boldsymbol{x})}{\mathrm{A} \cdot f_g(\boldsymbol{x})}$$
- This shows that the mean shift vector proceeds in the direction of the gradient

Mean shift vector

$$\boldsymbol{m}(\boldsymbol{x}) = \left[\boxed{\frac{\sum_{i=1}^{N} \boldsymbol{x}_i g(\boldsymbol{y}_i)}{\sum_{i=1}^{N} g(\boldsymbol{y}_i)}} - \boldsymbol{x}\right]$$

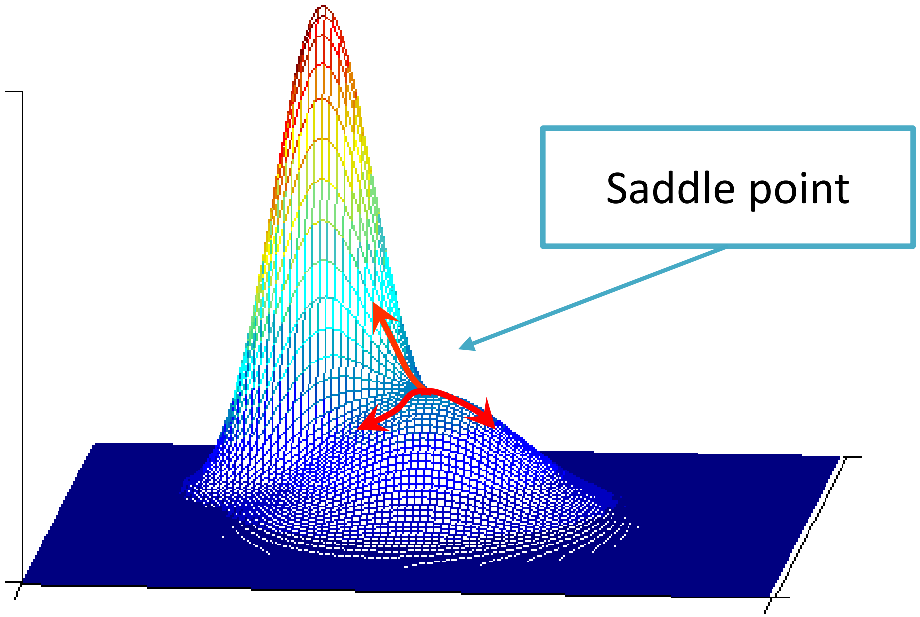


Mean shift procedure

- The mean shift vector is the direction to follow to find the mode
- Iterative procedure:
  - Compute mean shift vector $\boldsymbol{m}(\boldsymbol{x})$
  - Move the kernel window by $\boldsymbol{m}(\boldsymbol{x})$
  - Stop when the gradient is close to 0

Mode pruning

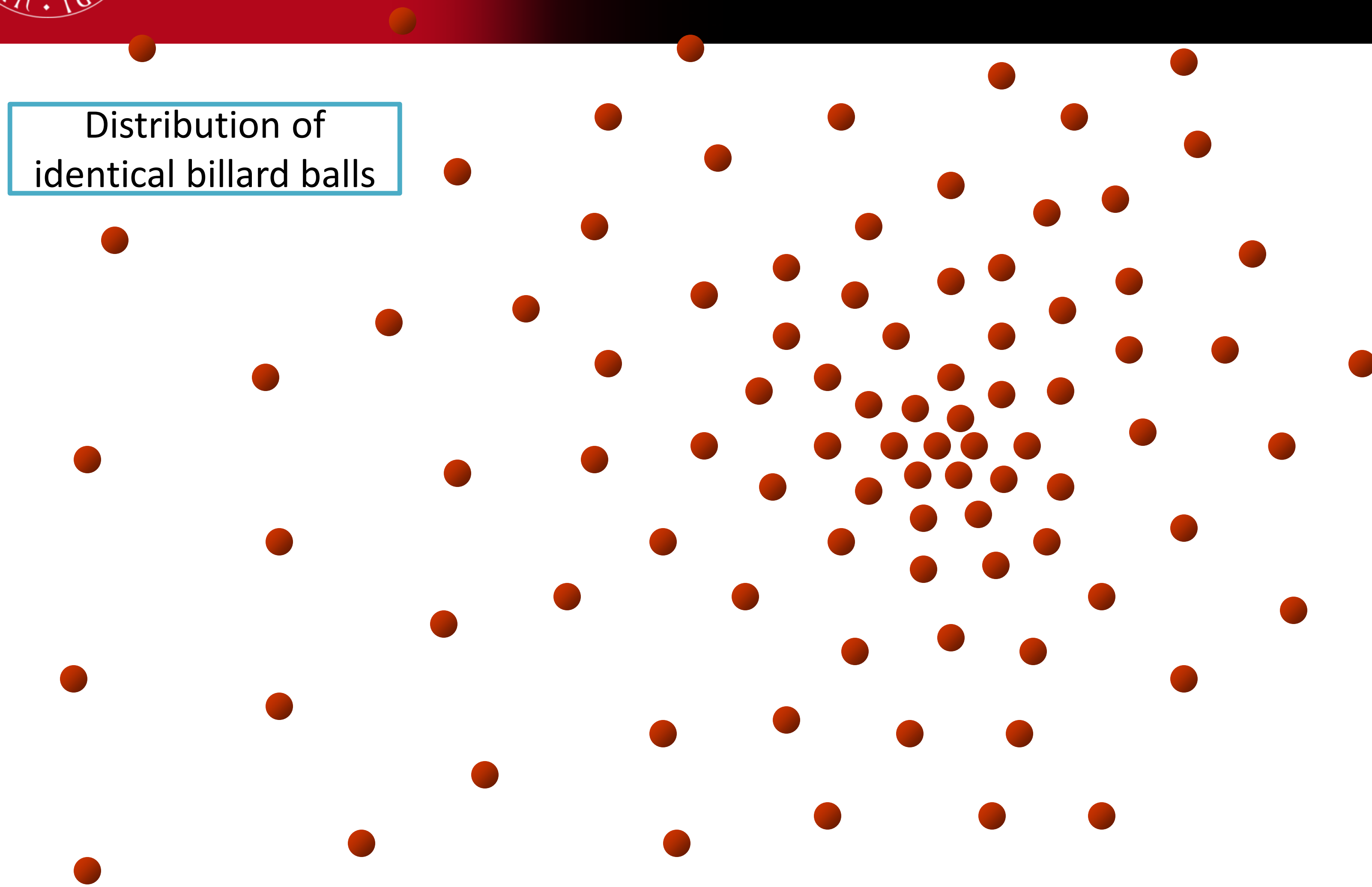
- Saddle points have zero-gradient
  - Unstable locations
  - A small perturbation cause the process to move away



