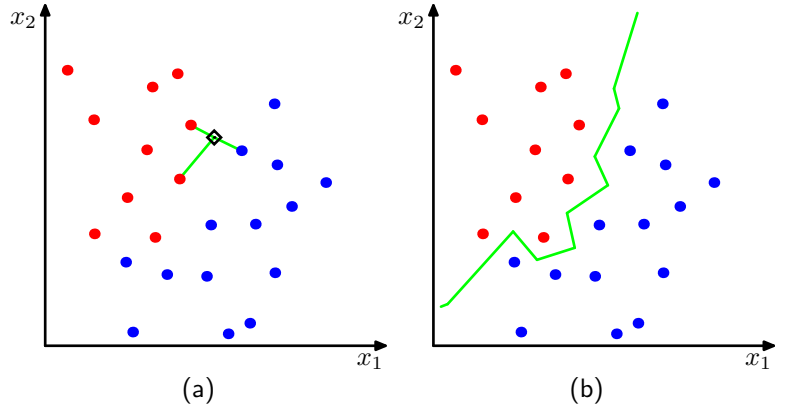


Figure 3.16 (a) In the K -nearest-neighbour classifier, a new point, shown by the black diamond, is classified according to the majority class membership of the K closest training data points, in this case $K = 3$. (b) In the nearest-neighbour ($K = 1$) approach to classification, the resulting decision boundary is composed of hyperplanes that form perpendicular bisectors of pairs of points from different classes.



with each class:

$$p(\mathbf{x}|\mathcal{C}_k) = \frac{K_k}{N_k V}. \quad (3.187)$$

Similarly, the unconditional density is given by

$$p(\mathbf{x}) = \frac{K}{NV} \quad (3.188)$$

and the class priors are given by

$$p(\mathcal{C}_k) = \frac{N_k}{N}. \quad (3.189)$$

We can now combine (3.187), (3.188), and (3.189) using Bayes' theorem to obtain the posterior probability of class membership:

$$p(\mathcal{C}_k|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_k)p(\mathcal{C}_k)}{p(\mathbf{x})} = \frac{K_k}{K}. \quad (3.190)$$

We can minimize the probability of misclassification by assigning the test point \mathbf{x} to the class having the largest posterior probability, corresponding to the largest value of K_k/K . Thus, to classify a new point, we identify the K nearest points from the training data set and then assign the new point to the class having the largest number of representatives amongst this set. Ties can be broken at random. The particular case of $K = 1$ is called the *nearest-neighbour* rule, because a test point is simply assigned to the same class as the nearest point from the training set. These concepts are illustrated in Figure 3.16.

An interesting property of the nearest-neighbour ($K = 1$) classifier is that, in the limit $N \rightarrow \infty$, the error rate is never more than twice the minimum achievable error rate of an optimal classifier, i.e., one that uses the true class distributions (Cover and Hart, 1967).

As discussed so far, both the K -nearest-neighbour method and the kernel density estimator require the entire training data set to be stored, leading to expensive