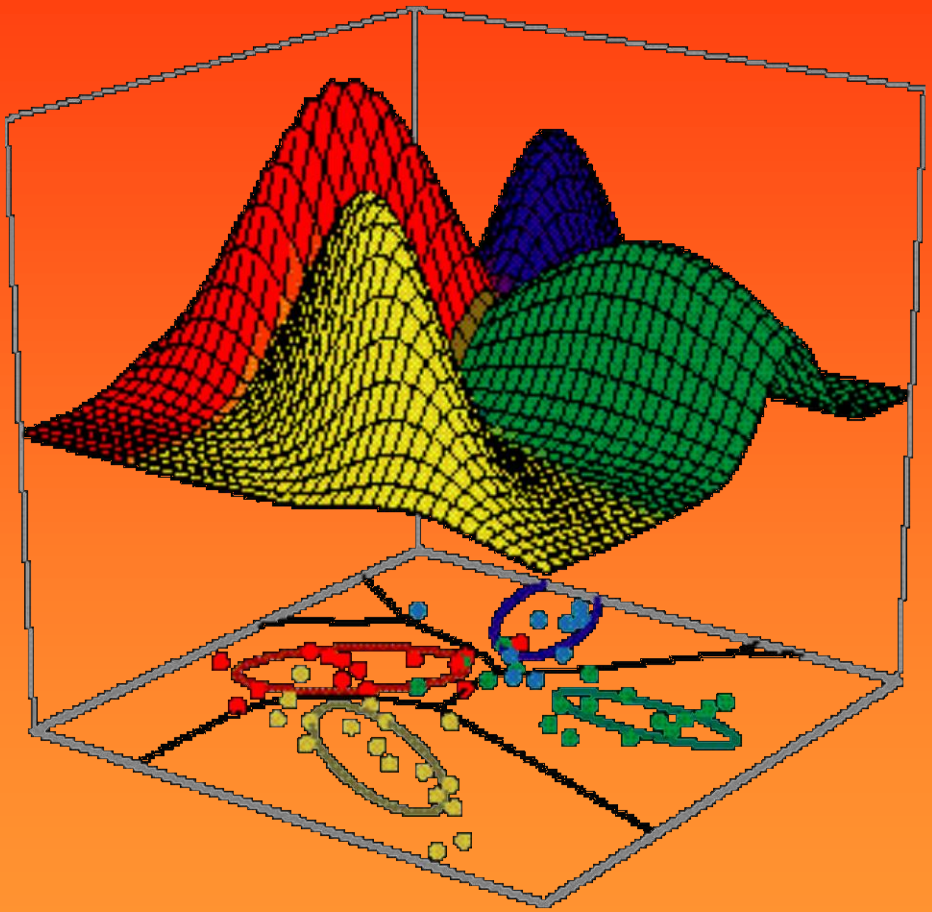


Pattern Classification

All materials in these slides were taken from *Pattern Classification (2nd ed) by R. O. Duda, P. E. Hart and D. G. Stork, John Wiley & Sons, 2000* with the permission of the authors and the publisher

Capitulo 4 (parte 2):
Clasificacion no parametrica
(Secciones 4.3-4.5)

