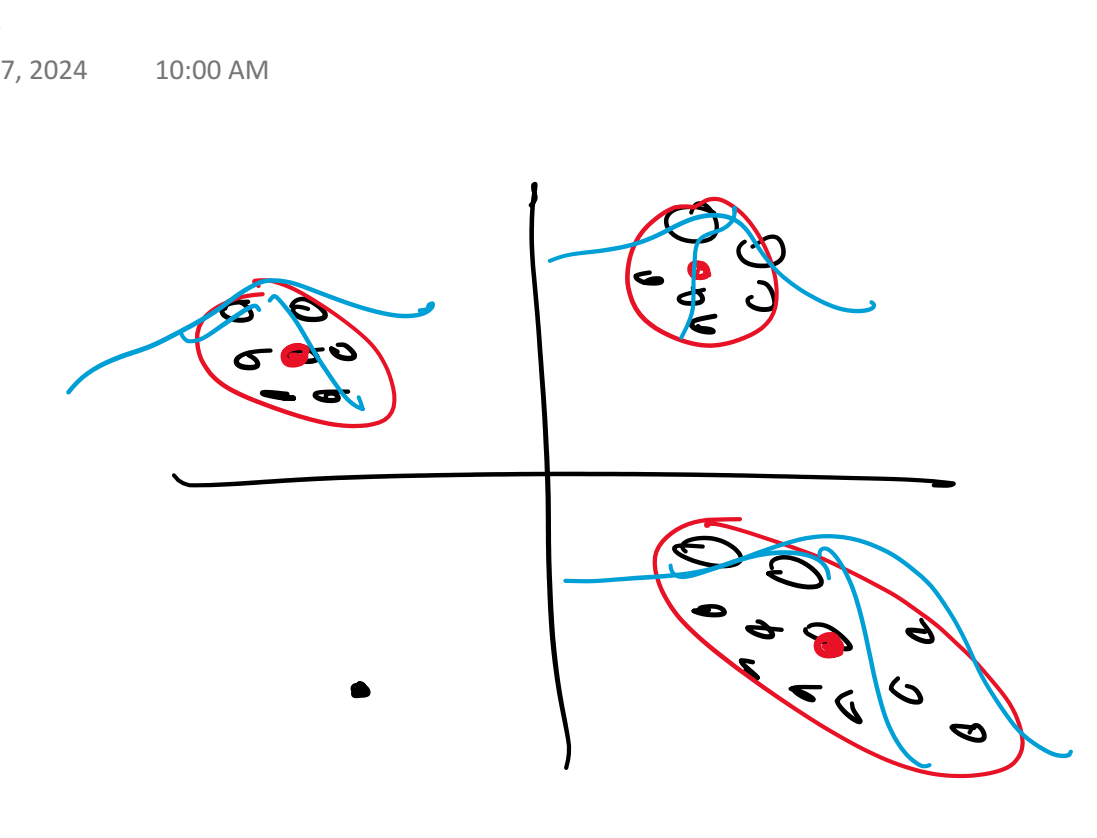
Lecture 11

Monday, October 7, 2024 10:00 AM



Parametric : k #clusters

k Normal distribution $(\vec{\mu}_i, \Sigma_i)$, $\mathbb{R}^d$, $\mathbb{R}^{d\times d}$

$P(C_i)$: prior

$f(\vec{x}) \equiv$ global model

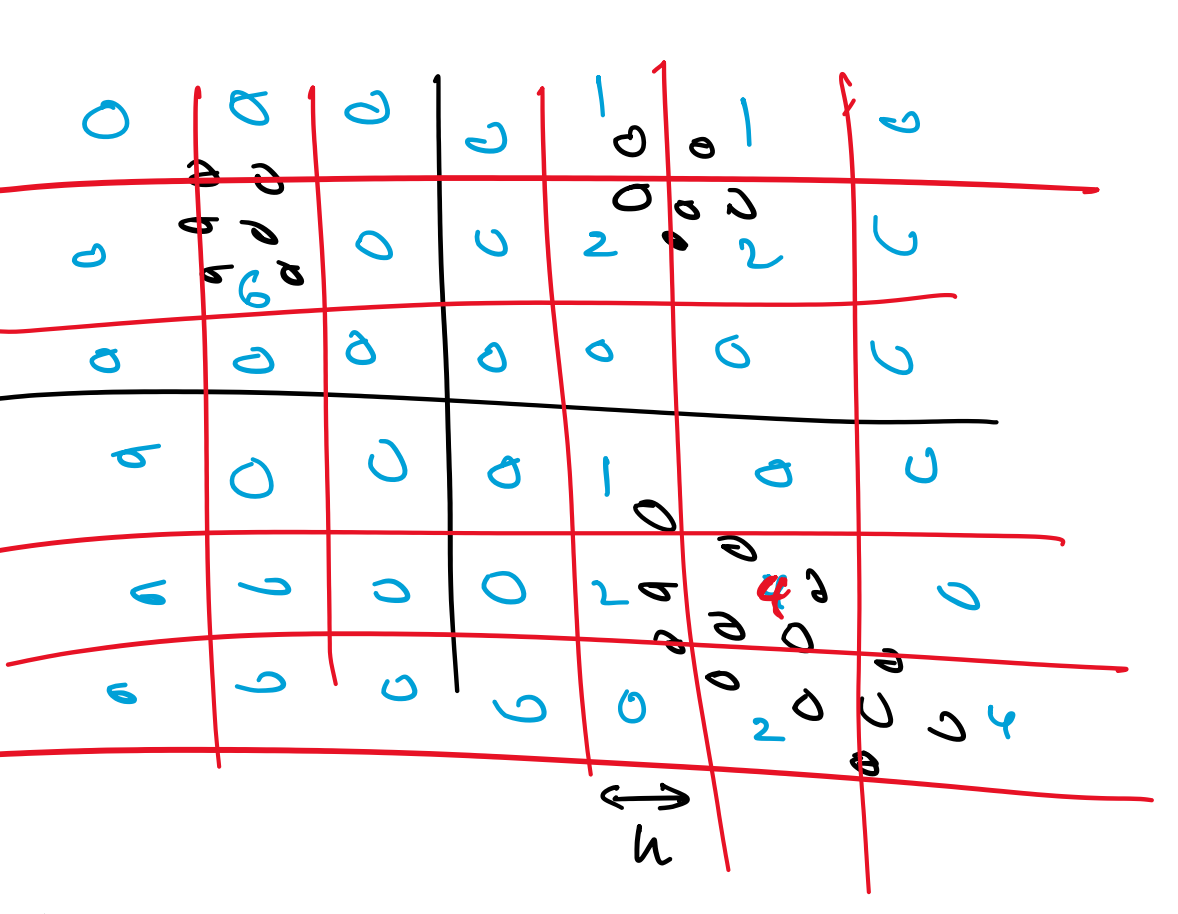
$$f(\vec{x}) = \sum_{i=1}^{k} \underbrace{f(\vec{x} \mid \vec{\mu}_i, \Sigma_i)}_{\text{density}} \cdot \underbrace{P(C_i)}_{\text{weight}}$$

