APRENDIZAJE AUTOMÁTICO: TERCER CURSO DEL GRADO
DE ING. INFORMÁTICA EN COMPUTACION

Clasificación: K-vecinos más cercanos: K-NN

César Hervás-Martínez
Grupo de Investigación AYRNA

Departamento de Informática y Análisis Numérico
Universidad de Córdoba
Campus de Rabanales. Edificio Einstein.
Email: chervas@uco.es

2019-2020

INTRODUCCION
K-NN (K-NEAREST NEIGHBOUR)

- **Un nuevo caso se va a clasificar en la clase más frecuente a la que pertenecen sus K vecinos más cercanos**

- **Idea muy simple e intuitiva**

- **Fácil implementación**

- **No hay modelo explícito**

- **Razonamiento basado en casos “*Case Based Reasoning*” (*CBR*)**

INTRODUCCION
K-NN (K-NEAREST NEIGHBOUR)
Notación para el paradigma K-NN

		X1	…	Xj	…	Xn	C
(x1,c1)	1	x11	…	x1j	…	x1n	c1
	. . .	. . .		. . .		. . .	. . .
(xi,ci)	i	xi1	…	xij	…	xin	ci
	. . .	. . .		. . .		. . .	. . .
(xN,cN)	N	xN1	…	xNj	…	xNn	cN
x	N+1	xN+1,1	…	xN+1,j	…	xN+1,n	?

Pseudocódigo para el clasificador K-NN

El algoritmo K-NN básico

COMIENZO

Entrada: D = {(x1, c1), . . . , (xN, cN)}

x = (x1, . . . , xn) nuevo caso a clasificar

PARA todo objeto ya clasificado (xi, ci)

calcular di = d(xi, x)

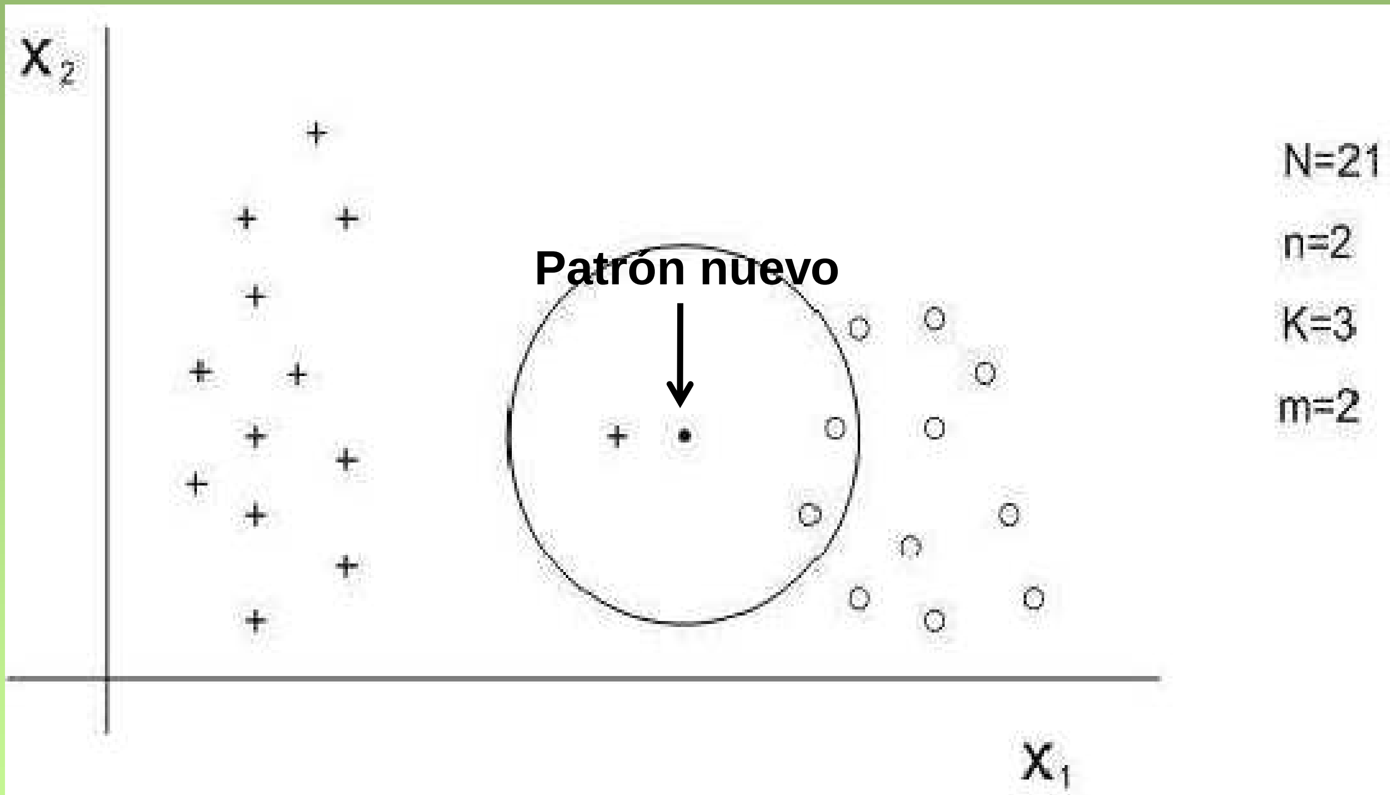
Ordenar di(i = 1, . . . ,N) en orden ascendente

