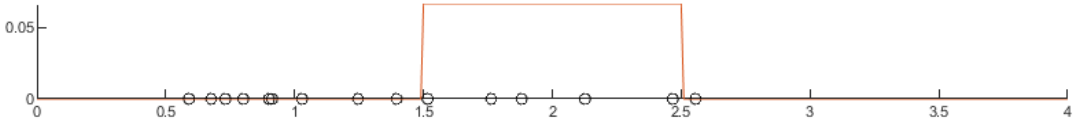
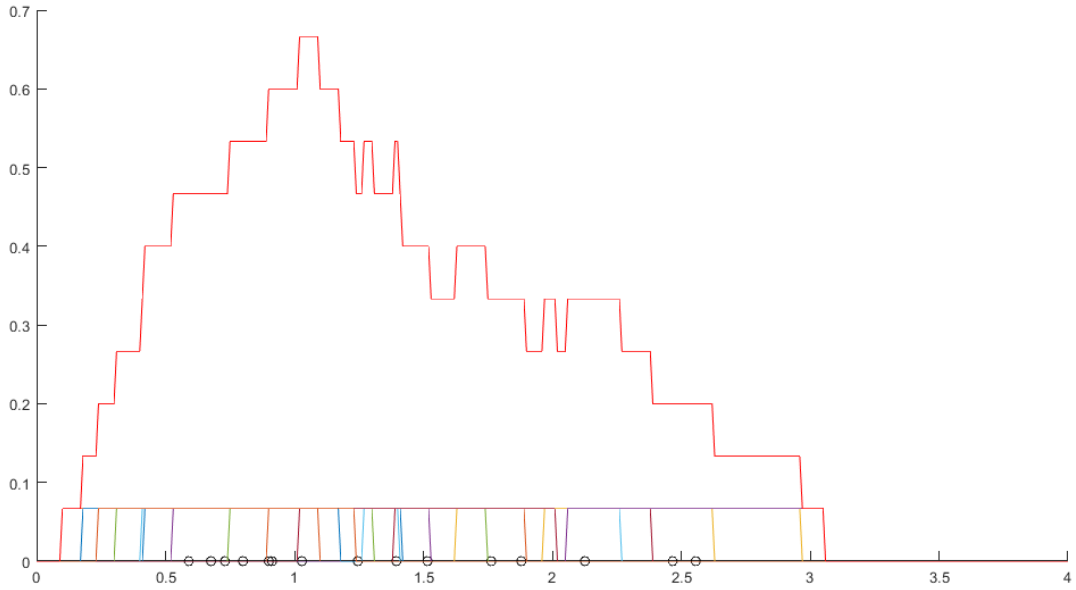
Parzen window density estimation

In the Parzen window method, we assume that for any given point, the probability density function (PDF) will depend on the number (or, more precisely, proportion) of training set data points that lie within a certain neighborhood of that point. Consider the example below (the window size is equal to 1):

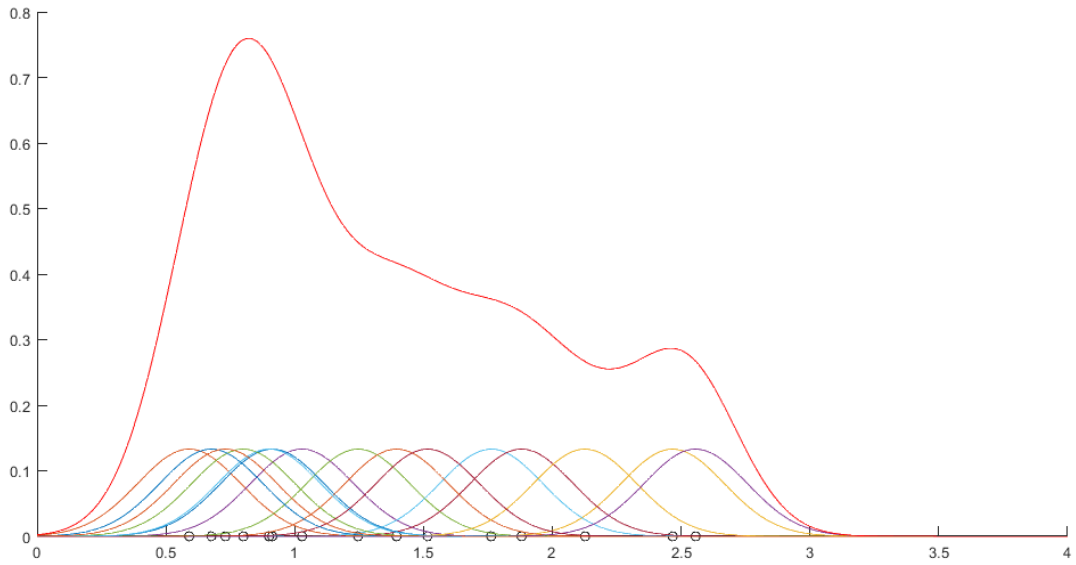


Suppose we want to calculate the PDF at the point 2. 5 of 15 points are inside the window (their distance from the point 2 is less than half the window size). This means, the PDF at 2 will be equal to 5/15 = 0.33.

Notice, that the simple interpretation is that each of the N data points contributes to the estimated PDF with a box centered around it with width equal to the window size (w) and height $\frac{1}{w*N}$ (see below)



The obtained density function is quite “jagged”. Therefore, various kernel functions may be used to smooth it (the total area under PDF must always be equal to 1, though!). One of the most popular is the gaussian kernel:



In naive Bayes classifier, every feature is analyzed separately. In other applications, however, the principle can easily be extended to multivariate distributions.