$f(\vec{x})$ — without parameter

density estimation

kernel density estimation

2D histogram



b bins per dim

2D: b^2 cells

d-dim : b^d cells

$n \ll b^d$

$b=2$, $d=100$ $\quad 2^{100}$

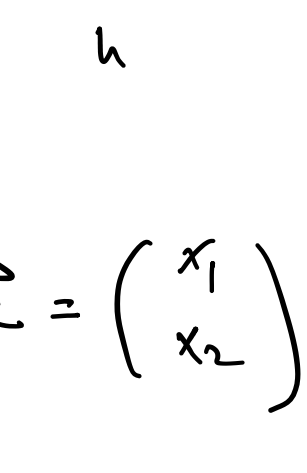
$b_{ij} \equiv \text{bin}(i,j)$

$f_{ij} \equiv$ frequency $= 4$

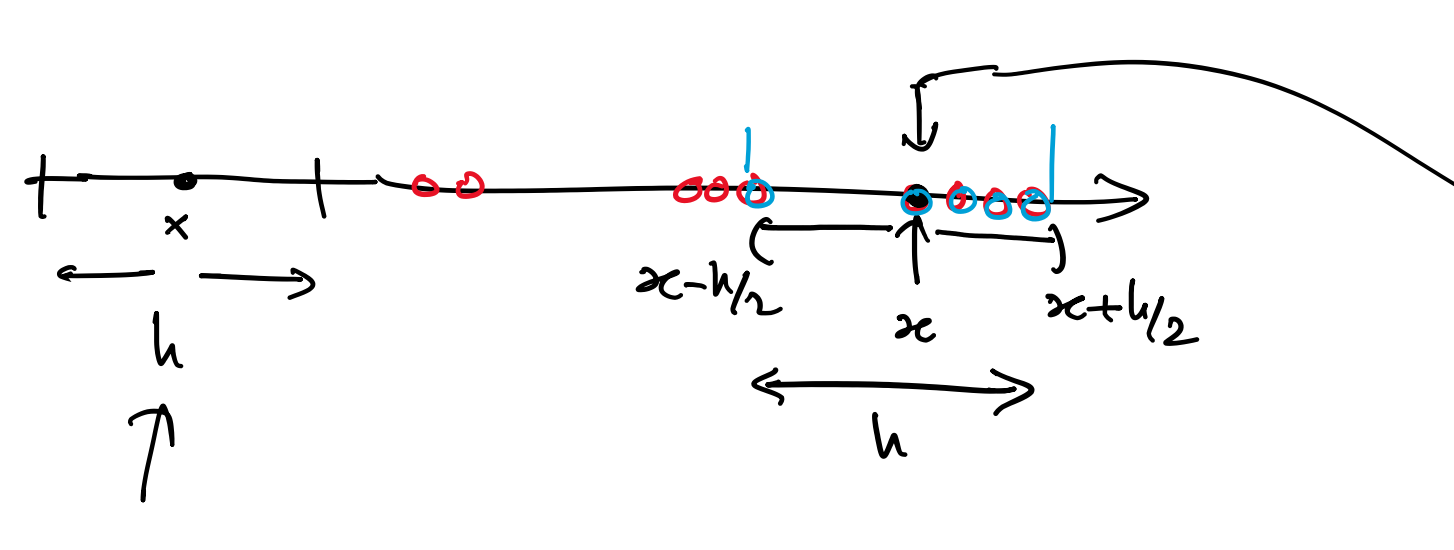
Probability $P_{ij} = \frac{f_{ij}}{n}$ ← # points $= 4/27$

density $\delta_{ij} \equiv \frac{P_{ij}}{\text{vol}} = \frac{P_{ij}}{h^2} = \frac{4/27}{h^2}$

$$\text{density} = \frac{\text{freq}/n}{h} = \frac{\text{freq}}{n \cdot h}$$



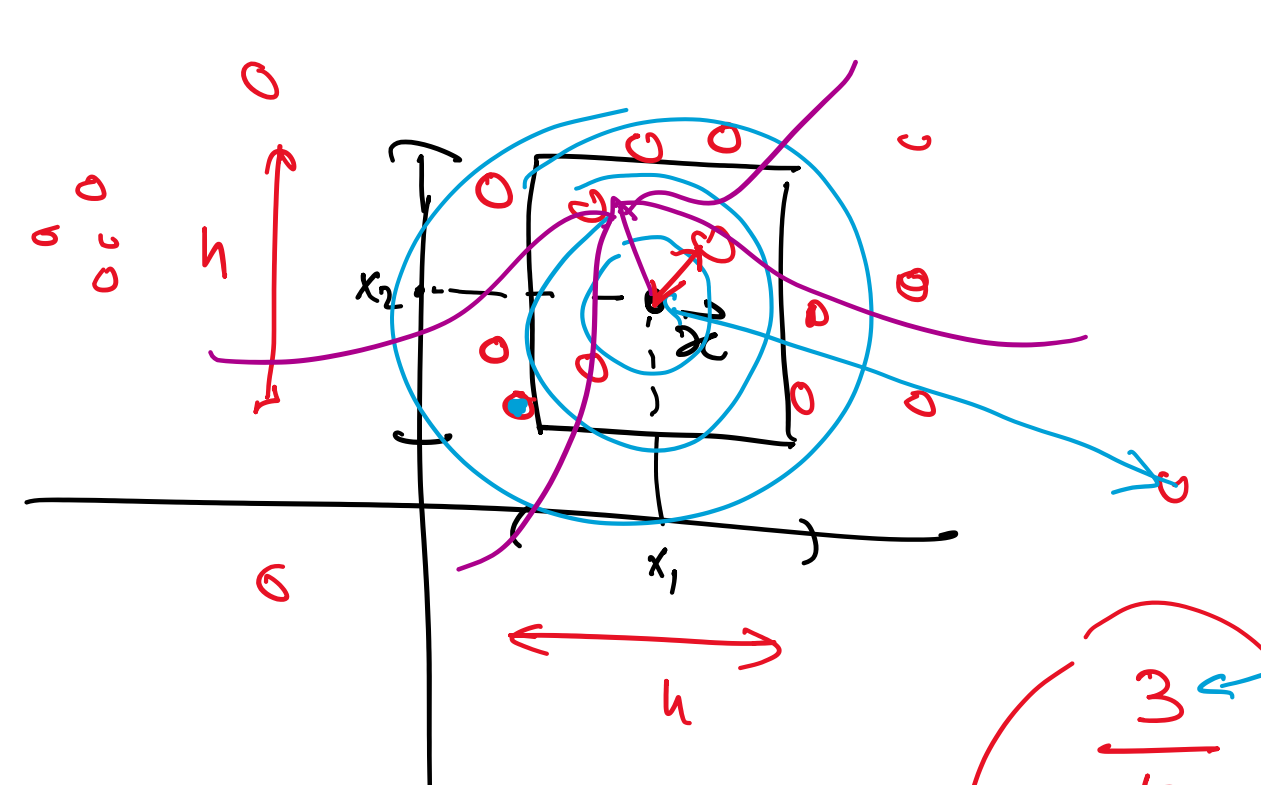
1D:



$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

5 = freq

density $= \frac{5/9}{h}$



$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$\frac{3/12}{h^2}$ ← hard count

discrete kernel

gaussian kernel ↳ 'smooth' counting

$f(\vec{x})$: to estimate the density ↳ frequency or counting

n: sample size

vol: h^d

discrete: each point $\vec{x}_i \in D$ will contribute 1 or 0 to the freq

gaussian: each point $\vec{x}_i \in D$ contribute $\propto e^{-\frac{\|x - x_i\|^2}{2h^2}}$