Mean shift iterative procedure

- The mean shift vector is the direction to follow to find the mode
- Iterative procedure:
  - Compute mean shift vector $\boldsymbol{m}(\boldsymbol{x})$
  - Move the kernel window by $\boldsymbol{m}(\boldsymbol{x})$
  - Stop when the gradient is close to 0
  - Prune the mode by perturbation

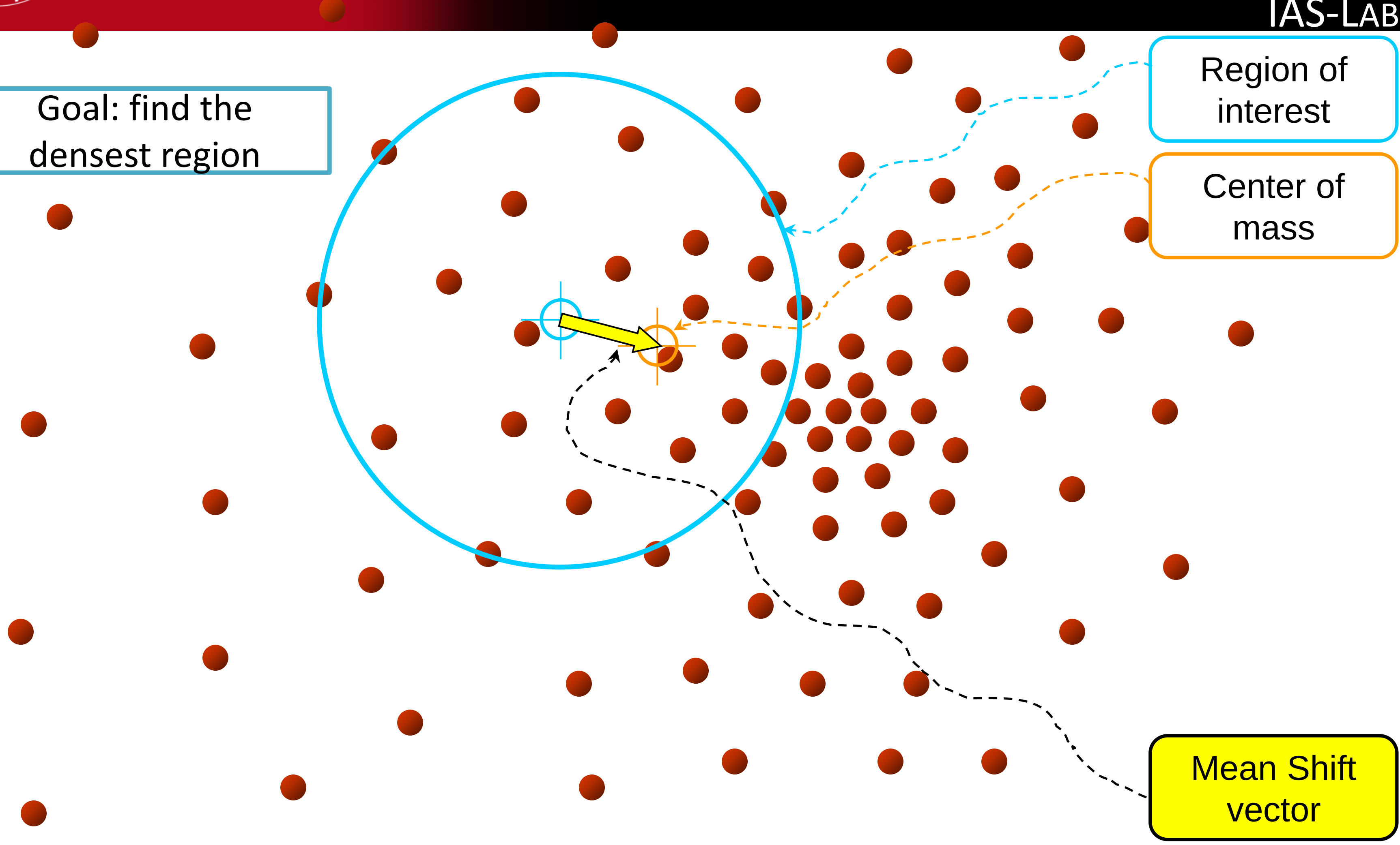
