

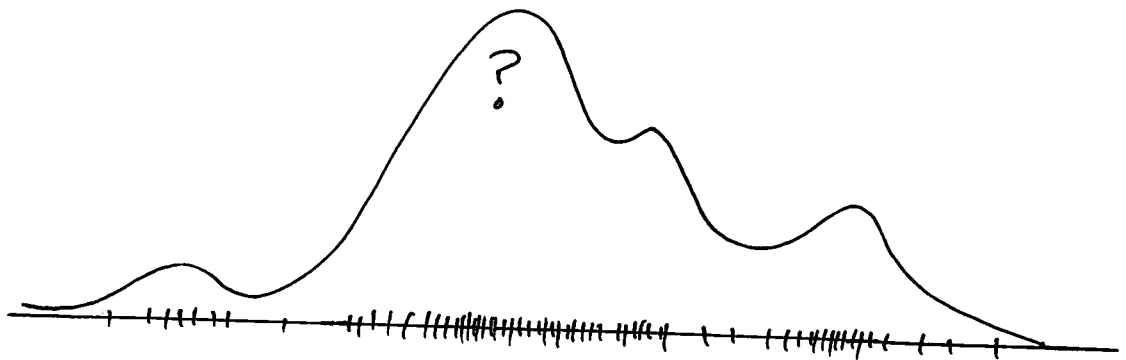
KERNEL DENSITY ESTIMATION

Density Estimation

In density estimation problems, we are given a random sample

$$x_1, \dots, x_n$$

from an unknown density $f(x)$



The objective is to estimate f .

Before examining this task let's first see why it is important.

Classification

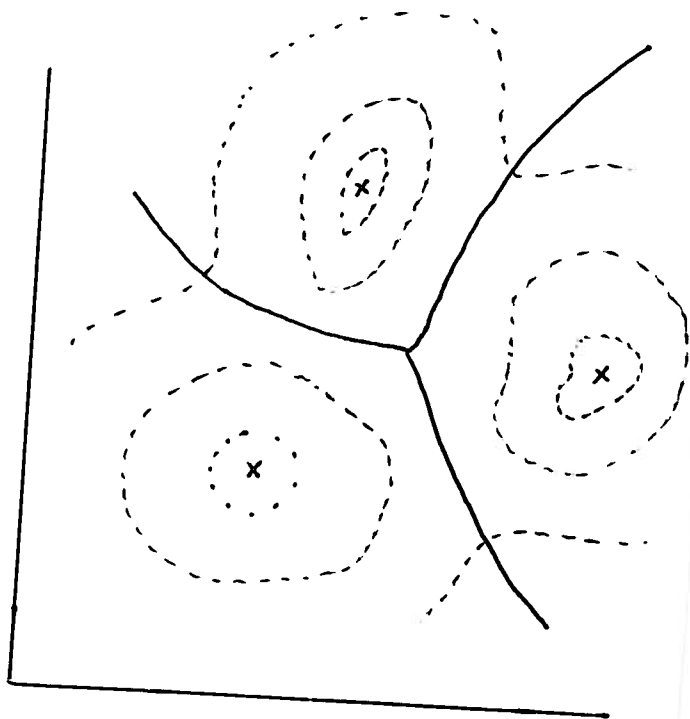
From the formula for the Bayes classifier, a "plug-in" classifier has the form

$$\hat{f}(x) = \arg \max_k \hat{\pi}_k \hat{g}_k(x)$$

where \hat{g}_k is an estimate of the class-conditional density

Clustering

Clusters can be defined by the modes of the density - given a point x , climb the density until you reach a mode. All x reaching the same mode form a cluster.



