Parzen windows

Probability density function (pdf)

The mathematical definition of a continuous probability function, $p(x)$, satisfies the following properties:

1. The probability that x is between two points a and b

$$P(a < x < b) = \int_a^b p(x)dx$$

2. It is non-negative for all real x.

3. The integral of the probability function is one, that is

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

The most commonly used probability function is Gaussian function (also known as Normal distribution)

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)$$

where c is the mean, σ^2 is the variance and σ is the standard deviation.

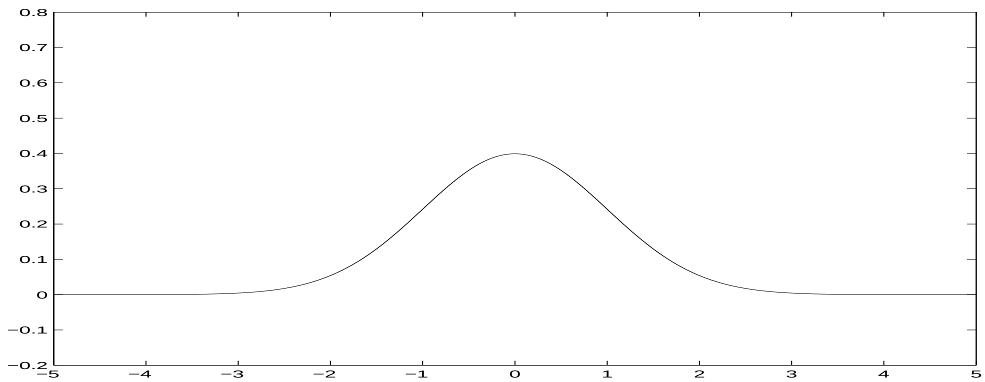


Figure: Gaussian pdf with $c = 0$, $\sigma = 1$.

Extending to the case of a vector $\mathbf{x}$, we have non-negative $p(\mathbf{x})$ with the following properties:

1. The probability that $\mathbf{x}$ is inside a region $\mathcal{R}$

$$P = \int_{\mathcal{R}} p(\mathbf{x}) d\mathbf{x}$$

2. The integral of the probability function is one, that is

$$\int p(\mathbf{x}) d\mathbf{x} = 1$$

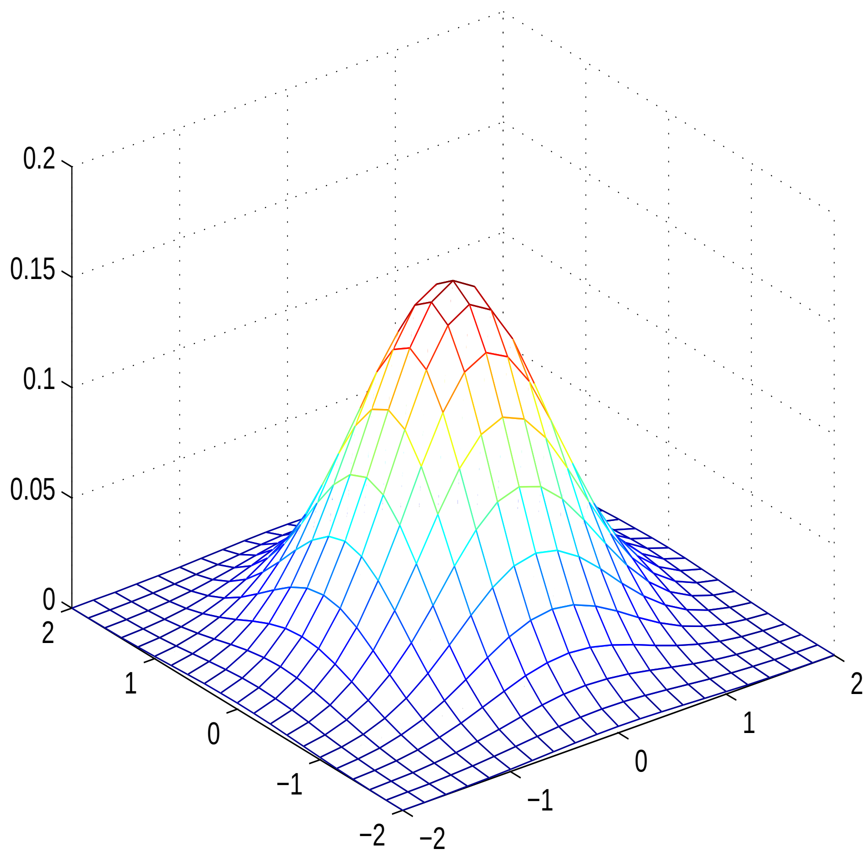


Figure: 2D Gaussian pdf.

Density estimation

Given a set of n data samples $\mathbf{x}_1$,..., $\mathbf{x}_n$, we can estimate the density function $p(\mathbf{x})$, so that we can output $p(\mathbf{x})$ for any new sample $\mathbf{x}$. This is called *density estimation*.