$x_i = x$ → $e^{-0} = 1$

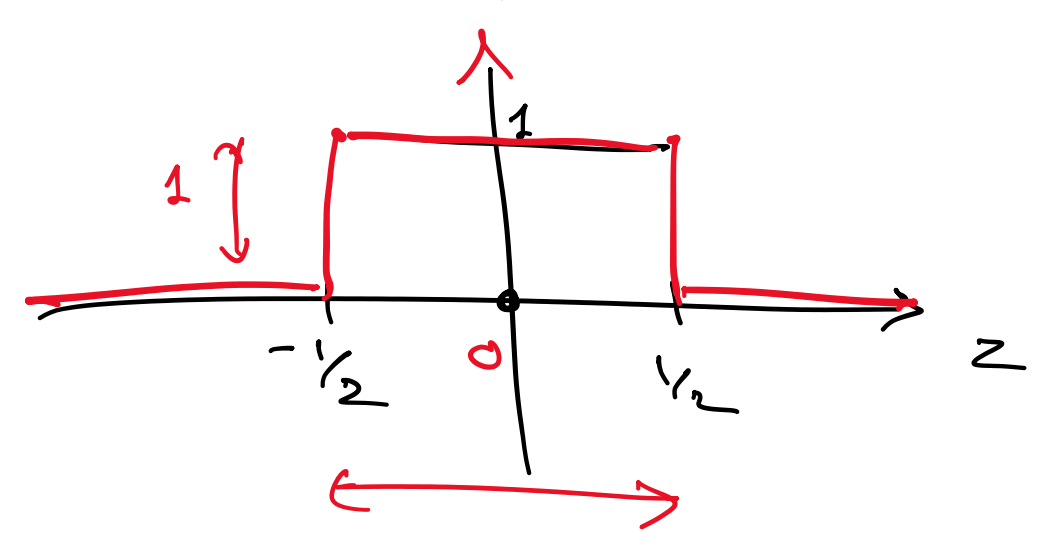
Kernel : K function

1) non-negative $K(\vec{z}) \geq 0$
2) symmetric $K(\vec{z}) = K(-\vec{z})$
3) $\int K(\vec{z})\, dz = 1$ proper density/probability distribution

Discrete kernel

$$\boxed{1d: \quad K(z) = \begin{cases} 1 & \text{if } |z| \leq 1/2 \\ 0 & \text{otherwise} \end{cases}}$$

$$\boxed{\text{d-dim:} \quad K(\vec{z}) = \begin{cases} 1 & \text{if } |z_i| \leq 1/2 \;\; \forall\, i = 1 \dots d \\ 0 & \text{otherwise} \end{cases}}$$



$\int K(z)\,dz = 1$

1×1

$$\boxed{1d \quad f(x) \equiv \frac{\text{freq}}{n \cdot h} = \frac{1}{n \cdot h} \underbrace{\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)}_{\text{freq / count (discrete)}}}$$

true for all $K(z)$!

$K(z) = \begin{cases} 1 & \text{if } |z| \leq 1/2 \\ 0 & \text{otherwise} \end{cases}$ ← $z = \frac{x - x_i}{h}$