Novelty Detection

Given a random sample x_1, \dots, x_n from the same nominal distribution, we can estimate its density f , and use

$$\hat{f}(x) \geq \gamma$$

to detect whether a future observation comes from the nominal distribution or a new one

Kernel Density Estimation

A kernel density estimate has the form

$$\textcircled{A} \quad \hat{f}(x) :=$$

where $k_{\sigma}(y)$ is called a kernel

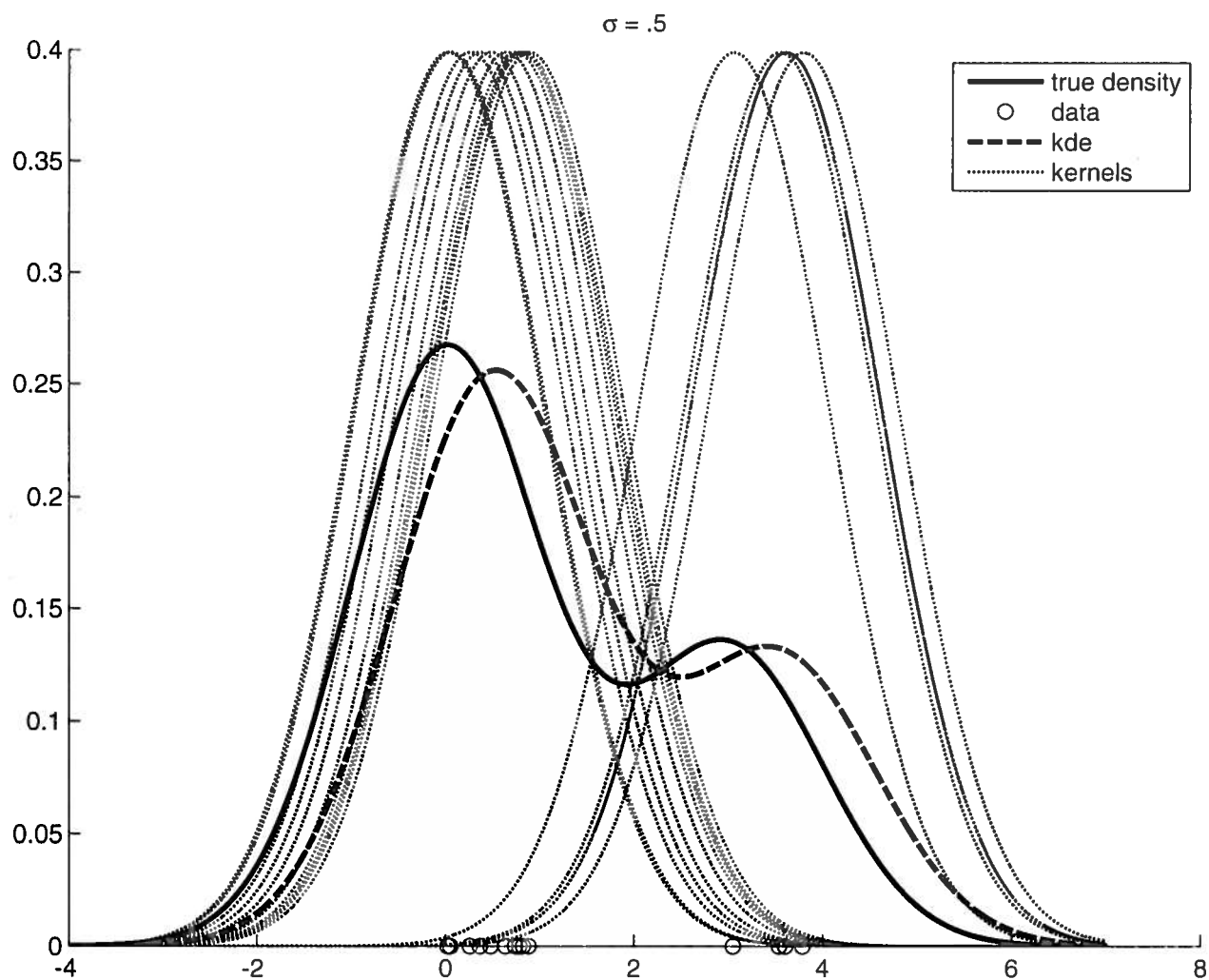
Example | Gaussian kernel

$$k_{\sigma}(y) =$$

Remarks |

- 1) Another term for a KDE is a
- 2) A KDE is nonparametric. Why?
- 3) The Gaussian kernel is the most common.
- 4) The parameter σ is called the