Mean shift procedure



Taken from Y. Ukrainitz & B. Sarel

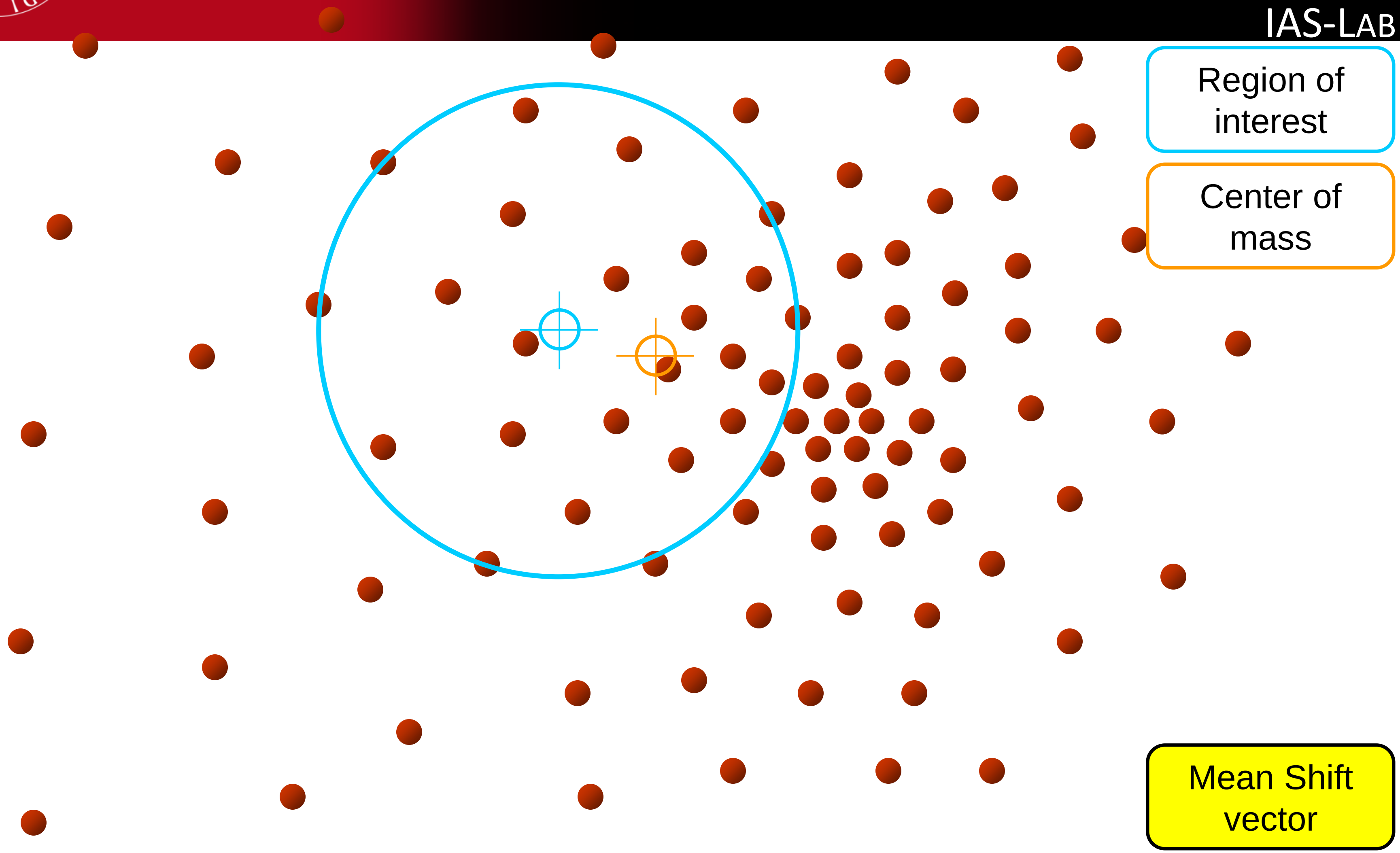
Mean shift procedure

Goal: find the densest region

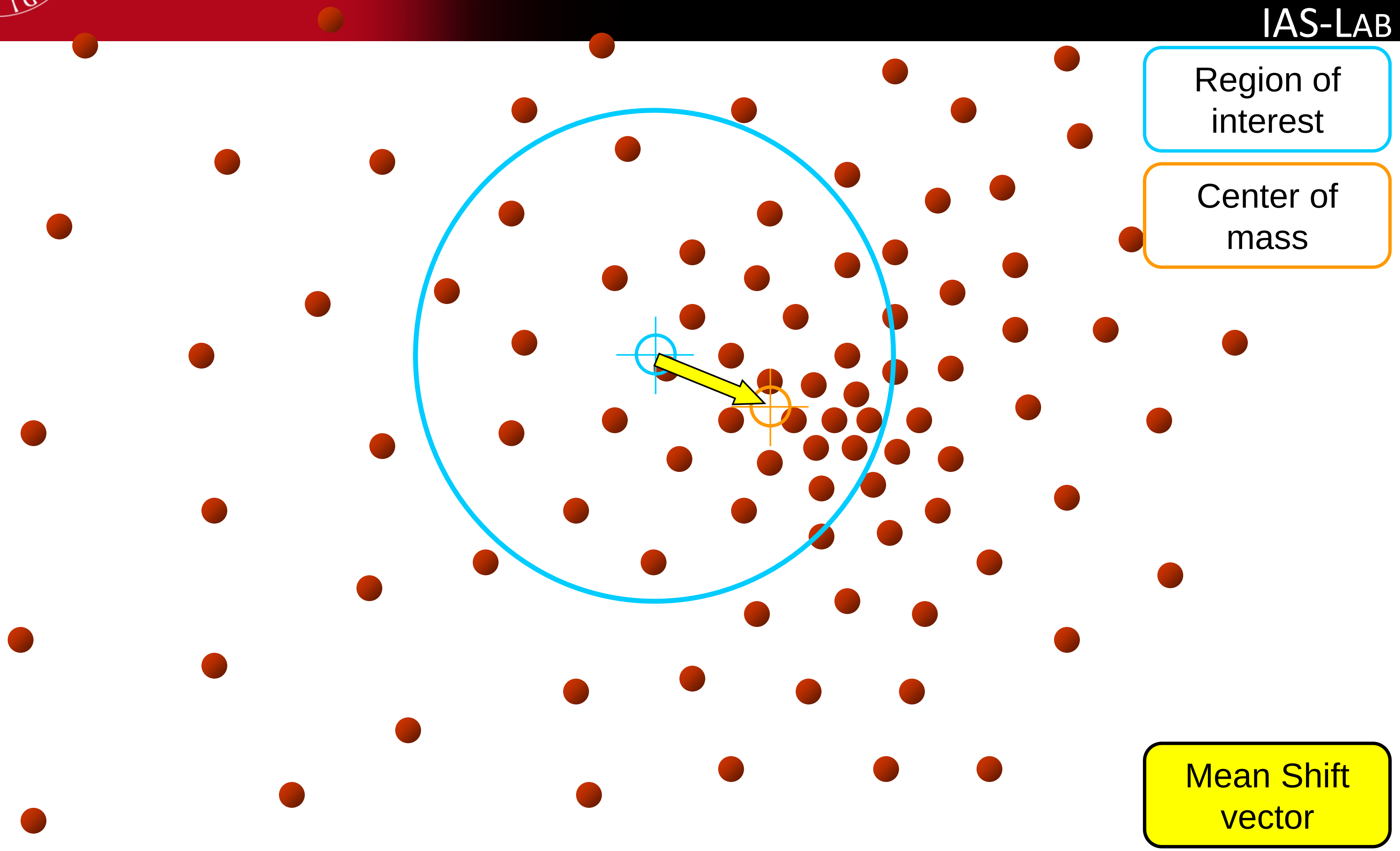


