Mean-shift

1 Kernel density estimators:

Let $P = \{p_1, \dots, p_n\} \subset \mathbb{R}^d$.

* General form: $\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_H(x - p_i)$

where: $K_H = (\det H)^{-1/2} \, K(H^{-1/2} x)$

H: $d \times d$ covariance matrix (puts anisotropy + scaling)

d-variate kernel $K : \mathbb{R}^d \to \mathbb{R}$ s.t.

- $\int_{\mathbb{R}^d} K(x)\, dx = 1$ (normalized)
- $\lim_{\|x\| \to +\infty} K(x) = 0$ (vanishes at infinity)
- $\int_{\mathbb{R}^d} x \, K(x) = 0$ (centered at the origin)
- $\int_{\mathbb{R}^d} x x^T K(x)\, dx = c_K \cdot I_d$ for some cst c_K. (isotropic)

* Specialization 1: take $H = \sigma^2 \cdot I_d$

↳ $\hat{f}(x) = \frac{1}{n\sigma^d} \cdot \sum_{i=1}^{n} K\left(\frac{x - p_i}{\sigma}\right)$

σ: bandwidth / scaling

* Specialization 2: take $K(x) = c_{k,d} \cdot k(\|x\|^2)$

k: symmetric function $\mathbb{R} \to \mathbb{R}$ (kernel profile)

radially symmetric kernel

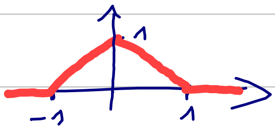
↳ $$\boxed{\hat{f}_{\sigma,k}(x) = \frac{c_{k,d}}{n\sigma^d} \cdot \sum_{i=1}^{n} k\left(\frac{\|x - p_i\|^2}{\sigma^2}\right)}$$

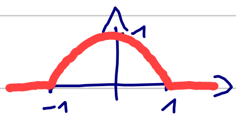
Examples:

- Epanechnikov: $k_E(t) = \begin{cases} 1 - |t| & \text{if } |t| \leq 1 \\ 0 & \text{if } |t| > 1 \end{cases}$

$$\hookrightarrow c_{k,d} = \frac{d+2}{2 \operatorname{Vol} B_d(0,1)}.$$

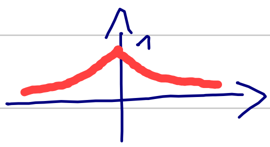
$$\left(\operatorname{Vol} B_d(0,1) = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)} \right)$$

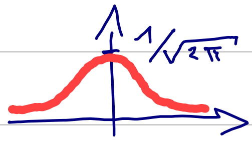
profile: 

kernel (1-d): 

- Gaussian (normal): $k_N(t) = \exp(-|t|/2)$.

$$\hookrightarrow c_{k,d} = \frac{1}{\left(\sqrt{2\pi}\right)^d}$$

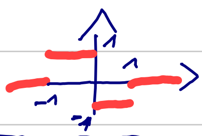
profile: 

kernel (1-d): 

② Differentiation (gradient):

Kernel profile k $\rightsquigarrow$ derivative: $k' := \dfrac{\partial k}{\partial t}$.

Examples:
- $k'_E = \begin{cases} \pm 1 & \text{if } |t| \leq 1 \\ 0 & \text{if } |t| > 1 \end{cases}$ flat kernel
- $k'_N \propto k_N$.



Then:
$$\boxed{\begin{aligned} \hat{\nabla} f_{\sigma,k}(x) &:= \nabla \hat{f}_{\sigma,k} \\ &= \frac{2 c_{k,d}}{n \sigma^{d+2}} \cdot \sum_{i=1}^{n} (x - p_i)\, k'\left(\frac{\|x-p_i\|^2}{\sigma^2}\right). \end{aligned}}$$

Let $g = -k'$ (profile) and $G(x) = c_{g,d}\ g(\|x\|^2)$ (kernel)

normalizing constant

Then: $\nabla \hat{f}_{\sigma,k}(x) = \frac{2 c_{k,d}}{n \sigma^{d+2}} \sum_{i=1}^{n} (p_i - x) \cdot g\left(\frac{\|x - p_i\|^2}{\sigma^2}\right)$

$$= \frac{2 c_{k,d}}{n \sigma^{d+2}} \cdot \left[\sum_{i=1}^{n} g\left(\frac{\|x-p_i\|^2}{\sigma^2}\right)\right] \cdot \left[\frac{\sum_{i=1}^{n} p_i \cdot g\left(\frac{\|x-p_i\|^2}{\sigma^2}\right)}{\sum_{i=1}^{n} g\left(\frac{\|x-p_i\|^2}{\sigma^2}\right)} - x\right]$$

$(**)$

First factor $\propto$ density estimate using G:

$$\hat{f}_{\sigma,g}(x) = \frac{c_{g,d}}{n \sigma^d} \sum_{i=1}^{n} g\left(\frac{\|x - p_i\|^2}{\sigma^2}\right).$$

