In density estimation problems, we are given a random sample

\[ x_1, \ldots, x_n \]

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**

The contours of a density can serve to define clusters in a natural way: All \(x\) in the same “connected component” of a contour are in the same cluster.

**Novelty Detection**

Given a random sample \(x_1, \ldots, x_n\) from the same nominal distribution, we can estimate its density \(\hat{f}\), 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

\[ \hat{f}(x) := \]

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 \( \sigma \) is called the
KDE = average of "local" density estimates $k_\sigma(x-x_i)$
Kernels

A kernel function should satisfy

1)
2)
3)

Examples

- Uniform kernel
  \[ k_\sigma(y) = c_\sigma \mathbb{1}_{||y|| \leq \sigma} \]

- Triangular kernel

- Epanechnikov kernel
  \[ \text{(parabolic)} \]

- Cauchy kernel
The accuracy of a KDE depends critically on the

Automatically setting $\sigma$ is a nontrivial problem to which we will return later in the course.
Theorem. Let $\hat{f}_n(x)$ be a KDE based on the kernel $k_0$. Suppose $\sigma = \sigma_n$ is such that

- $\sigma_n \to 0$ as $n \to \infty$
- $n \cdot \sigma_n^{-d} \to \infty$ as $n \to \infty$.

Then

$$E\left\{ \int |\hat{f}_n(x) - f(x)| \, dx \right\} \to 0$$

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

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

\[
k_{\sigma}(y) = \left(2\pi\sigma^2\right)^{-\frac{d}{2}} \exp \left\{ -\frac{y^2}{2\sigma^2} \right\}
\]

\[
= \phi(y; 0, \sigma^2 I)
\]

Parzen window, \( \sigma = \text{bandwidth} \)

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

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

3) \( k_{\sigma}(-y) = k_{\sigma}(y) \)

C. \( \text{bandwidth} \)