Taken from Y. Ukrainitz & B. Sarel

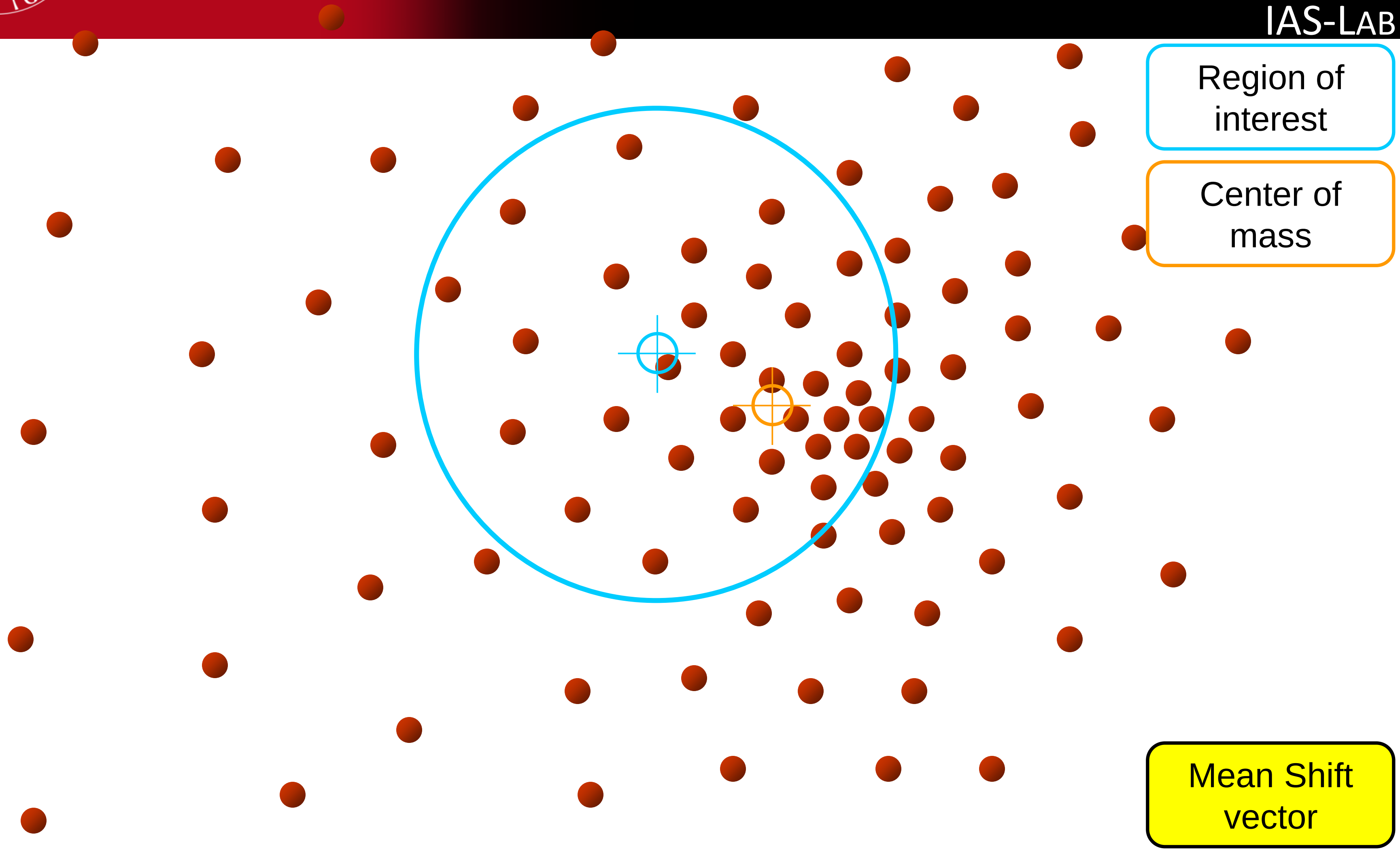
Mean shift procedure



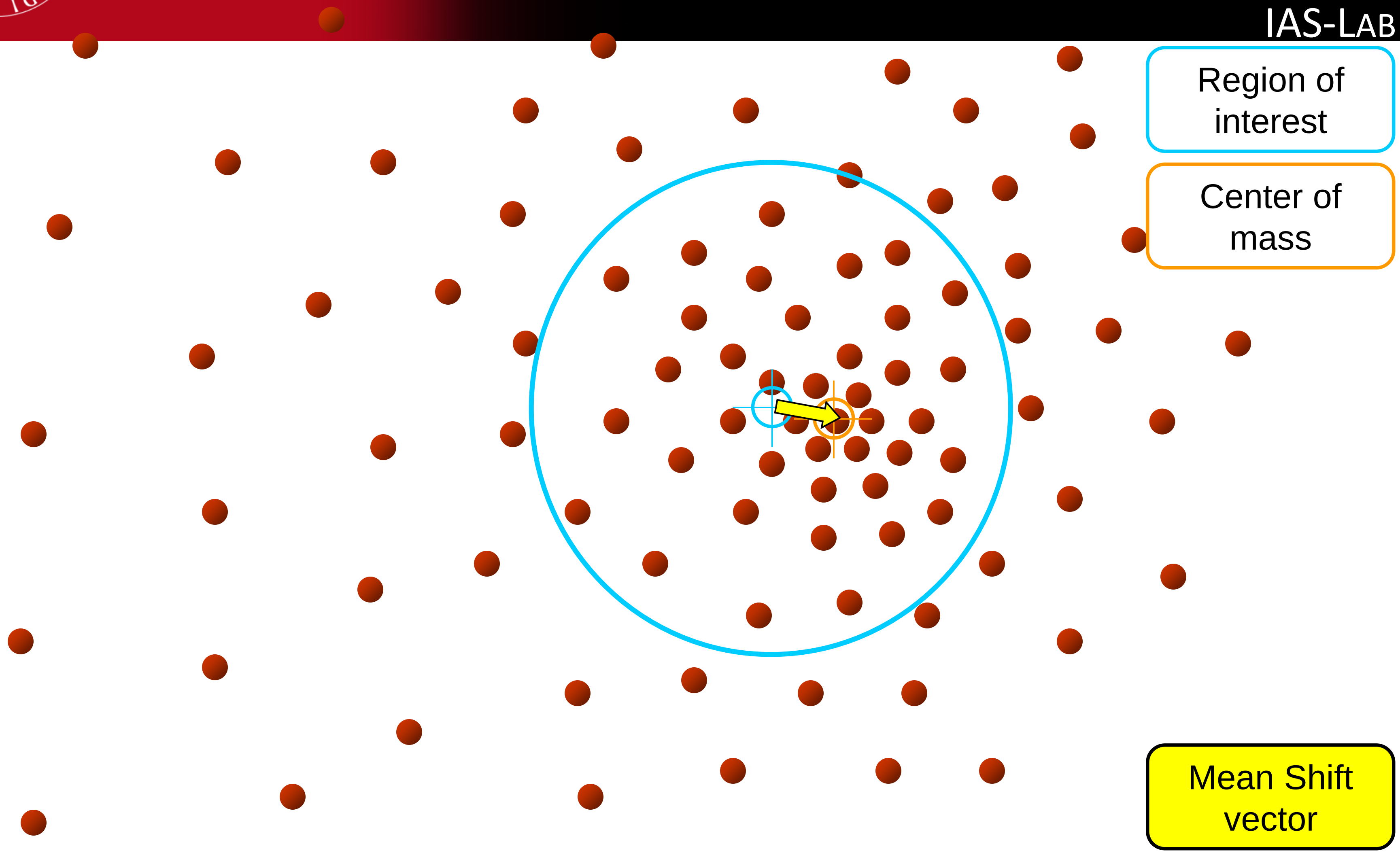
Mean shift procedure