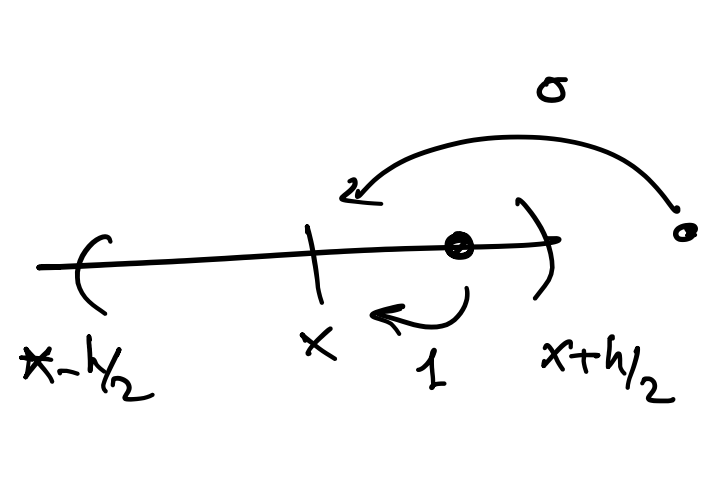
$|z| \leq 1/2$

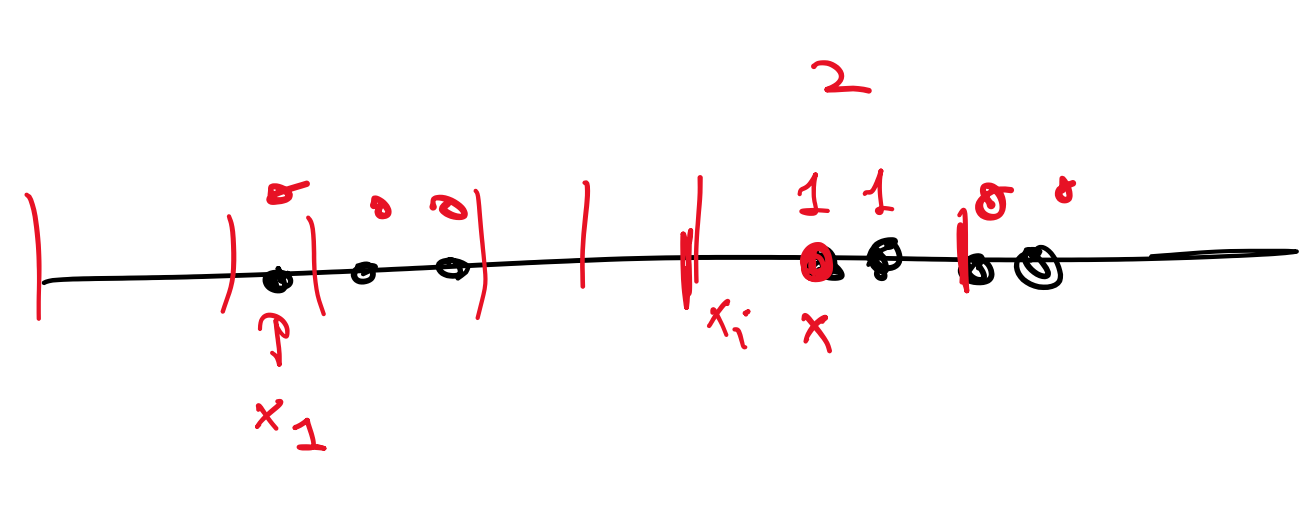
$\Rightarrow -1/2 \leq z \leq 1/2$

$\Rightarrow -1/2 \leq \frac{x - x_i}{h} \leq 1/2$

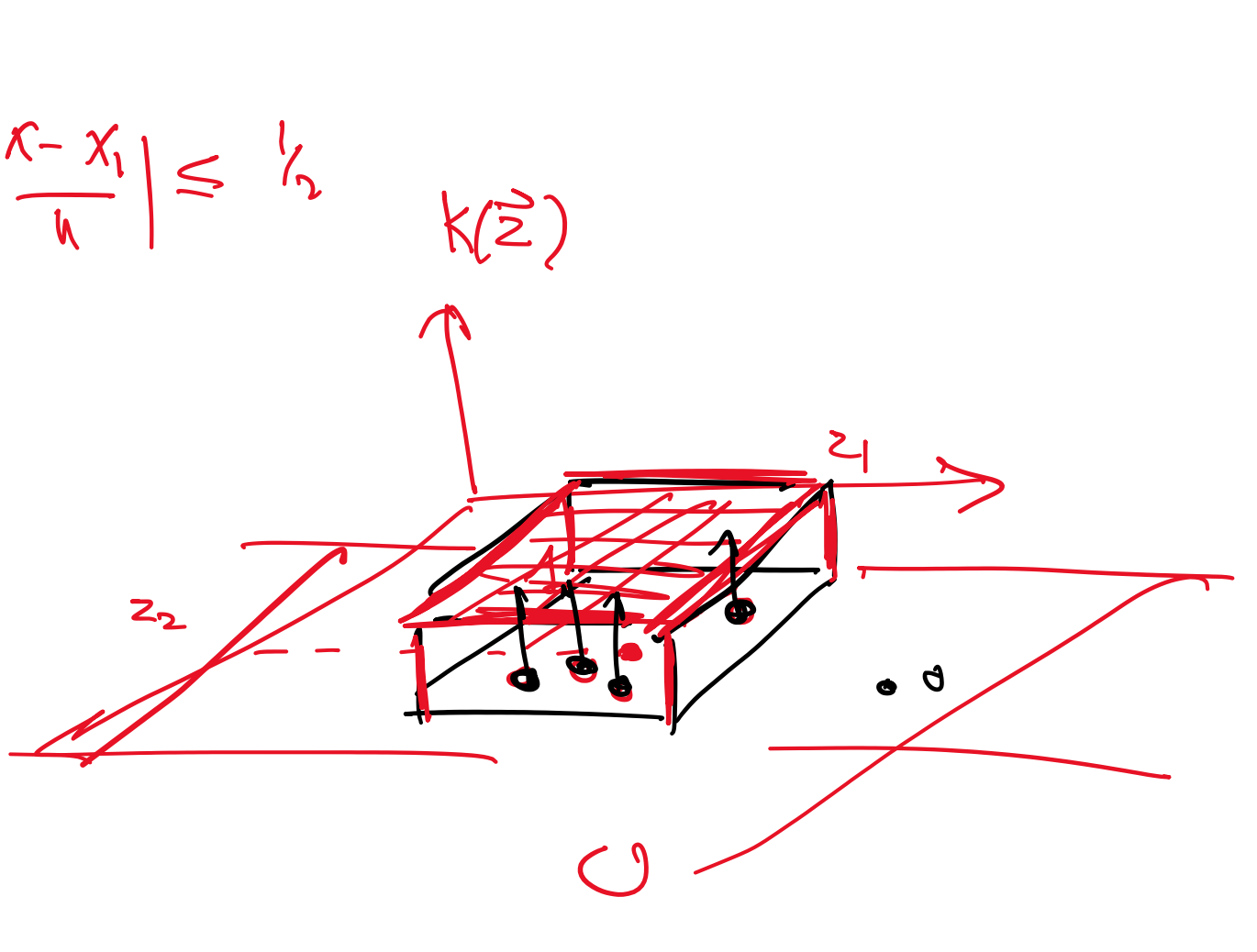
$\Rightarrow -\frac{h}{2} \leq x - x_i \leq h/2$

$\Rightarrow \left(x - h/2 \leq x_i \leq x + h/2\right)$





$\left|\frac{x - x_1}{h}\right| \leq 1/2$



d-dim:

$$\boxed{\vec{x} \in \mathbb{R}^d \quad f(\vec{x}) = \frac{1}{\underset{\text{only change}}{n \cdot h^d}} \sum_{i=1}^{n} K\left(\frac{\vec{x} - \vec{x}_i}{h}\right)}$$

---

gaussian kernel

1d : standard normal centered at 0

$$K(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$f(x) = \frac{1}{n \cdot h} \underbrace{\left(\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)\right)}_{\text{freq}}$$