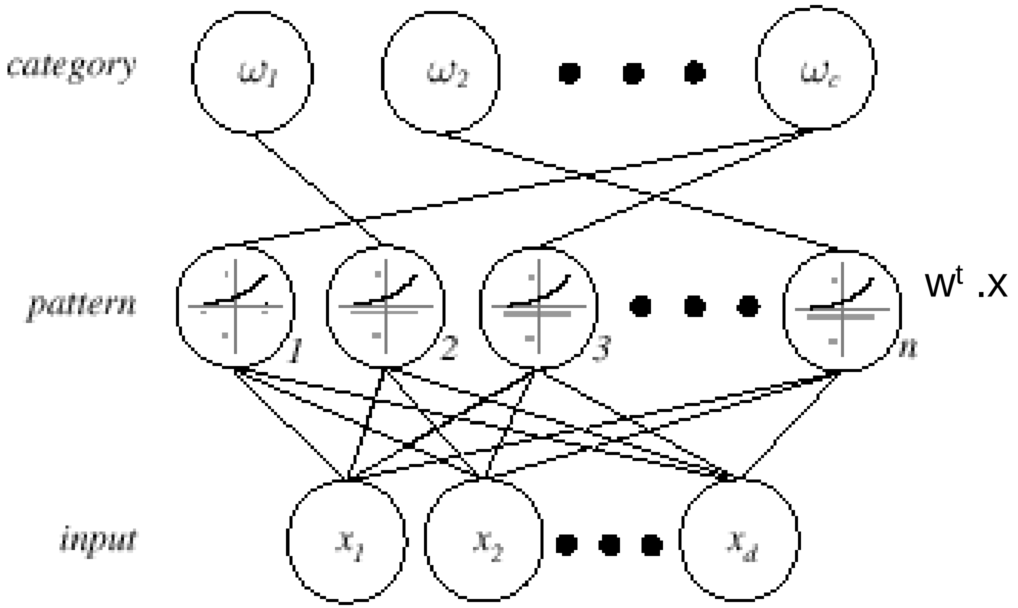
Parzen Window

Kn –Nearest Neighbor Estimation

The Nearest-Neighbor Rule

Probabilistic Neural Networks

- Los vectores a clasificar (ingreso) estan conectados con n patrones, con pesos w estimados con la muestra de entrenamiento.
- Los patrones estan conectados con las c clases, tambien de acuerdo a la muestra de entrenamiento



• Algoritmo de entrenamiento

1. Normalizar cada muestra de entrenamiento, $|| \mathbf{x_i} ||=1$

2. Poner la primera muestra de entrenamiento $\mathbf{x_1}$ en las unidades de entrada

3. Definir los pesos de los enlaces de la muestra $\mathbf{x_1}$ con el primer patron como $w_{1k} = x_{1k}$ *con* $k=1,..,d.$

4. Hacer un enlace desde el patron 1 a la clase conocida de la muestra **:** $\boldsymbol{x_1} \in \omega_j$ *entonces* $p_1 \dashrightarrow \omega_j$

5. Repetir el proceso para todos los patrones de entrenamiento

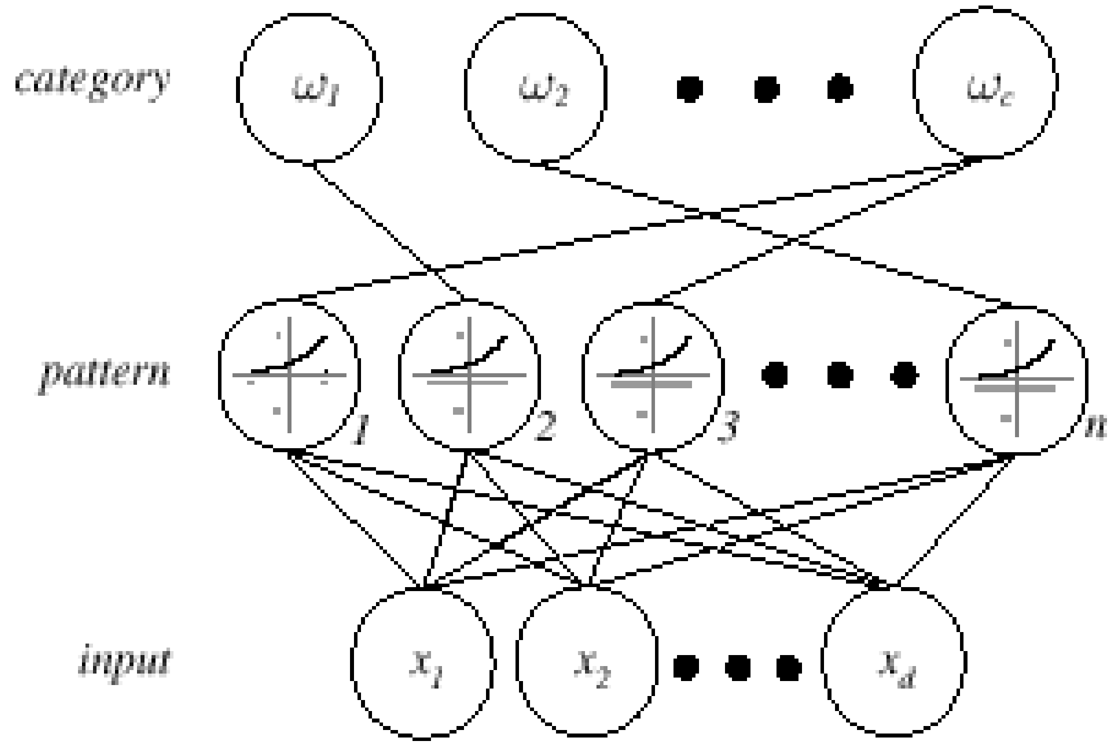


FIGURE 4.9. A probabilistic neural network (PNN) consists of d input units, n pattern units, and c category units.

