

These two main types of nonparametric methods, the kernel method and nearest-neighbor method, while appearing to be quite different, are related [Hand81]. The general approach to classification involves estimation of  $p(\mathbf{x}_0)$ , the value of the distribution density at  $\mathbf{x}_0$ . This, then, is the important quantity to estimate. In the previous sections we assumed  $f$  to be known, so that we only needed to estimate parameters to obtain an estimate of the density. With the nonparametric methods we directly estimate  $p(\mathbf{x}_0)$ .

Suppose for simplicity that there are only two classes and a single variable. Then a simple approach is to estimate  $p$  by a histogram. From the histograms for both classes we can estimate  $p(\mathbf{x}_0|c_1)$  and  $p(\mathbf{x}_0|c_2)$  and make an assignment. However, when the number of features is large, it is difficult to obtain a meaningful histogram. Consider so simple a case as a set of two-valued (binary) variables. Then the number of bins in which the histogram is to be evaluated will be  $2^d$ ,