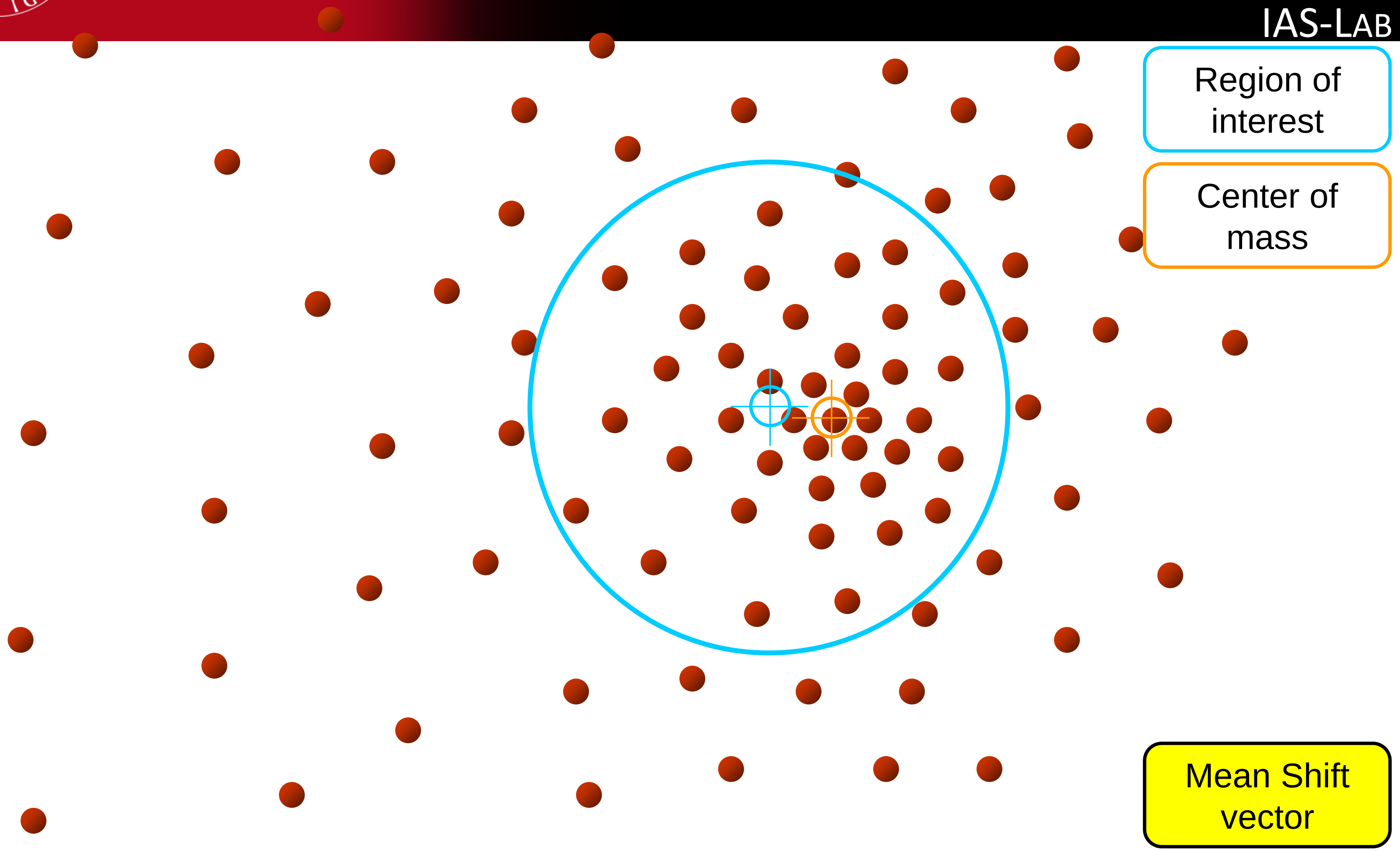
Mean shift procedure



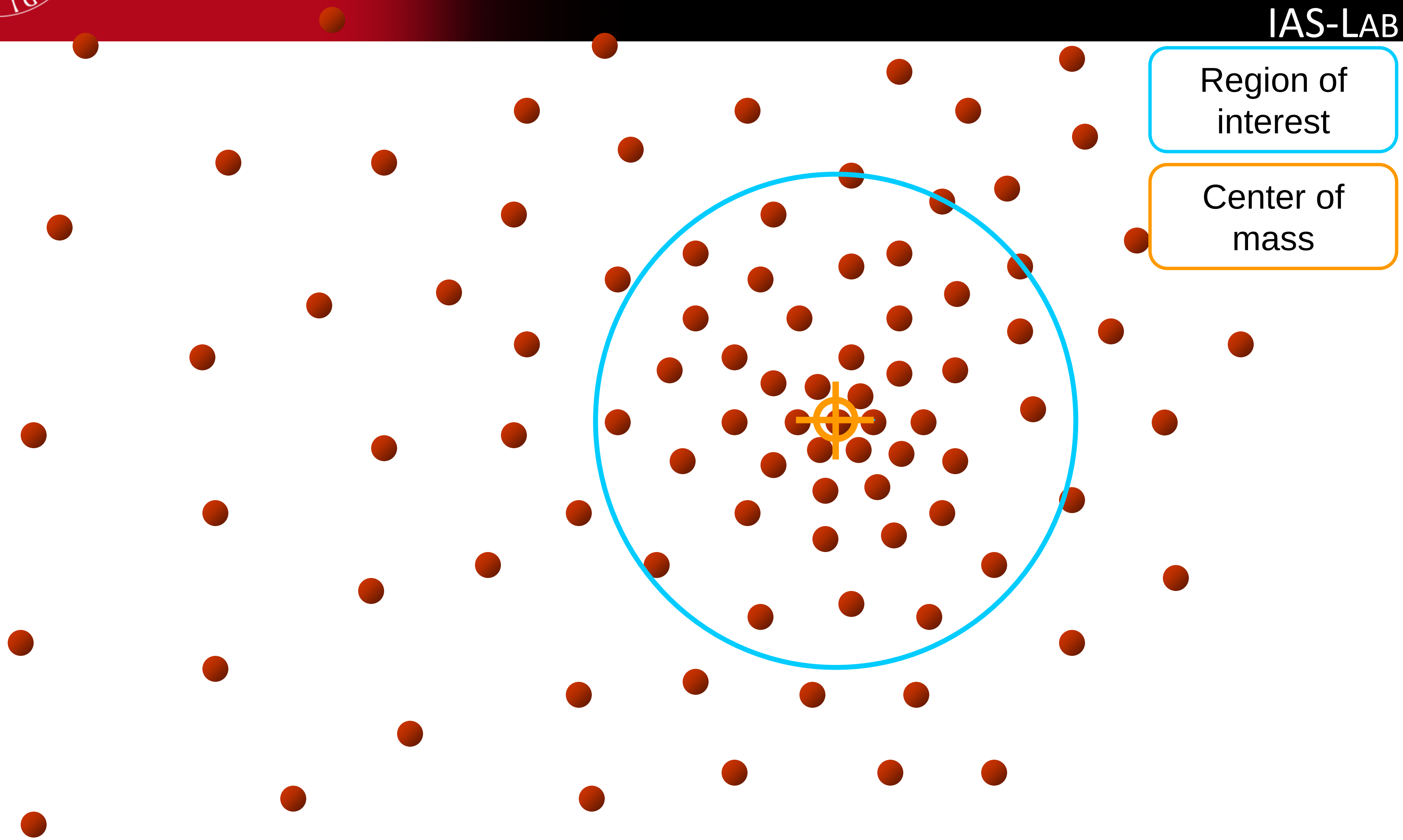
Mean shift procedure



Mean shift procedure

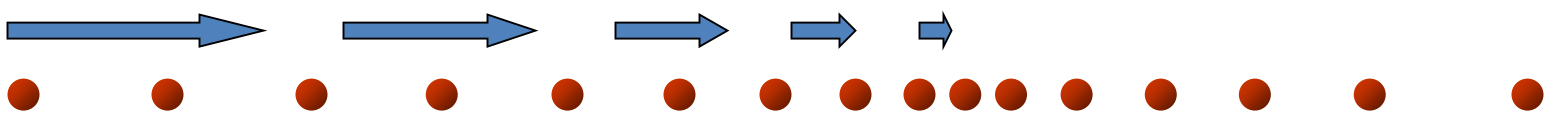