Cada unidad patron tiene como vector de pesos a una muestra de entrenamiento. La etiqueta de dicha muestra produce un enlace con las unidades de categoria.

Algoritmo de clasificacion

- Normalizar el vector x a clasificar y ponerlo en la unidad de entrada
- Cada patron calcula el producto interno con sus pesos para armar la red de activacion

$$net_k = w_k^t . x$$

$$f(net_k) = \exp\left[\frac{net_k - 1}{\sigma^2}\right] \propto \varphi\left(\frac{x - w_k}{h_n}\right)$$

  y emite una funcion no lineal hacia la categoria ω_j donde esta conectada
- Cada categoria suma las contribuciones de todos los patrones conectados con ella

$$P_n(x \mid \omega_j) = \sum_{i \to j} \varphi_i \propto P(\omega_j \mid x)$$

- Clasifica seleccionando el maximo valor de $P_n(x \mid \omega_j)$

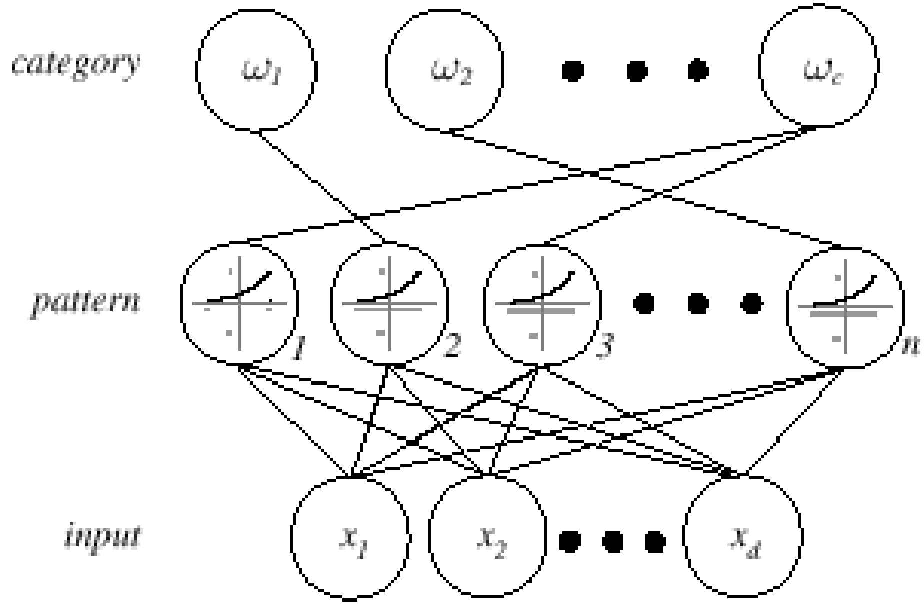


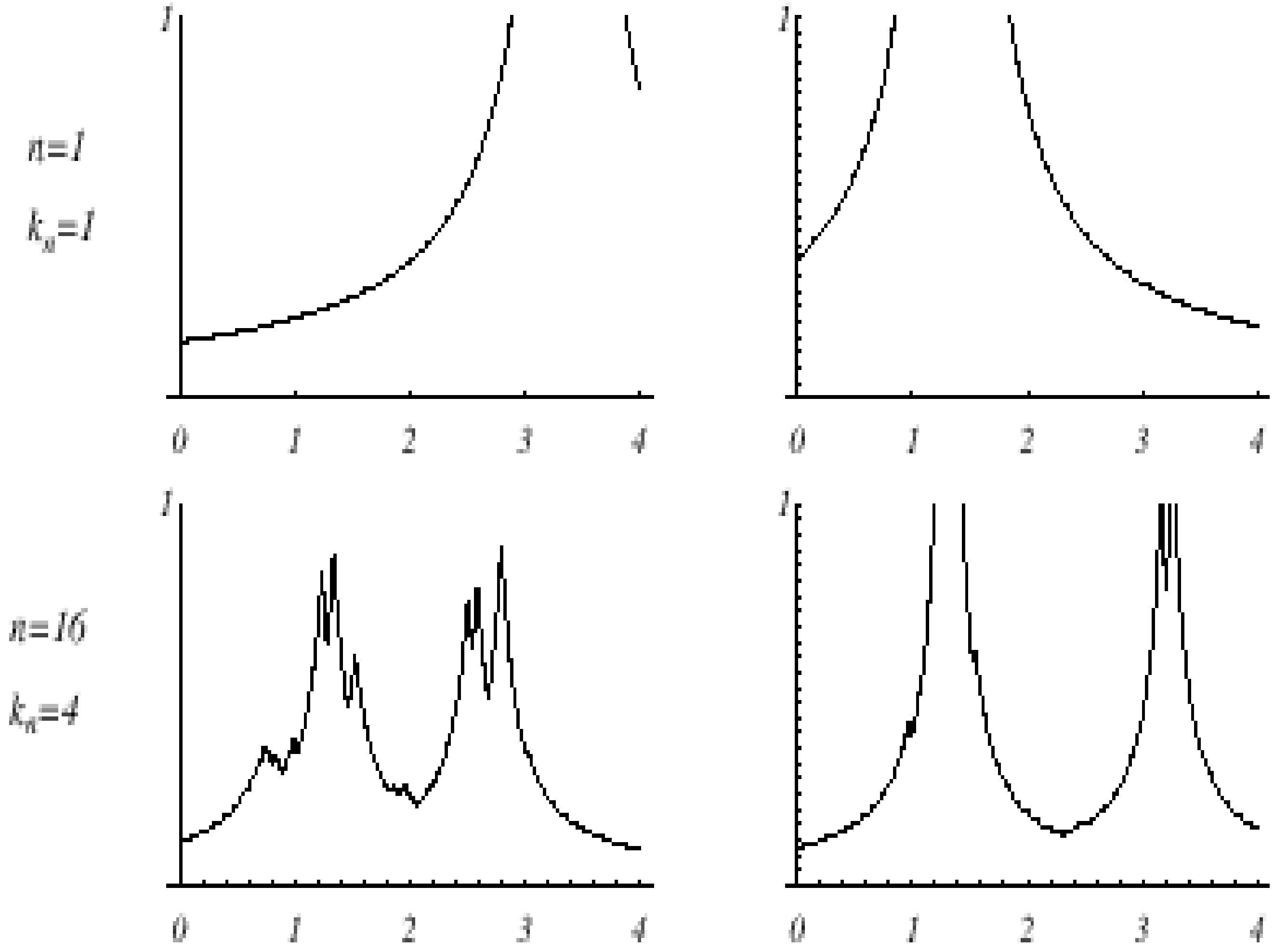
FIGURE 4.9. A probabilistic neural network (PNN) consists of d input units, n pattern units, and c category units. Each pattern unit forms the inner product of its weight vector and the normalized pattern vector $\mathbf{x}$ to form $z = \mathbf{w}^t\mathbf{x}$, and then it emits $\exp[(z-1)/\sigma^2]$. Each category unit sums such contributions from the pattern unit connected to it. This ensures that the activity in each of the category units represents the Parzen-window density estimate using a circularly symmetric Gaussian window of covariance $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the $d \times d$ identity matrix. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

- K_n – estimacion de vecinos mas cercanos
  - Objetivo: solucionar el problema de la mejor funcion de ventana
    - Dejemos que el volumen sea una funcion de los datos de entrenamiento
    - Centremos una celda en x y dejemos crecer el volumen hasta capturar k_n muestras$(k_n = f(n))$
    - k_n son llamados k_n vecinos mas cercanos x

Pueden ocurrir dos cosas:

- Densidad es alta cerca de x; por lo tanto la celda puede ser pequeña lo cual da una buen resolucion.
- Densidad es pequeña; por lo tanto la celda y se detendra cuando encuentre una region de alta densidad.