KDE = average of "local" density estimates $k_{\sigma}(x-x_i)$



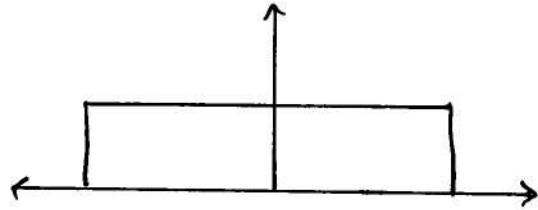
Kernels

A kernel function should satisfy

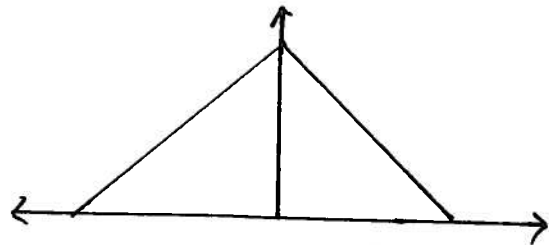
- (B)
- 1)
 - 2)
 - 3)

Examples | (in 1-d)

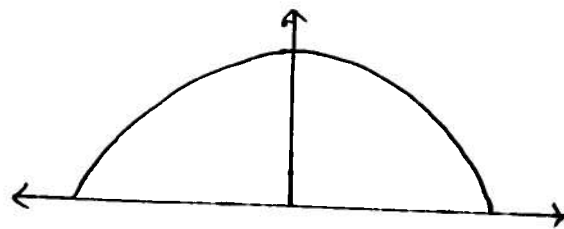
• Uniform kernel



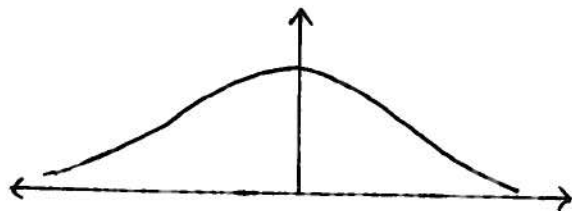
• Triangular kernel



• Epanechnikov kernel
(parabolic)