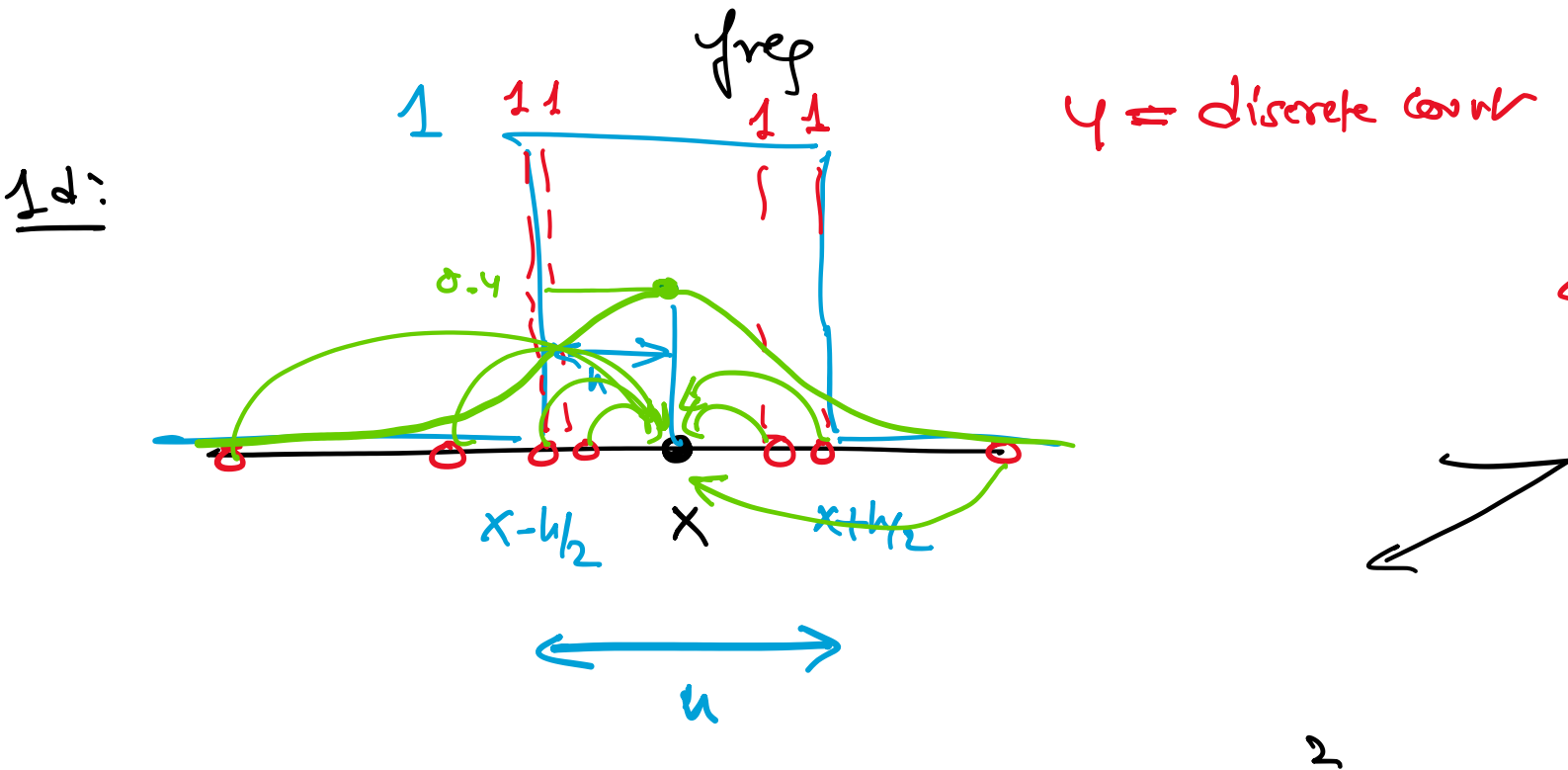
d-dim

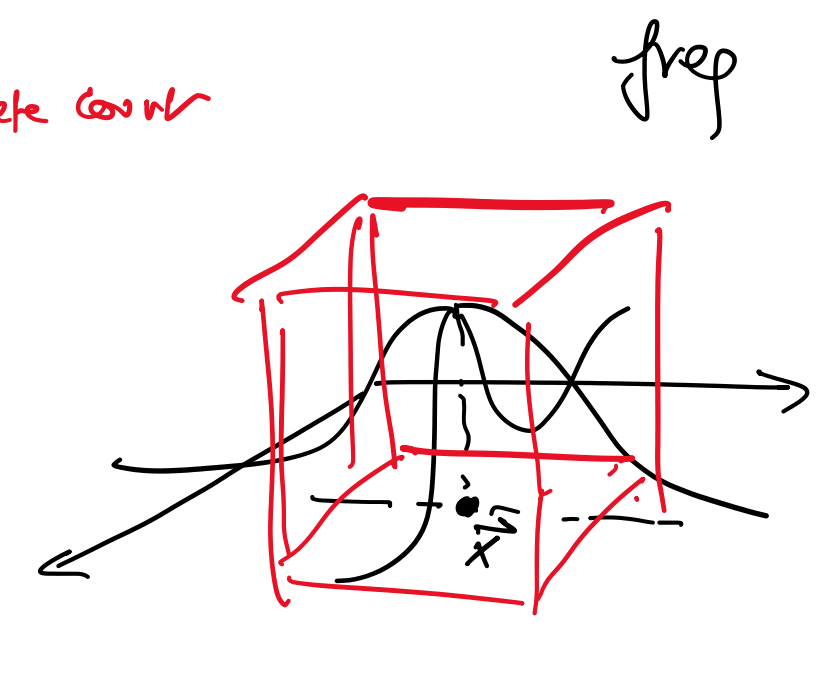
Standard multivariate normal centered at $\vec{0}$

$$K(\vec{z}) = \frac{1}{(2\pi)^{d/2}} e^{-\|\vec{z}\|^2/2}$$

$$f(\vec{x}) = \frac{1}{n h^d} \underbrace{\sum_{i=1}^{n} K\left(\frac{\vec{x} - \vec{x}_i}{h}\right)}_{\text{freq}}$$

1d:





$$K(z) = K\left(\frac{x - x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x - x_i}{h}\right)^2 / 2}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x - x_i)^2}{2h^2}}$$

variance $\equiv h^2$

Std $\equiv h$

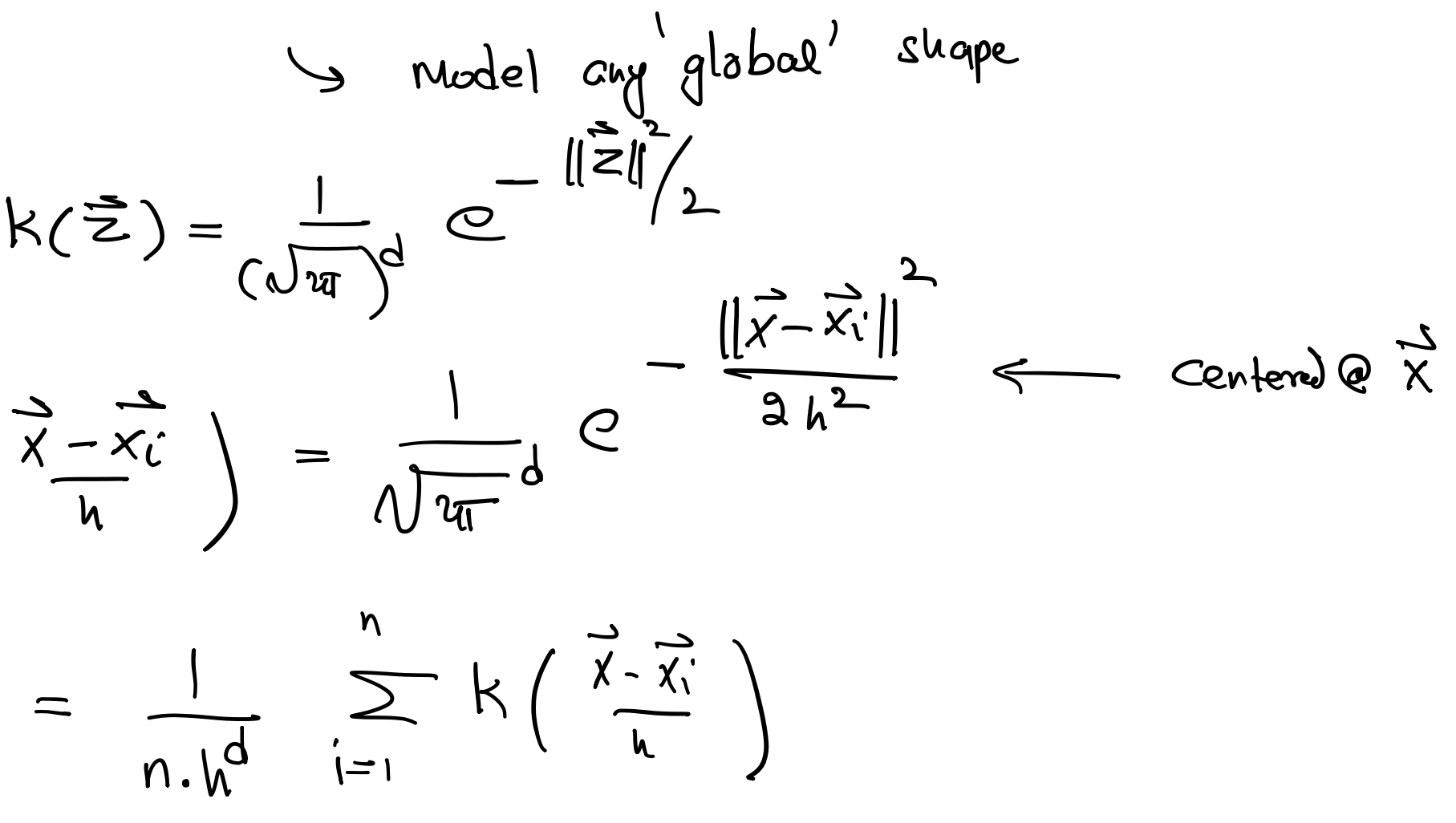
---

let's assume Gaussian kernel ↳ model any 'global' shape

$$K(\vec{z}) = \frac{1}{(\sqrt{2\pi})^d} e^{-\|\vec{z}\|^2/2}$$

$$K\left(\frac{\vec{x} - \vec{x}_i}{h}\right) = \frac{1}{\sqrt{2\pi}^d} e^{-\frac{\|\vec{x} - \vec{x}_i\|^2}{2h^2}} \longleftarrow \text{centered @ } \vec{x}$$

$$f(\vec{x}) = \frac{1}{n \cdot h^d} \sum_{i=1}^{n} K\left(\frac{\vec{x} - \vec{x}_i}{h}\right)$$



clusters:

a set of points that have the same density attractor

$f(x) \longrightarrow \nabla_x = \frac{\partial f(x)}{\partial x}$

$x^0 = x$

gradient ascent $\quad x^t = x^{t-1} + \eta \cdot \nabla_x$ (a small step in the direction of the gradient)

