First Draft KDE

Owain Morgan

29/11/2021

Kernel Desity Estimation

Suppose we have n independent, identically distributed random variables $X_1, X_2, \ldots, X_n$ (known as samples) from a distribution whose density function, $f(x)$ is not known but is assumed to exist.

Then Kernel Density Estimation (also known as Parzen Windows or the empirical probability density function) provides an easy way of estimating the true density function and as shown by Parzen in 1962, will converge to the true distribution as the number of samples $n \to \infty$.

The Method

Intuitively, one expects areas with a high density of samples to have a higher density function, and those areas which are more sparse to have a lesser density function.

To that end, a kernel function $K(y)$ is introduced which in general is radially symmetric and unimodal. The two most common kernel functions are the rectangular (or window) which weights every point within a certain distance equally and the Gaussian.