Quedarnos con los K casos D_x^K ya clasificados más cercanos a x

Asignar a x la clase más frecuente en D_x^K

FIN

ALGORITMO K-NN BÁSICO



Nº de patrones

Nº de variables

Nº de clases

Ejemplo de aplicación del algoritmo K-NN básico

ALGORITMO K-NN BÁSICO: Ejemplo 1NN

Dado el siguiente ejemplo con 4 instancias, 3 variables independientes y dos clases:

- **(x1)=(1.4,0.8,1.2), y (x2)=(1.2,0.7,1.9) de la clase positiva**
- **(x3)=(1.0,0.8,0.9) y (x4)=(0.8,1.0,0.0) de la clase negativa**

Calcular la etiqueta del clasificador del vecino más cercano (1NN) con respecto al patrón de etiqueta desconocida (x)=(0.7,1.2,1.3)

$$\mathrm{d}(\mathbf{x}_1,\mathbf{x})=\sqrt{\sum_{i=1}^{3}(x_{1i}-x_i)^2}=\sqrt{((1.4-0.7)^2+(0.8-1.2)^2+(1.2-1.3)^2}=0.81$$

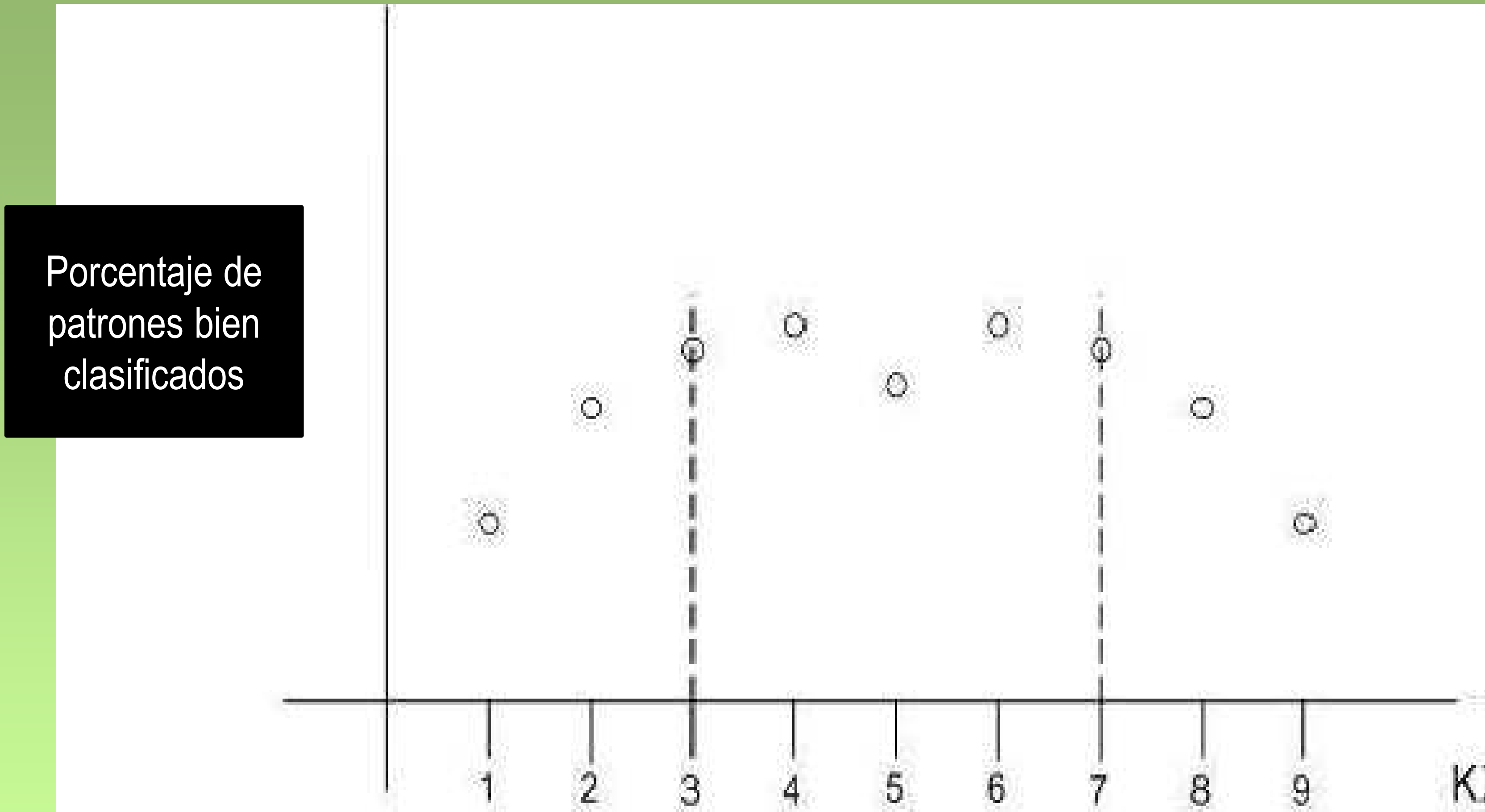
$$\mathrm{d}(\mathbf{x}_2,\mathbf{x})=\sqrt{\sum_{i=1}^{3}(x_{2i}-x_i)^2}=\sqrt{((1.2-0.7)^2+(0.7-1.2)^2+(1.9-1.3)^2}=0.93$$

