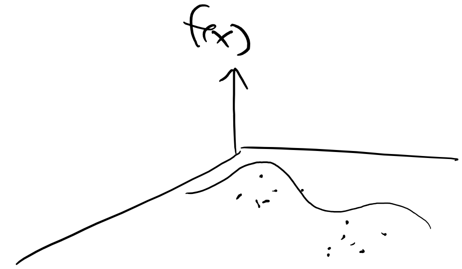
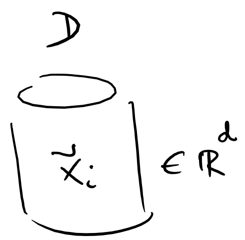
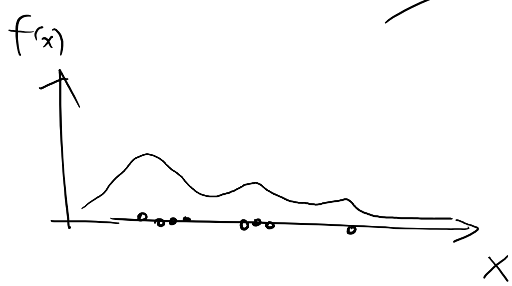
# Kernel Density Estimation

→ non-parametric approach




CDF: Cumulative Distribution Function

$$F(\underline{x}) = P(X \le \underline{x})$$

empirical CDF →

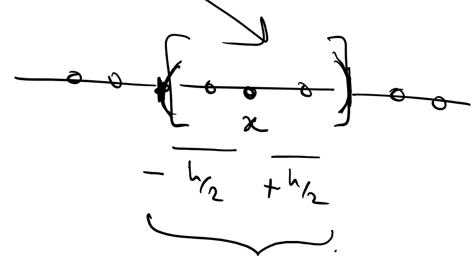
$$\hat{F}(x) = \underline{\underline{\frac{1}{n}}} \sum_{i=1}^{n} \underbrace{I(x_i \le x)}_{\text{Indicator function}}$$

← fraction of points $\le x$.

density function $\hat{f}(x)$ is the derivative of $\hat{F}(x)$

$$\boxed{\hat{f}(x) = \frac{\hat{F}(x+h/2) - \hat{F}(x-h/2)}{h}}$$

← width




$h \to 0$

$$\underbrace{\hat{f}(x) = \frac{k/n}{h}}_{\text{density of the points}}$$

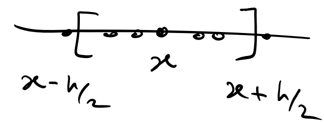
$k/n$ ← fraction of points in h-width interval around x

$h$ ← width

equivalent

equivalent $\downarrow$ $\underbrace{\qquad\qquad\qquad}_{\text{density of the points}}$

① Discrete kernel

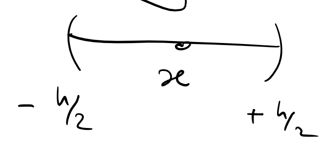



different from the pair-wise kernel

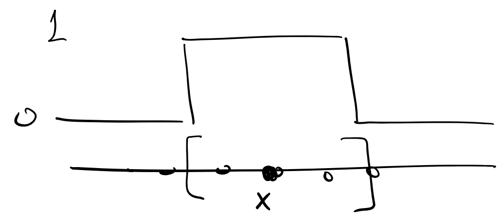
Kernel is a function
1. non-negative $\quad k(x) \geq 0$
2. symmetric $\quad k(x) = k(-x)$
3. Integrates to 1 $\quad \int k(x)\,dx = 1 \quad \Big| \quad \sum_{-\infty}^{\infty} k(x) = 1$

$$\hat{f}(x) = \frac{k/n}{h} = \frac{k}{nh}$$

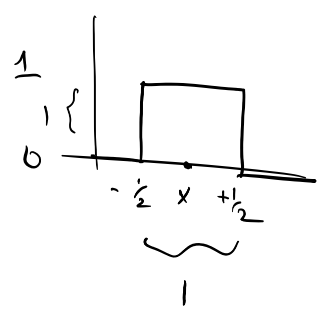
$$\hat{f}(x) = \frac{1}{nh} \underbrace{\sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right)}$$




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




$$k\left(\frac{x - x_i}{h}\right) = 1 \quad \text{if} \quad \left|\frac{x - x_i}{h}\right| \leq 1/2$$