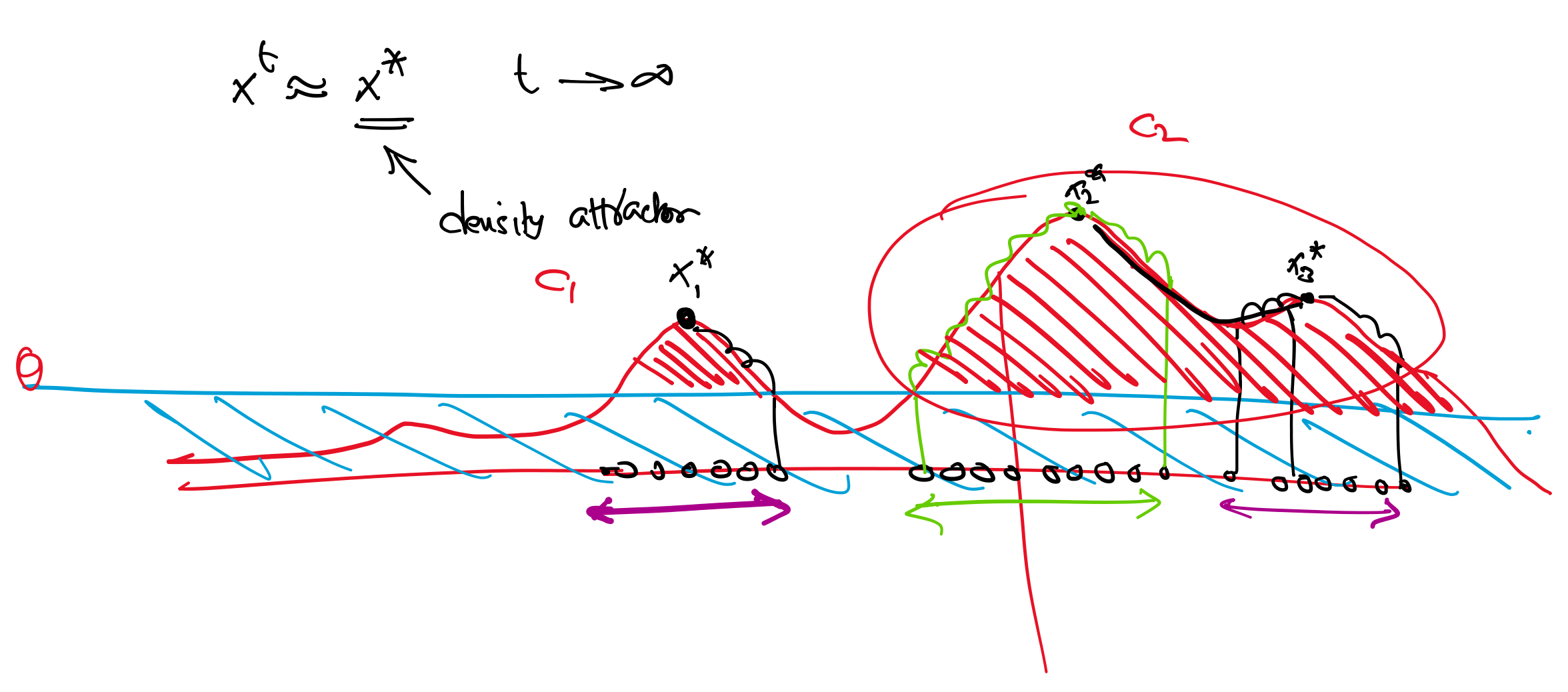
↓

$x^t \simeq \underline{x^*}$ $\quad t \to \infty$

↖ density attractor

1) $\forall\, x_i \in D$ find the density attractor
2) discard x^* if $f(x^*) \leq \theta$
3) merge "connected" attractors

Denclue algorithm



$$f(\vec{x}) = \frac{1}{n \cdot h^d} \sum_{i=1}^{n} K\left(\frac{\vec{x} - \vec{x}_i}{h}\right)$$

$$\boxed{\text{Gaussian kernel} \quad K(\vec{z}) = \frac{1}{\sqrt{2\pi}^d} e^{-\frac{\|\vec{z}\|^2}{2}}}$$

$$\nabla_{\vec{x}} = \frac{\partial f(\vec{x})}{\partial \vec{x}} = \frac{1}{n \cdot h^{d+2}} \underbrace{\sum_{i=1}^{n} \underbrace{K\left(\frac{\vec{x} - \vec{x}_i}{h}\right)}_{\text{scalar}} \underbrace{(\vec{x}_i - \vec{x})}_{\text{vector difference}}}_{\text{sum of weighted vector differences}}$$

1) gradient ascent

Start at a given $\vec{x}$ & update

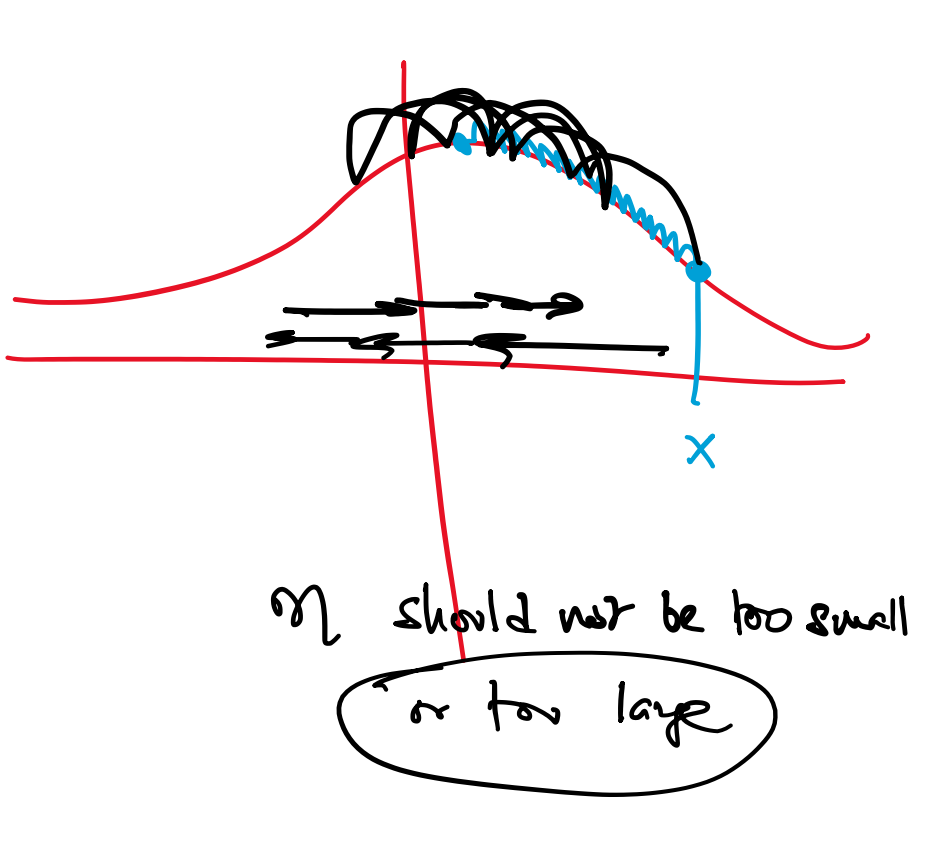
$\vec{x}^{(t)} = \vec{x}^{(t-1)} + \eta\, \nabla_{\vec{x}^{(t-1)}}$

repeat until convergence (t times)