Mean shift procedure



Mean shift properties

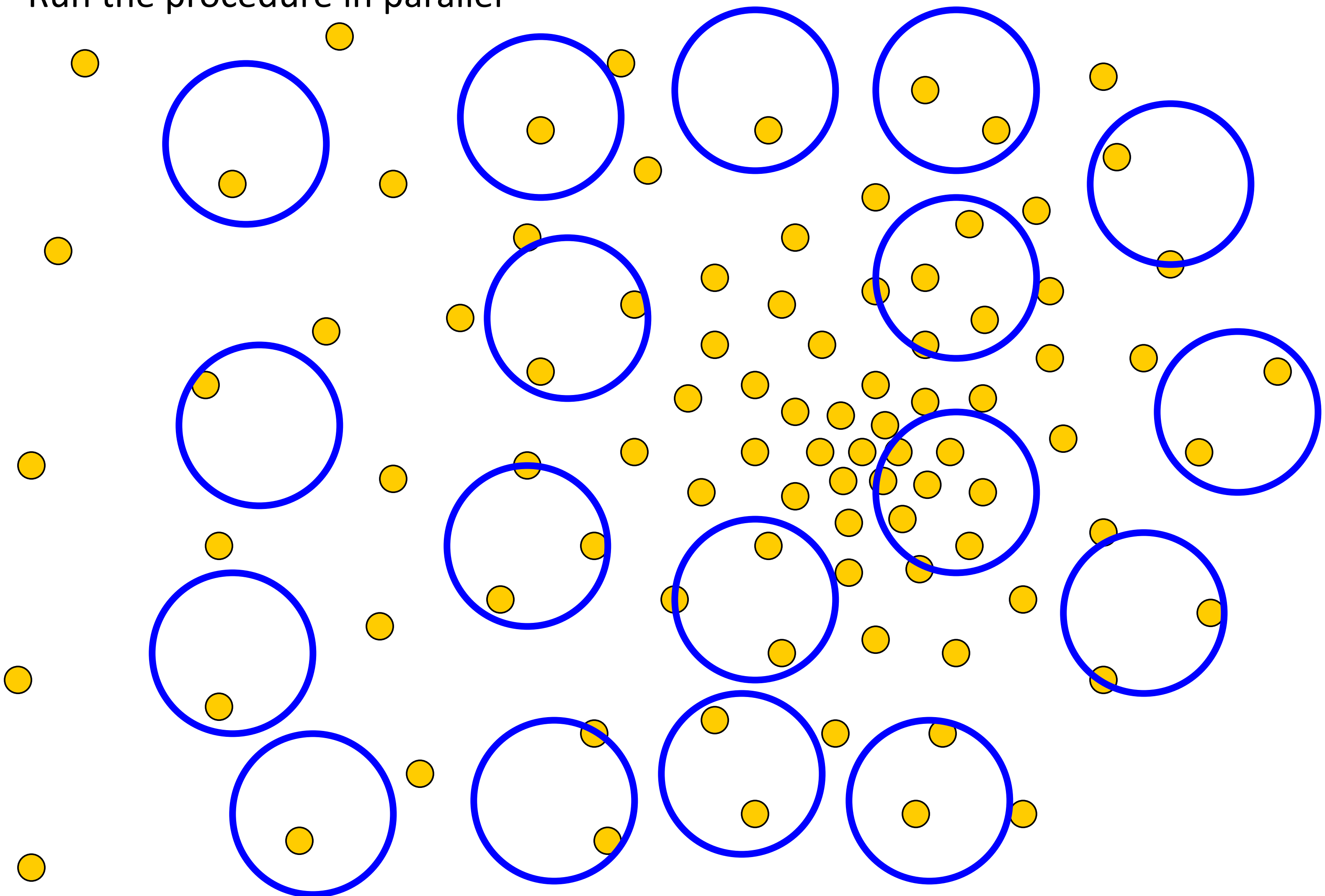
- Automatic convergence speed
  - The mean shift vector magnitude depends on the gradient