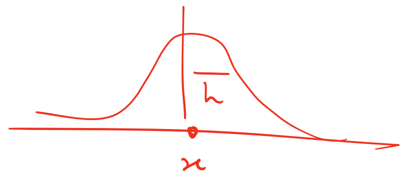

$$-\frac{1}{2} \leq \frac{x - x_i}{h} \leq \frac{1}{2}$$

$$-h/2 \leq x - x_i \leq h/2$$

$$\underbrace{x - h/2 \leq x_i \leq x + h/2}$$

② Gaussian kernel

## (2) gaussian kernel



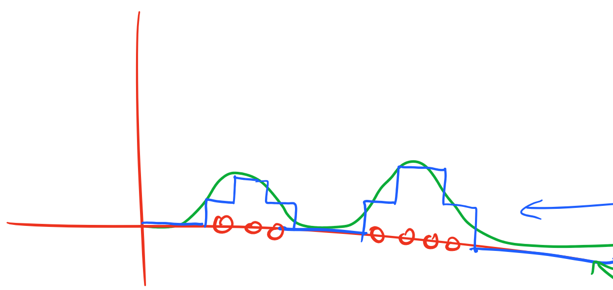

h

user-defined

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

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

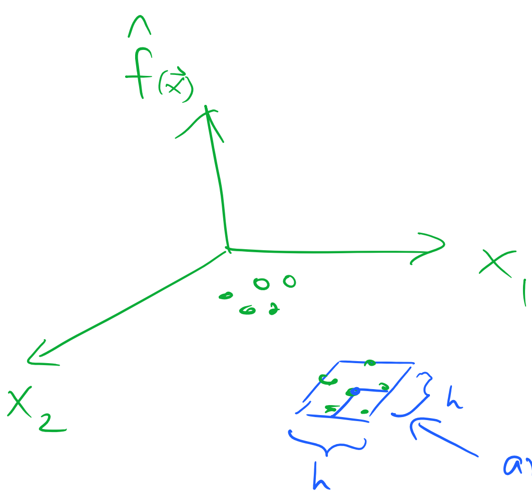
gaussian kernel



discrete kernel

gaussian

## 2D or higher




around x a hypercube with edge length h

$$\hat{f}(x) = \frac{\text{fraction of points}}{\quad}$$

vol of hypercube

$$= \frac{1}{n \cdot h^d} \left( \text{count} \right)$$

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

$\vec{x}_i \in \mathbb{R}^d$

① Discrete kernel

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

$$\vec{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_d \end{pmatrix}$$

② Gaussian kernel

$$K(\vec{z}) = \underbrace{\frac{1}{(2\pi)^{d/2}} e^{-\frac{\vec{z}^T \vec{z}}{2}}}_{\text{Standard multivariate normal}}$$

$\vec{\mu} = 0, \quad \underbrace{\Sigma = I}_{\text{Covariance matrix is identity}}$

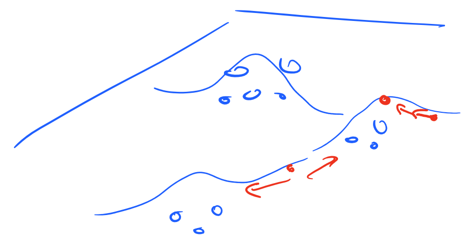
---

Use the density function $\hat{f}(x)$ for clustering

gradient ascent
hill climbing

① extract local maxima
also called attractors

$$\nabla \hat{f}(\vec{x}) = \frac{\partial \hat{f}(\vec{x})}{\partial \vec{x}}$$



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

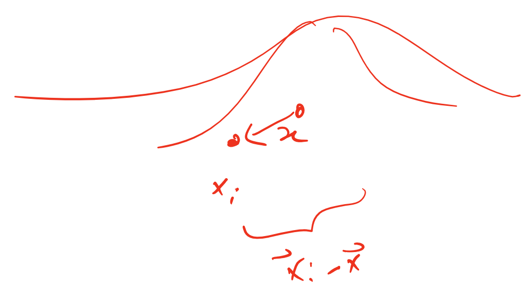
assume $K(\vec{z}) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{\vec{z}^T \vec{z}}{2}}$

$$\frac{\partial K(\vec{z})}{\partial \vec{x}} = \frac{\partial K(\vec{z})}{\partial \vec{z}} \cdot \frac{\partial \vec{z}}{\partial \vec{x}} = \underbrace{\frac{1}{(\sqrt{2\pi})^d} e^{-\frac{z^T z}{2}}} \cdot \underbrace{\frac{\partial \vec{z}}{\partial \vec{x}}}_{\frac{1}{h}} \; - \frac{\cancel{2}\vec{z}}{\cancel{2}}$$