Podemos obtener una familia de estimadores poniendo $k_n=k_1/\sqrt{n}$ y eligiendo diferentes valores de k_1


n=1
k_n=1
n=16
k_n=4
1
0
1
2
3
4

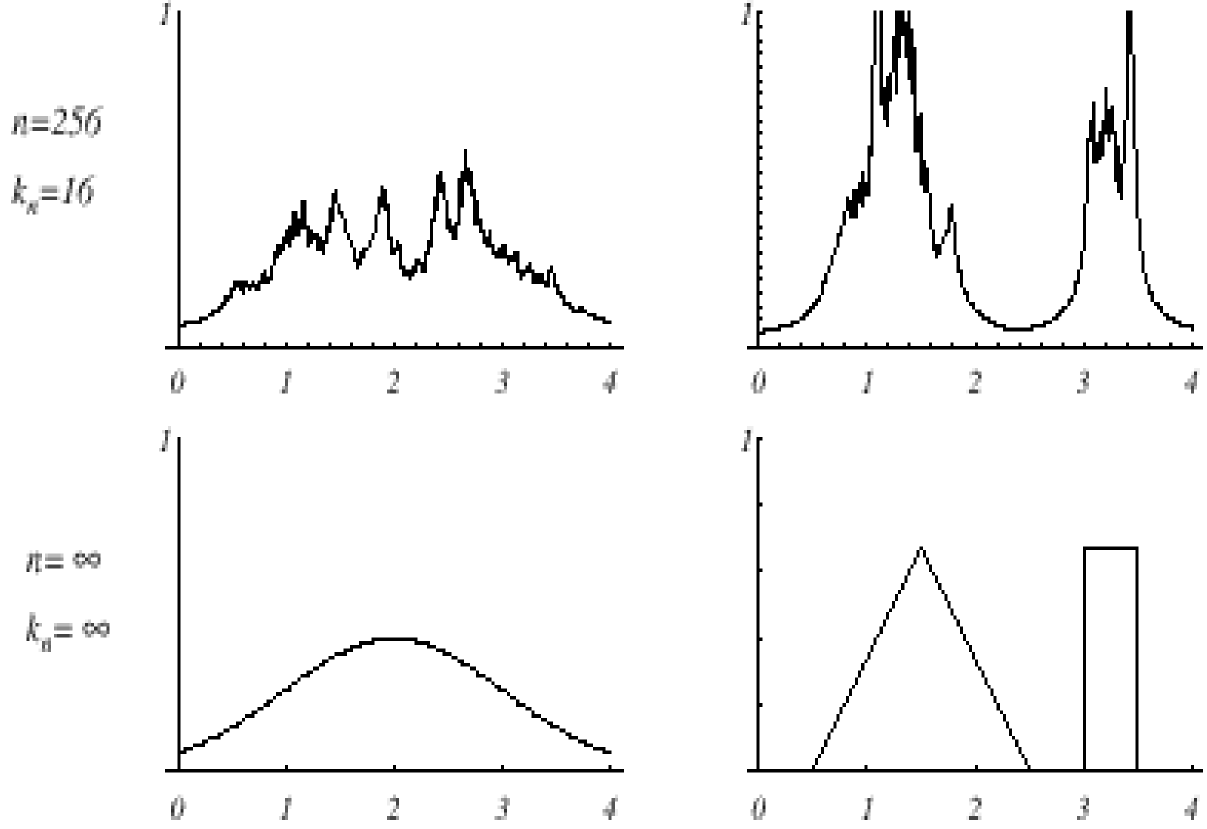


FIGURE 4.12. Several k-nearest-neighbor estimates of two unidimensional densities: a Gaussian and a bimodal distribution. Notice how the finite n estimates can be quite "spiky." From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

- Estimation de las probabilidades a posteriori

  - Objetivo: estimar $P(\omega_i \mid x)$ de un conjunto de n muestras etiquetadas
    - Pongamos un volumen V alrededor de x y capturemos k muestras
    - k_i muestras alrededor k resultaron de la clase ω_i entonces:

$$p_n(x, \omega_i) = k_i / n.V$$

Un estimador de $p_n(\omega_i \mid x)$ es:

$$p_n(\omega_i \mid x) = \frac{p_n(x, \omega_i)}{\sum_{j=1}^{j=c} p_n(x, \omega_j)} = \frac{k_i}{k}$$

- k_i/k es la fraccion de las muestras dentro de la celda que estan etiquetadas con ω_i

- Para una tasa de error minima, la categoria con mas frecuencia de representacion dentro de la celda es seleccionada

- Si k es grande y la celda es suficientemente chica , el desempeño se acercara al mejor posible.