Second factor = $\boxed{\text{mean shift}}$:

shift from x

$$m_{\sigma,g}(x) = \frac{\sum_{i=1}^{n} p_i \cdot g\left(\frac{\|x-p_i\|^2}{\sigma^2}\right)}{\sum_{i=1}^{n} g\left(\frac{\|x-p_i\|^2}{\sigma^2}\right)} - x.$$

barycenter of P w.r.t. k', σ, x

↳ passing 1^{st} factor and c^t to left-hand side in $(**)$ yields:

$$\boxed{m_{\sigma,g}(x)} = \frac{n\sigma^{d+2}}{2 c_{k,d}} \cdot \frac{\nabla \hat{f}_{\sigma,k}(x)}{\sum_{i=1}^{n} g\left(\frac{\|x-p_i\|^2}{\sigma^2}\right)}$$

$$\boxed{= \frac{\sigma^2}{2} \frac{c_{g,d}}{c_{k,d}} \cdot \frac{\nabla \hat{f}_{\sigma,k}(x)}{\hat{f}_{\sigma,g}(x)}}$$

gradient estimate normalized by k'

(hence $m_{\sigma,g} \propto \nabla \hat{f}_{\sigma,k}$)

③ Mean-shift procedure:

Input: $\underline{P} = \{p_1, \dots, p_n\} \subset \mathbb{R}^d$, $x \in \mathbb{R}$ (query)

Parameters: profile $k : \mathbb{R} \to \mathbb{R}$, bandwidth $\sigma > 0$.

Code (gradient ascent):

$x_0 := x$

Repeat:

$$x_{j+1} := x_j + m_{\sigma,g}(x_j) = \frac{\sum_{i=1}^{n} p_i \cdot g\left(\frac{\|x_j - p_i\|^2}{\sigma^2}\right)}{\sum_{i=1}^{n} g\left(\frac{\|x_j - p_i\|^2}{\sigma^2}\right)}.$$

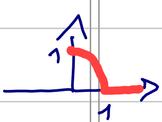
Until convergence.

④ Guarantees:

Thm: [Comaniciu, Meer 2002]

If the profile k is convex and monotonically decreasing above 0, then the sequence $\{\hat{f}_{\sigma,k}(x_j)\}_{j \in \mathbb{N}}$ is monotonically increasing and converges.

Notes:
- both Epanechnikov and Normal profiles are convex above 0.



- convexity is important: $k(t) = \begin{cases} 0 & \text{for } t > 1 \\ c.(1 - t^2) & \text{for } |t| \leq 1 \end{cases}$

yields a kernel $K(x) = \begin{cases} c.(1 - \|x\|^4) & \text{for } \|x\| \leq 1 \\ 0 & \text{for } \|x\| > 1 \end{cases}$

$$\hookrightarrow g(t) = \begin{cases} 2c\,t & \text{for } |t| \leq 1 \\ 0 & \text{for } |t| > 0 \end{cases}$$

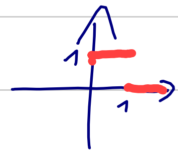
$\Rightarrow$ for $\sigma = 1$ and $P = \{0, 1\} \subset \mathbb{R}$:

$$\begin{cases} \text{if } x_j = 0 \text{ then } x_{j+1} = \dfrac{1 \cdot 2c}{2c} = 1 \\ \text{if } x_j = 1 \text{ then } x_{j+1} = \dfrac{-0 \cdot 2c}{2c} = 0 \end{cases}$$

$\Rightarrow$ Mean-shift oscillates $0 \leftrightarrow 1 \in \mathbb{R}$.

$\boxed{Q}$ Does the sequence $\{x_j\}_{j \in \mathbb{N}}$ itself converge?

Case of the Epanechnikov kernel: $g(t) = 1_{t \leq 1}$ (for $t \geq 0$)



$$\hookrightarrow \forall x \in \mathbb{R}^d, \quad m_{\sigma,g}(x) = \frac{\sum_{i \,|\, \|x - p_i\| \leq \sigma} p_i}{\sum_{i \,|\, \|x - p_i\| \leq \sigma} 1} - x$$

$$= \text{mean}\{p_i \in P \,|\, \|x - p_i\| \leq \sigma\} - x$$

$\Rightarrow$ Whatever x_0, the x_j ($j > 0$) belong to the set of isobarycenters of (subsets of) the pts of P.

$\Rightarrow$ finitely many positions for the x_j ($j > 0$).

Since $m_{\sigma,g} \propto \nabla \hat{f}_{\sigma,k} = 0$ once a maximum of $\hat{f}$ has been reached, the sequence $\{x_j\}_j$ cvges.

General case:

Note: [CM '02] announce cogence in general