The basic ideas behind many of the methods of estimating an unknown probability density function are very simple. The most fundamental techniques rely on the fact that the probability P that a vector falls in a region $\mathcal{R}$ is given by

$$P = \int_{\mathcal{R}} p(\mathbf{x}) d\mathbf{x}$$

If we now assume that $\mathcal{R}$ is so small that $p(\mathbf{x})$ does not vary much within it, we can write

$$P = \int_{\mathcal{R}} p(\mathbf{x}) d\mathbf{x} \approx p(\mathbf{x}) \int_{\mathcal{R}} d\mathbf{x} = p(\mathbf{x}) V$$

where V is the "volume" of $\mathcal{R}$.

On the other hand, suppose that n samples $\mathbf{x}_1$,..., $\mathbf{x}_n$ are independently drawn according to the probability density function $p(\mathbf{x})$, and there are k out of n samples falling within the region $\mathcal{R}$, we have

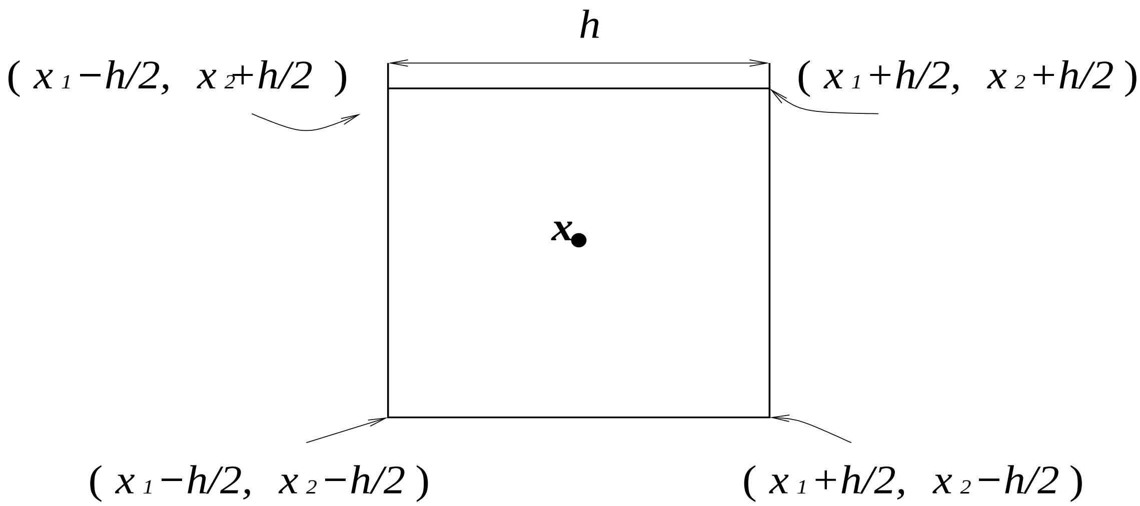
$$P = k/n$$

Thus we arrive at the following obvious estimate for $p(\mathbf{x})$,

$$p(\mathbf{x}) = \frac{k/n}{V}$$

Parzen window density estimation

Consider that $\mathcal{R}$ is a hypercube centered at $\mathbf{x}$ (think about a 2-D square). Let h be the length of the edge of the hypercube, then $V = h^2$ for a 2-D square, and $V = h^3$ for a 3-D cube.



Introduce

$$\phi(\frac{\mathbf{x}_i - \mathbf{x}}{h}) = \begin{cases} 1 & \frac{|x_{ik}-x_k|}{h} \leq 1/2, \quad k = 1, 2 \\ 0 & otherwise \end{cases}$$

which indicates whether $\mathbf{x}_i$ is inside the square (centered at $\mathbf{x}$, width h) or not.

The total number k samples falling within the region $\mathcal{R}$, out of n, is given by

$$k = \sum_{i=1}^{n} \phi(\frac{\mathbf{x}_i - \mathbf{x}}{h})$$

The Parzen probability density estimation formula (for 2-D) is given by

$$p(\mathbf{x}) = \frac{k/n}{V}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^2} \phi(\frac{\mathbf{x}_i - \mathbf{x}}{h})$$

$\phi(\frac{\mathbf{x}_i - \mathbf{x}}{h})$ is called a window function. We can generalize the idea and allow the use of other window functions so as to yield other Parzen window density estimation methods. For example, if Gaussian function is used, then (for 1-D) we have

$$p(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - x)^2}{2\sigma^2}\right)$$

This is simply the average of n Gaussian functions with each data point as a center. σ needs to be predetermined.

Example: Given a set of five data points $x_1 = 2$, $x_2 = 2.5$, $x_3 = 3$, $x_4 = 1$ and $x_5 = 6$, find Parzen probability density function (pdf) estimates at $x = 3$, using the Gaussian function with $\sigma = 1$ as window function.