$\|\vec{x}^{(t)} - \vec{x}^{(t-1)}\| \leq \varepsilon$

$\vec{x}^{(t)} = x^*$ ↖ attractor

$\vec{x}^0 = \vec{x}$ $\quad t = 0$



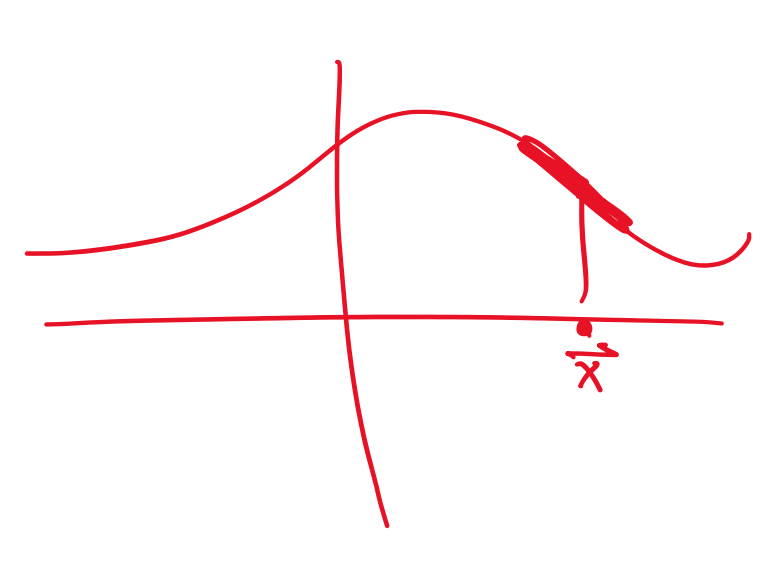
η should not be too small or too large

2) Solve for $\nabla_x = 0$

$$\nabla_x = \cancel{\frac{1}{n h^{d+2}}} \sum_{i=1}^{n} K\left(\frac{\vec{x} - \vec{x}_i}{h}\right)(\vec{x}_i - \vec{x}) = 0$$

$$\Rightarrow \underbrace{\sum_{i=1}^{n} K\left(\frac{\vec{x} - \vec{x}_i}{h}\right)}_{\text{scalar}} \vec{x} = \sum_{i=1}^{n} K\left(\frac{\vec{x} - \vec{x}_i}{h}\right) \cdot \vec{x}_i$$

$$\Rightarrow \vec{x} = \frac{\sum K\left(\frac{\vec{x} - \vec{x}_i}{h}\right) \cdot \vec{x}_i}{\sum K\left(\frac{\vec{x} - \vec{x}_i}{h}\right)}$$



$$\vec{x} = \frac{\sum w_i\, \vec{x}_i}{\sum w_i}$$

$w_i = K\left(\frac{\vec{x} - \vec{x}_i}{h}\right)$

$w_i^{(t-1)} = K\left(\frac{x^{(t-1)} - x_i}{h}\right)$

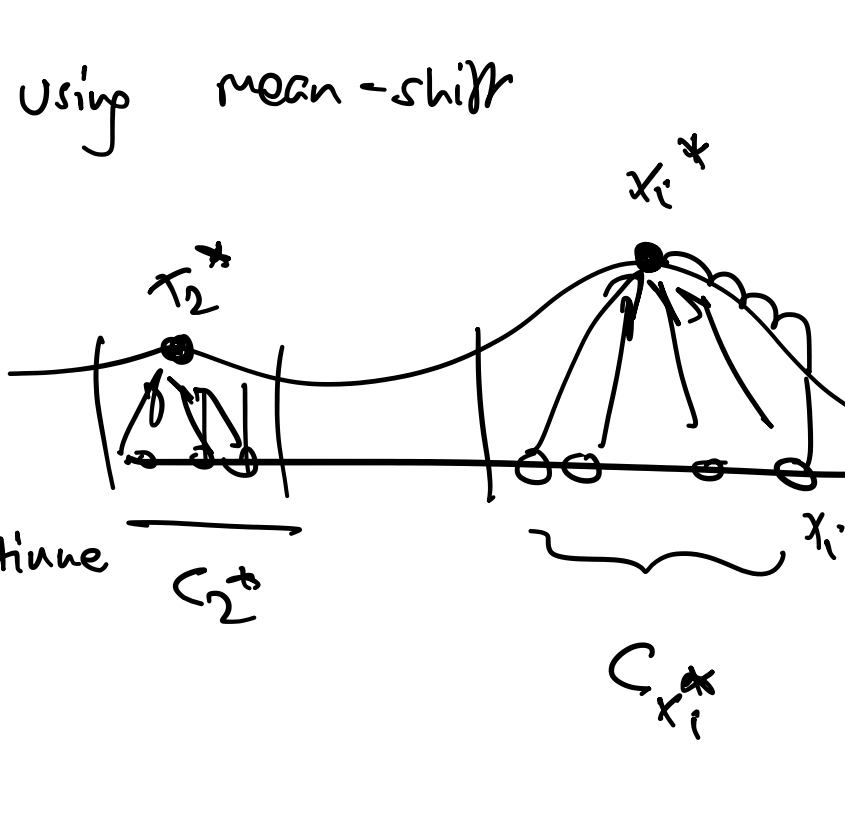
new update for $\vec{x}$ is a weighted mean of all $\vec{x}_i \in D$

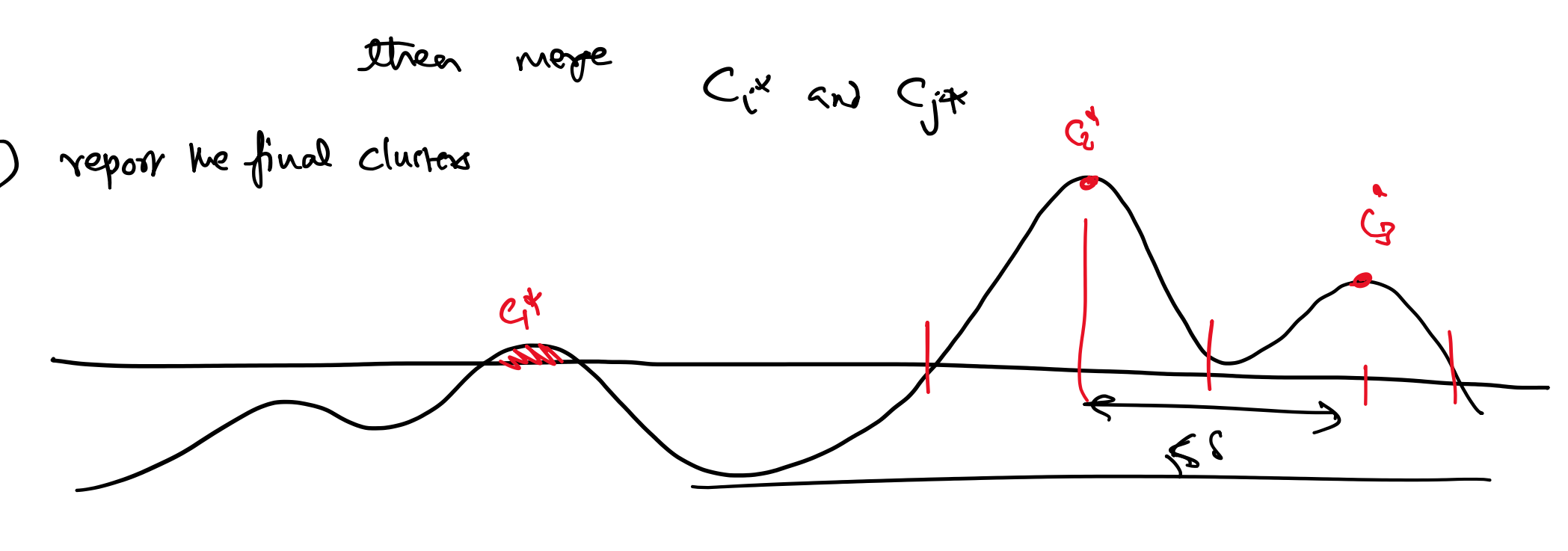
$$\boxed{\vec{x}^{(t)} = \frac{\sum w_i^{t-1} \cdot \vec{x}_i}{\sum w_i^{t-1}}}$$