$$\frac{\partial K(\vec{z})}{\partial \vec{x}} = -\frac{1}{h} K(\vec{z})\, \vec{z}$$

$$\nabla \hat{f}(x) = \frac{1}{n \cdot h^{d+2}} \sum_{i=1}^{n} \underbrace{K\left(\frac{x - x_i}{h}\right)} \cdot \underbrace{(\vec{x}_i - \vec{x})}$$

① find attractors




$\forall\ \vec{x}_i \in D$

// gradient ascent

$\vec{v}_0 = \vec{x}_i$

$\vec{V}(t+1) = V(t) + \underbrace{\eta \cdot \nabla \hat{f}(\vec{v})}$

step size

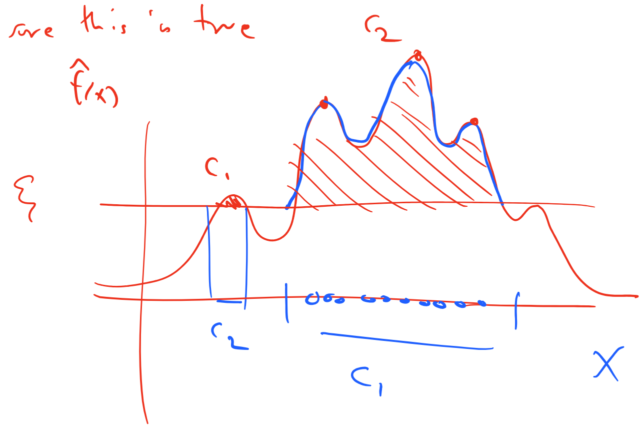
$\vec{x}_i$ find $\underbrace{A(\vec{x}_i)}_{\text{attractor for } \vec{x}_i}$

replace with mean-shift step

attractor for $\vec{x}_i$

(2) discard attractors whose density is below $\xi$

$\boxed{\hat{f}(A(x_i)) \ge \xi}$ make sure this is true




(3) extract the connected components of attractors

we should be able to reach from $A(\vec{x}_i)$ to $A(\vec{x}_j)$ without the density falling below $\xi$

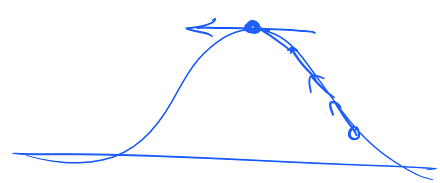
Spatial Index → e.g. core points & neighborhood

---

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

at optimal value

$\nabla \hat{f}(x) = 0$



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

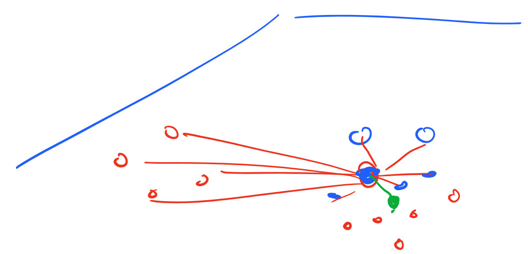
$$\vec{x} = \frac{\sum_{i=1}^{n} K\left(\frac{\vec{x}-x_i}{h}\right)\cdot \vec{x_i}}{\sum_{i=1}^{n} K\left(\frac{\vec{x}-x_i}{h}\right)}$$

weighted mean

$\vec{x}_0 = x$

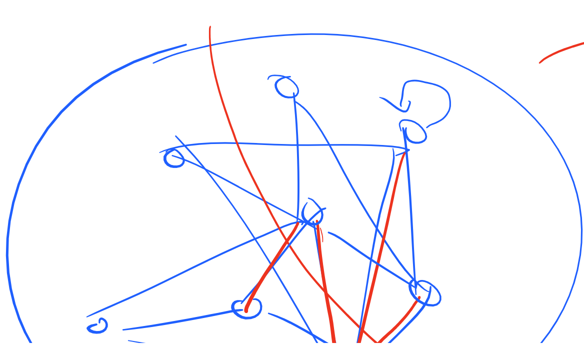
$$\underbrace{\vec{x}_{(t+1)} = \frac{\sum K\left(\frac{x_t - x_i}{h}\right) \vec{x_i}}{\sum K\left(\frac{x_t - x_i}{h}\right)}}_{\text{mean-shift rule}}$$