$$\mathrm{d}(\mathbf{x}_3,\mathbf{x})=\sqrt{\sum_{i=1}^{3}(x_{3i}-x_i)^2}=\sqrt{((1-0.7)^2+(0.8-1.2)^2+(0.9-1.3)^2}=0.64$$

$$\mathrm{d}(\mathbf{x}_4,\mathbf{x})=\sqrt{\sum_{i=1}^{3}(x_{4i}-x_i)^2}=\sqrt{((0.8-0.7)^2+(1.0-1.2)^2+(0.0-1.3)^2}=1.32$$

De esta forma el patrón x pertenece a la clase negativa. ¿Y si K fuera igual a 2?

ALGORITMO K-NN BÁSICO



Ejemplo de la no monotonicidad del porcentaje de patrones bien clasificados en función de K

VARIANTES DEL ALGORITMO K-NN

Hay distintas variantes del algoritmo en función de cómo calculamos la clase a partir de los K vecinos más cercanos:

- **Voto por mayoría (*majority voting*).**
- **Voto ponderado: en función de la distancia di el voto vale más o menos (más peso cuando la distancia sea más pequeña).**
- **En el caso del uso de votos ponderados, es frecuente que relajemos la restricción de los K vecinos y en muchos casos se usan todos los ejemplos del conjunto de datos.**

Variantes del algoritmo K-NN básico

- **K-NN con rechazo**
- **K-NN con distancia media**
- **K-NN con distancia mínima**
- **K-NN con ponderación de vecinos**
- **K-NN con ponderación de variables**

K-NN CON RECHAZO

Para clasificar un caso se exigen ciertas garantías, por ejemplo distancia mínima.

- **Si no se tienen puedo dejar el caso sin clasificar**
- **Umbral prefijado**
- **Mayoría absoluta**

K-NN CON DISTANCIA MEDIA

Ejemplo de ilustración del algoritmo K-NN con distancia media

K-NN con distancia mínima

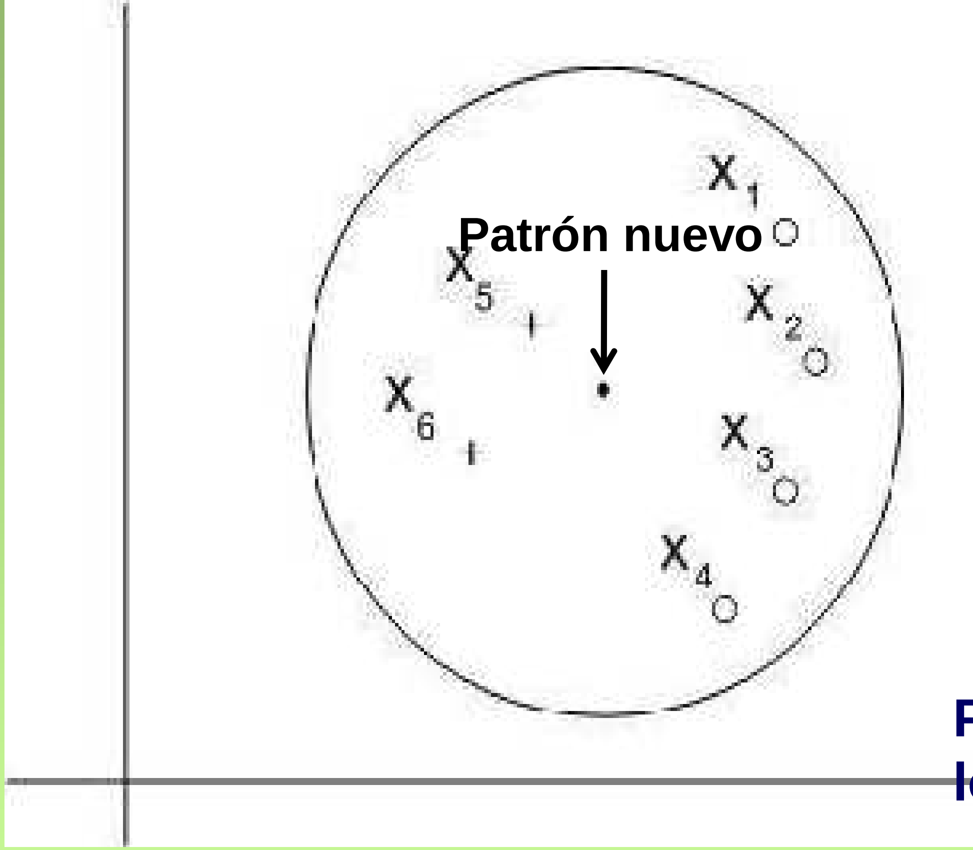
Seleccionar un caso por clase (p.e. el más cercano al centroide de la clase)