- Steps are smaller towards a mode
- Convergence depends on appropriate step size

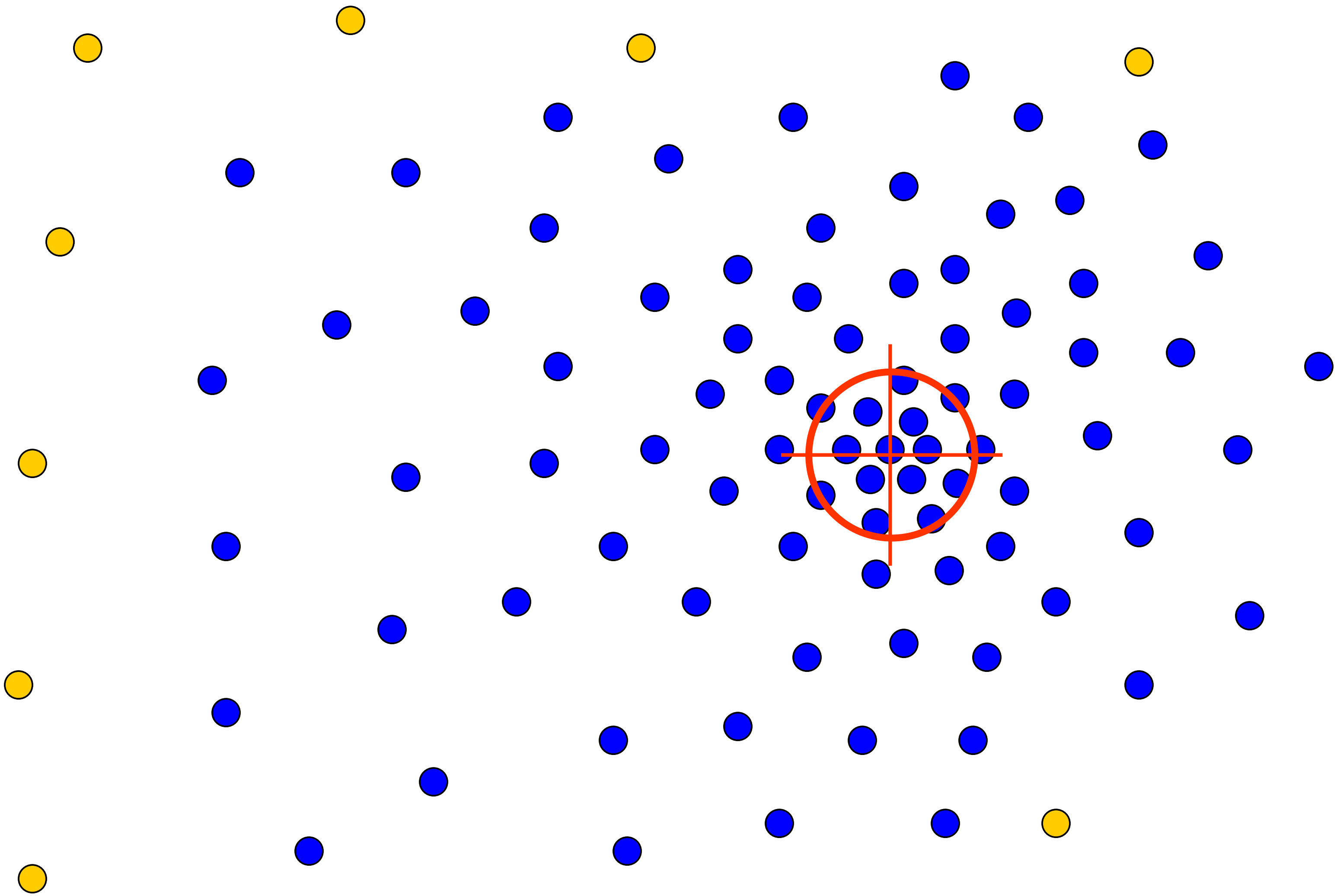
Optimization

- Divide the space into windows
- Run the procedure in parallel



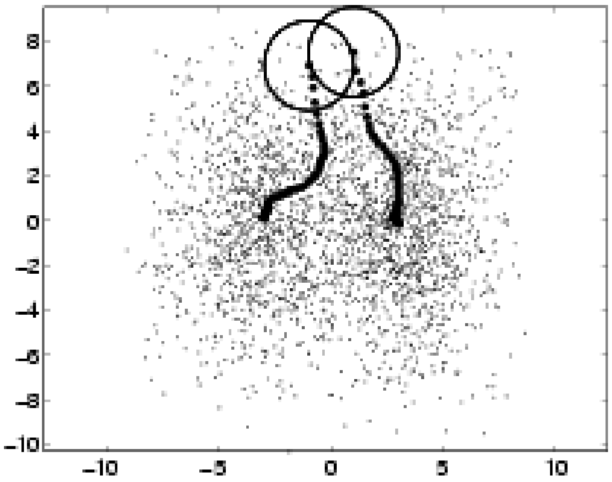
Optimization

- The blue points were traversed by the windows – same cluster



Window trajectories

- The windows trajectories follow the steepest ascent directions



Mean shift: pros and cons

- Pros
  - Does not assume clusters to be spherical
  - One single parameter (window size r)
    - It has a physical meaning
  - Finds a variable number of modes
  - Robust to outliers
- Cons
  - Depends on window size
    - Not always easy to find the best value
    - Inappropriate choice can cause modes to be merged
  - Computationally expensive
  - Does not scale well with dimension of feature space

Università degli Studi di Padova

Mean shift

Stefano Ghidoni