t is iteration



# Graph & Spectral clustering

D



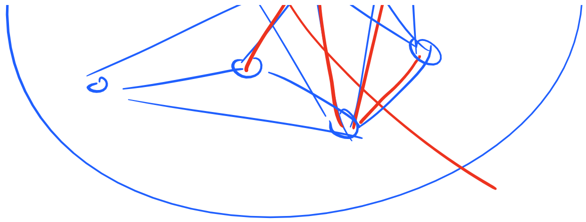
graph / divisive clustering

≡ graph-cuts

≡ graph partitioning

≡ k-way cuts

pair-wise distance matrix



distance matrix

weighted adjacency matrix

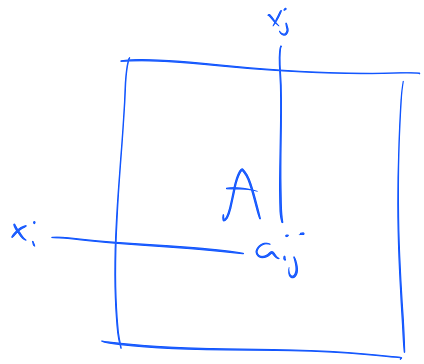
$K \leftarrow$ kernel matrix

Similarities

PSD

$D \to G$

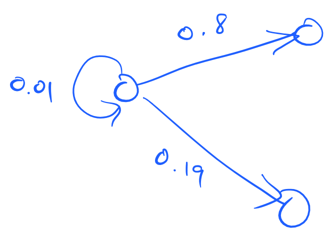
$G$

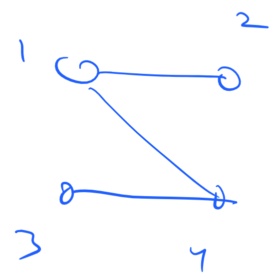



$a_{ij}$ = weight / similarity between nodes $x_i$ & $x_j$

$a_{ij} \geq 0$

symmetric




$$\begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 0 & 1 & 0 & 1 \\ 2 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 1 \\ 4 & 1 & 0 & 1 & 0 \end{array}$$

normal adjacency

graph $G = (V, E)$

$\Downarrow$

matrices that represent aspects of $G$

① Adjacency Matrix $A = \{ a_{ij} \mid (x_i, x_j) \in E \}$

$a_{ij} \geq 0$

$a_{ij} = a_{ji}$

square, symmetric

$n \times n$

Undirected graph

② Degree matrix : $\Delta$

$$\Delta = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{pmatrix}$$

$$d_i = \sum_{j=1}^{n} a_{ij}$$

weighted degree

③ Normalized Adjacency Matrix.

$$M = \Delta^{-1} A = \begin{pmatrix} \frac{a_{11}}{d_1} & \frac{a_{12}}{d_1} & \cdots & \frac{a_{1n}}{d_1} \\ \frac{a_{21}}{d_2} & \frac{a_{22}}{d_2} & \cdots & \frac{a_{2n}}{d_2} \\ \vdots & & & \\ \frac{a_{n1}}{d_n} & \frac{a_{n2}}{d_n} & \cdots & \frac{a_{nn}}{d_n} \end{pmatrix}$$

markov chain

1

Pagerank

$M^T$

$m_{ij}$ = prob of jumping from node $x_i$ to node $x_j$

Sum of row 1 :
$$\sum_{i=1}^{n} \frac{a_{1i}}{d_1} = \frac{d_1}{d_1} = 1$$

M : largest eigenvalue $\lambda_1 = 1$

all other eigenvalues $|\lambda_i| \leq 1$

$$M \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$\lambda_1 = 1$

$u_1 = \left.\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\right\} n$

$$\boxed{M\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} = 1\cdot\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}}$$

$\lambda_1 = 1$

$u_1 = \left.\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\right\} n$

Laplacian Matrix

$$L = \Delta - A$$

$$= \begin{pmatrix} d_1 - a_{11} & -a_{12} & -a_{13} & \cdots & -a_{1n} \\ -a_{21} & d_2 - a_{22} & -a_{23} & \cdots & -a_{2n} \\ & \vdots & & & \end{pmatrix}$$

$M = \Delta^{-1} A$

PSD matrix

Positive semidefinite