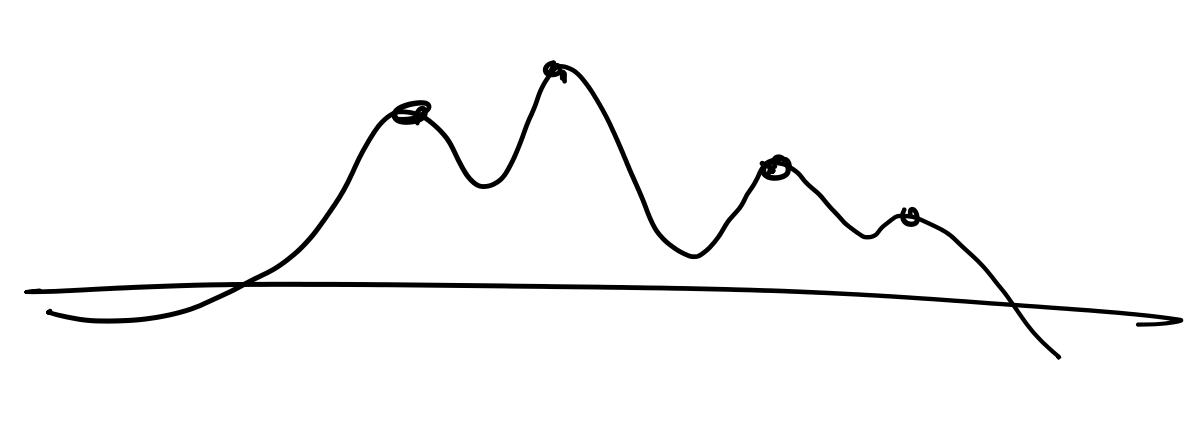
mean-shift equation

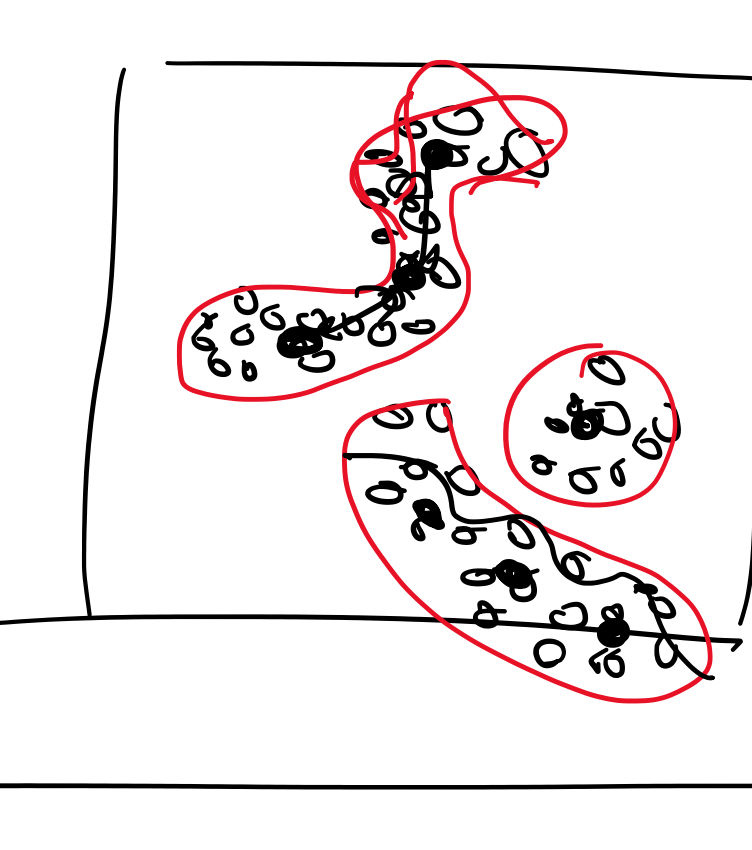
Denclue (D, θ, h)

1) $\forall\, \vec{x}_i \in D$
   - a) find the density attractor x_i^* using mean-shift
   - c) assign $\vec{x}_i$ to $C_{x_i^*}$ cluster
   - b) if $f(x_i^*) \leq \theta$, discard/continue
2) several local maximas, attractors $C_1^*, C_2^*, \dots, C_m^*$
3) merge clusters ?
   → if $\|x_i^* - x_j^*\|^2 \leq \delta$ (they are close) then merge C_i^* and C_j^*
4) report the final clusters









---

h ← variance for gaussian kernel

$$K\left(\frac{x - x_i}{h}\right) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{\|x - x_i\|^2}{2 h^2}}$$

variance along each axis!

Σ^{-1} cov inverse

0) fixed h^2 diagonal cov

$$\Sigma = \begin{pmatrix} h^2 & & 0 \\ & h^2 & \\ 0 & & h^2 \;\ddots \end{pmatrix}$$

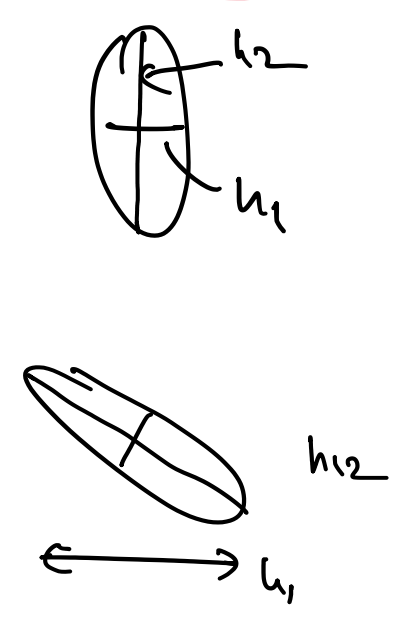
$(x - x_i)^T \Sigma^{-1} (x - x_i)$

$\frac{1}{h}$ — h^2

1) one h_i^2 per dim

$$\Sigma = \begin{pmatrix} h_1^2 & & 0 \\ & h_2^2 \;\ddots & \\ 0 & & h_d^2 \end{pmatrix}$$

2) full cov

$$\Sigma = \begin{pmatrix} h_1^2 & h_{12} & \cdots & h_{1d} \\ h_{12} & h_2^2 & & \vdots \\ \cdots & & \cdots & h_d^2 \end{pmatrix}$$