- La regla del vecino mas cercano

  - Sea $D_n = \{x_1, x_2, \ldots, x_n\}$ un conjunto de n patrones etiquetados

  - Sea $x' \in D_n$ el patron mas cercano al punto de test x entonces la regla del vecino mas cercano para clasificar x es asignar la etiqueta asociada con x'

  - La regla del vecino mas cercano lleva a un error mayor que el minimo posible, la tasa de Bayes.

  - Si el numero d epatrones es largo (ilimitado), el error del clasificador de vecinos mas cercano no es nunca peor que dos veces la tasa de Bayes

  - Si $n \rightarrow \infty$, siempre es posible encontrar x' suficientemente cerca de x tal que:

$$P(\omega_i \mid x') \cong P(\omega_i \mid x)$$

Ejemplo:

$x = (0.68, 0.60)^t$

Patrones	Etiquetas	Probabilidades a posteriori estimadas
(0.50, 0.30)	ω_2	0.25
	ω_3	$0.75 = P(\omega_m / x)$
(0.70, 0.65)	ω_5	0.70
	ω_6	0.30

<u>Decision:</u> ω_5 es la etiqueta asignada a x

Si $P(\omega_m \mid x) \simeq 1$, entonces la regla del vecino mas cercano es casi siempre la eleccion de Bayes