Solution:

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_1 - x)^2}{2}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(2 - 3)^2}{2}\right) = 0.2420$$

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_2 - x)^2}{2}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(2.5 - 3)^2}{2}\right) = 0.3521$$

$$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x_3-x)^2}{2}\right)=0.3989$$

$$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x_4-x)^2}{2}\right)=0.0540$$

$$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x_5-x)^2}{2}\right)=0.0044$$

so

$$p(x=3)=(0.2420+0.3521+0.3989$$

$$+0.0540+0.0044)/5 \quad =0.2103$$

The Parzen window can be graphically illustrated next. Each data point makes an equal contribution to the final pdf denoted by the solid line.

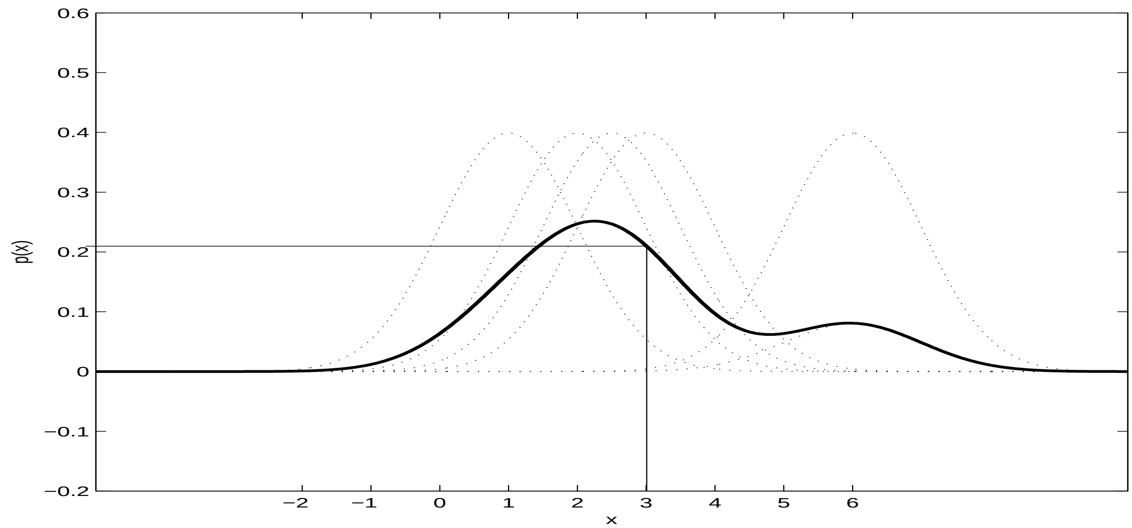


Figure above: The dotted lines are the Gaussian functions centered at 5 data points.

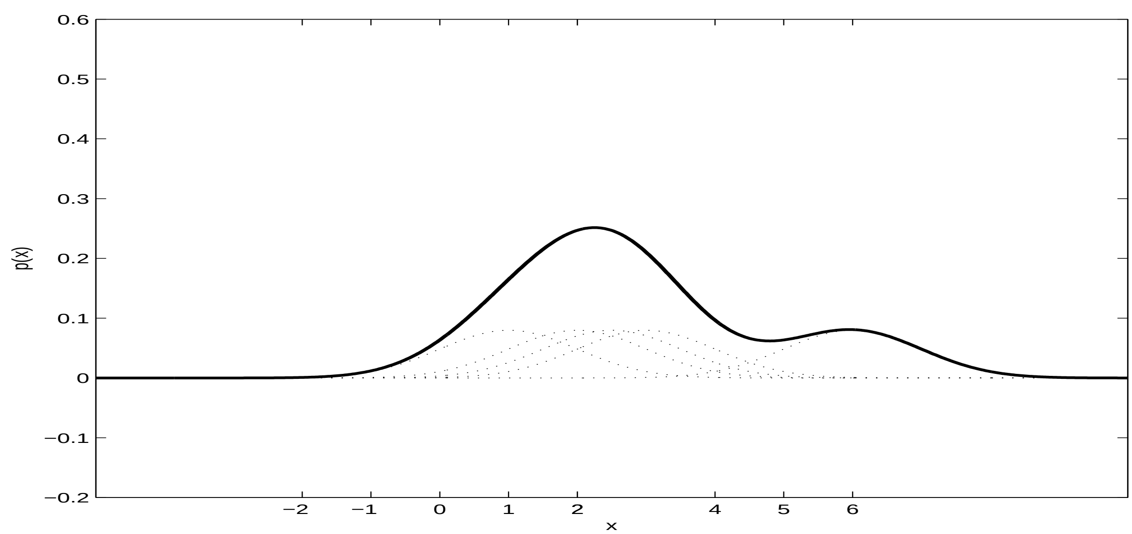


Figure above: The Parzen window pdf function sums ups 5 dotted line.