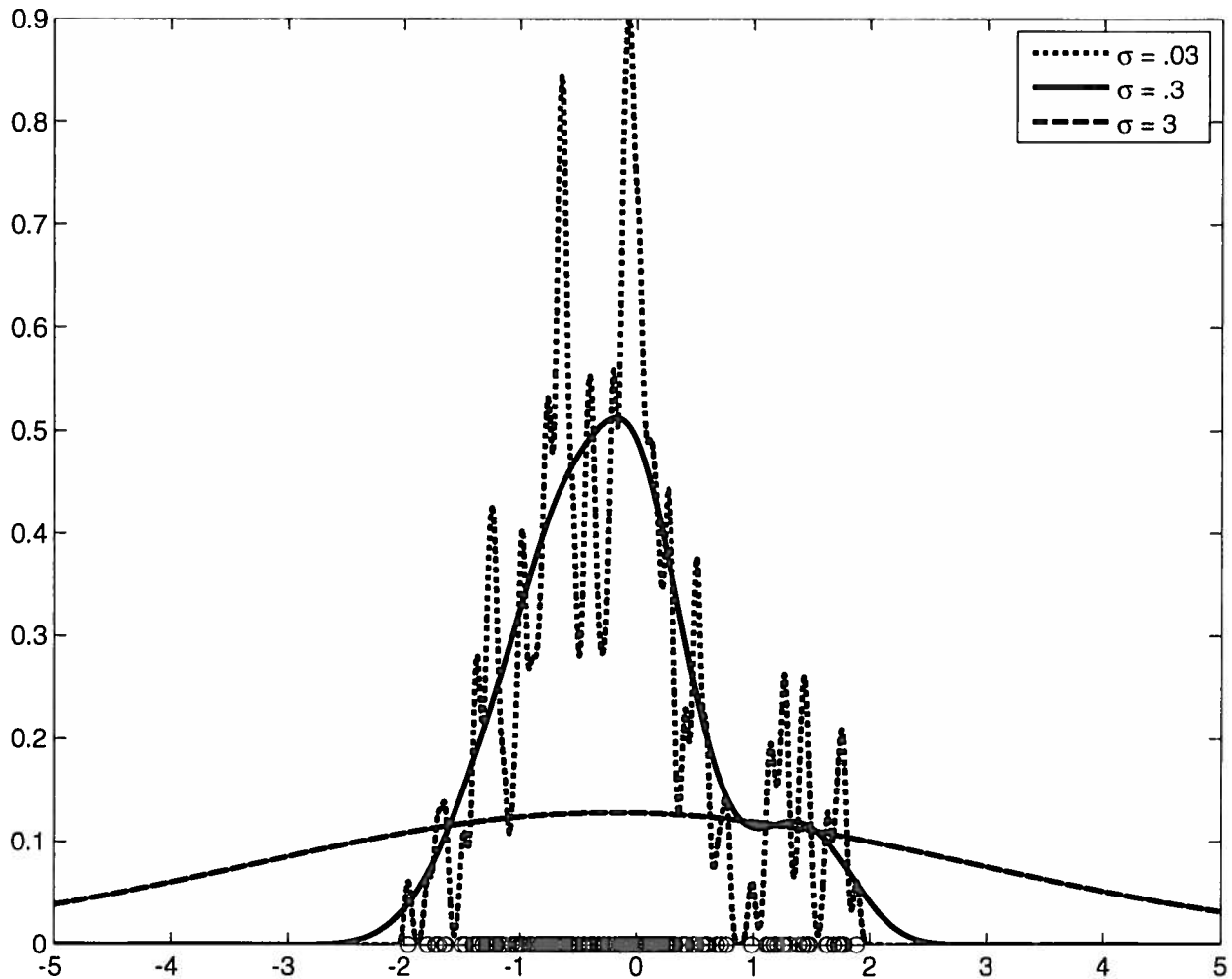


• Cauchy kernel



The accuracy of a KDE depends critically on the

③



Automatically setting σ is a nontrivial problem to which we will return later in the course.

Theorem | Let $\hat{f}_\sigma(x)$ be a KDE based on the kernel k_σ . Suppose $\sigma = \sigma_n$ is such that

$$\bullet \sigma_n \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

$$\bullet n \cdot \sigma_n^d \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

Then

$$E \left\{ \int |\hat{f}_\sigma(x) - f(x)| dx \right\} \rightarrow 0$$

as $n \rightarrow \infty$, regardless of the true density f .

Proof: Devroye and Lugosi, *Combinatorial Methods in Density Estimation*.

Key

$$A. \hat{f}(x) = \frac{1}{n} \sum_{i=1}^n k_{\sigma}(x-x_i)$$

$$k_{\sigma}(y) = (2\pi\sigma^2)^{-\frac{d}{2}} \exp\left\{-\frac{y^2}{2\sigma^2}\right\}$$
$$= \phi(y; 0, \sigma^2 I)$$

Parzen window, $\sigma = \text{bandwidth}$

$$B. 1) \int k_{\sigma}(y) dy = 1$$

$$2) k_{\sigma}(y) \geq 0$$

$$3) k_{\sigma}(y) = \frac{1}{\sigma^d} D\left(\frac{\|y\|}{\sigma}\right) \text{ for some } D$$

C. bandwidth