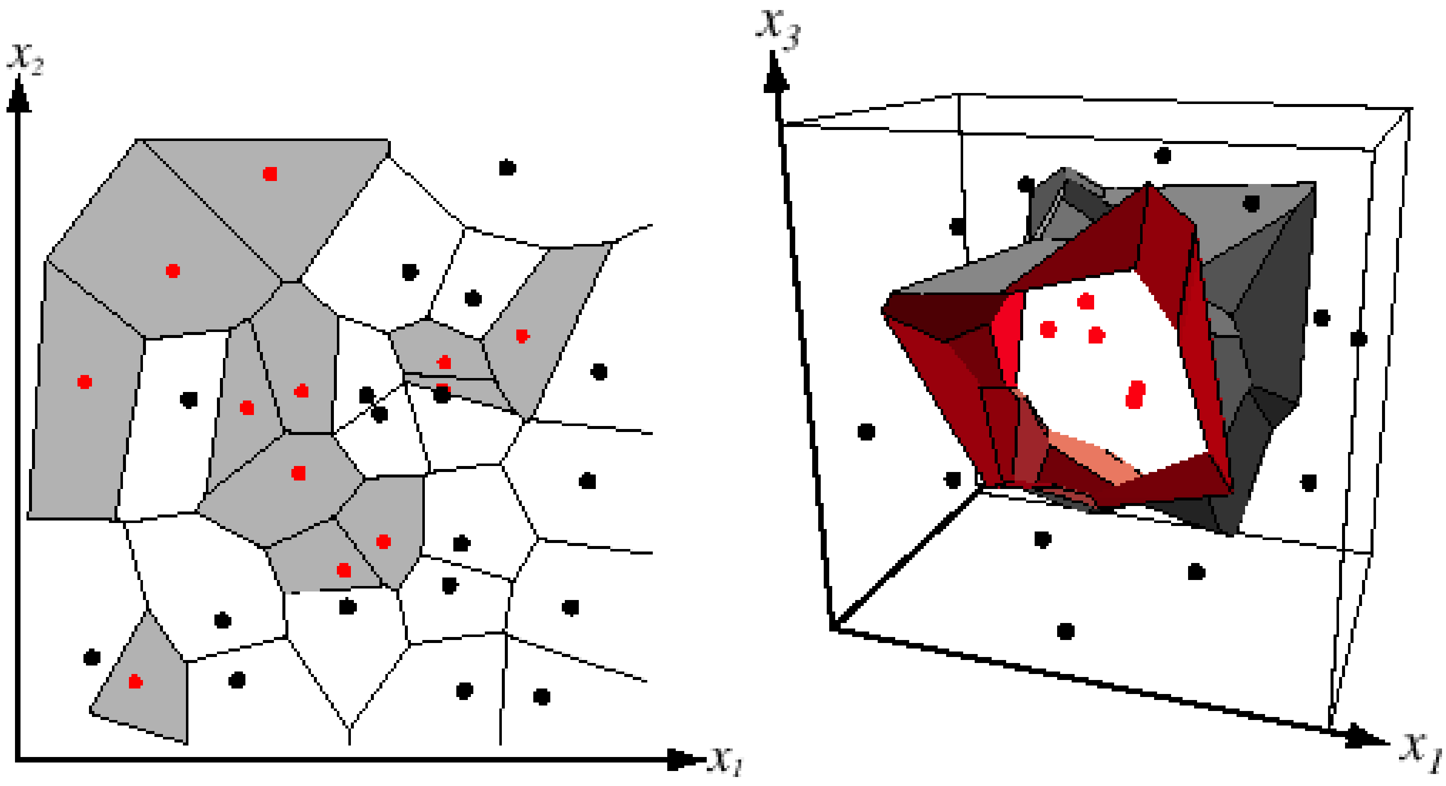


FIGURE 4.13. In two dimensions, the nearest-neighbor algorithm leads to a partitioning of the input space into Voronoi cells, each labeled by the category of the training point it contains. In three dimensions, the cells are three-dimensional, and the decision boundary resembles the surface of a crystal. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

- La regla de k– vecinos mas cercanos

  - Objetivo: Clasificar a x con la etiqueta mas representada (en frecuencia) en los k patrones mas cercanos a x de la muestra de entrenamiento global

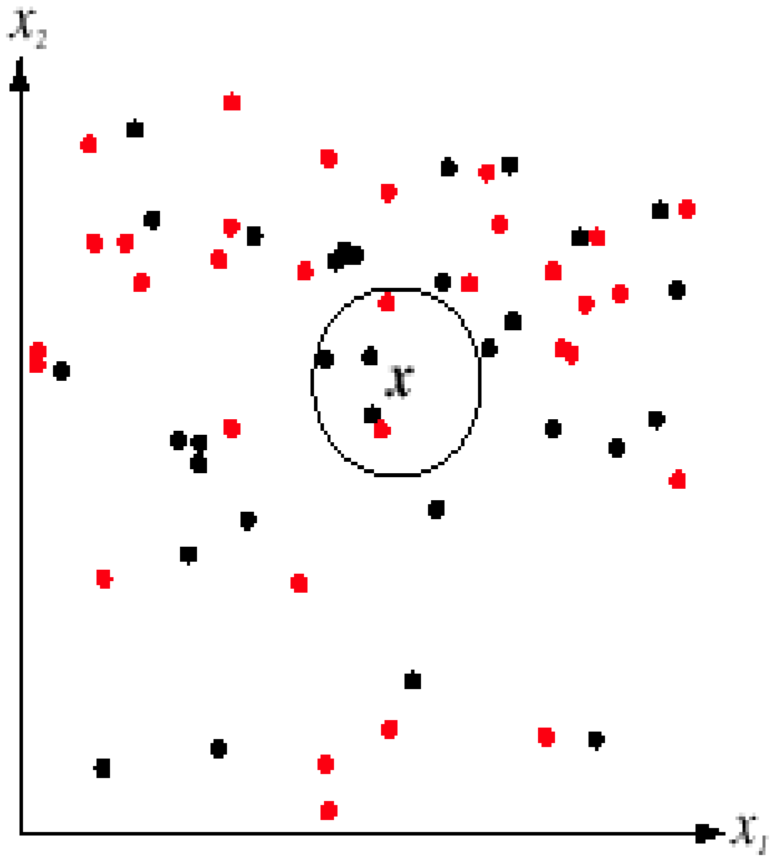


FIGURE 4.15. The k-nearest-neighbor query starts at the test point $\mathbf{x}$ and grows a spherical region until it encloses k training samples, and it labels the test point by a majority vote of these samples. In this $k = 5$ case, the test point $\mathbf{x}$ would be labeled the category of the black points. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Ejemplo:

$k = 3$ (valor impar) y $x = (0.10, 0.25)^t$

Prototypes	Labels
(0.15, 0.35)	ω_1
(0.10, 0.28)	ω_2
(0.09, 0.30)	ω_5
(0.12, 0.20)	ω_2

Vectores mas cercanos a x donde sus etiquetas son:

$$\{(0.10, 0.28, \omega_2); (0.12, 0.20, \omega_2); (0.15, 0.35, \omega_1)\}$$

El esquema de votacion asigna la etiqueta ω_2 a x pues ω_2 es el mas representado