- Reducción de la dimensión del fichero almacenado de N a m

- Ejecutar un 1-NN a dicho fichero reducido

- Efectividad condicionada a la homogeneidad dentro de las clases (menor varianza). A mayor homogeneidad más efectivo

K-NN con ponderación de vecinos



	D(xi ,x)	wi
x1	2	½=0,5
x2	2	0,5
x3	2	0,5
x4	2	0,5
x5	0,7	1/0,7
x6	0,8	1/0,8

K=6

Peso a asignar a cada uno de los 6 casos seleccionados

Ejemplo de ilustración del K-NN con ponderación de casos seleccionados

Votos para clase “o”: 0.5 + 0.5 + 0.5 + 0.5 = 2
Votos para clase “+”: 1.43 + 1.25 = 2.68
Predicción “+”

K-NN con ponderación de vecinos

Mismo peso a todas las variables:

$$d(\mathbf{x},\mathbf{x}_{\mathrm{r}}) = \sum_{j=1}^{n}\left(x_j - x_{\mathrm{r}j}\right)^2$$

- **Distinto peso a cada variable:**

$$d(\mathbf{x},\mathbf{x}_{\mathrm{r}}) = \sum_{j=1}^{n} w_j\left(x_j - x_{\mathrm{r}j}\right)^2$$

- **Determinar w_j a partir de la Cantidad de Información mutua entre X_j y C** $I(X_j, C)$

K-NN con ponderación de variables

X1	X2	C
0	0	1
0	0	1
0	0	1
1	0	1
1	0	1
1	1	1
0	1	0
0	1	0
0	1	0
1	1	0
1	1	0
1	0	0

$$I(X_1,C)=\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_{1i},\mathrm{c}_j)\log_2 \frac{p(x_{1i},\mathrm{c}_j)}{p(x_{1,i})\,p(\mathrm{c}_j)}$$

$$I(X_1,C)=P_{(x_1,C)}(0,0)\log_2\frac{P_{(x_1,C)}(0,0)}{P_{x_1}(0)P_C(0)}+P_{(x_1,C)}(0,1)\log_2\frac{P_{(x_1,C)}(0,1)}{P_{x_1}(0)P_C(1)}+$$

$$+P_{(x_1,C)}(1,0)\log_2\frac{P_{(x_1,C)}(1,0)}{P_{x_1}(1)P_C(0)}+P_{(x_1,C)}(1,1)\log_2\frac{P_{(x_1,C)}(1,1)}{P_{x_1}(1)P_C(1)}=$$

$$=\frac{3}{12}\log_2\frac{\frac{3}{12}}{\frac{6}{12}\cdot\frac{6}{12}}+\frac{3}{12}\log_2\frac{\frac{3}{12}}{\frac{6}{12}\cdot\frac{6}{12}}+\frac{3}{12}\log_2\frac{\frac{3}{12}}{\frac{6}{12}\cdot\frac{6}{12}}+\frac{3}{12}\log_2\frac{\frac{3}{12}}{\frac{6}{12}\cdot\frac{6}{12}}=$$

$$=\frac{3}{12}\log_2\frac{12}{12}+\frac{3}{12}\log_2\frac{12}{12}+\frac{3}{12}\log_2\frac{12}{12}+\frac{3}{12}\log_2\frac{12}{12}=0$$

La variable X1 no es relevante para determinación de la clase C

K-NN con ponderación de variables

$$I(X_2,C)=\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_{2i},\mathrm{c}_j)\log_2 \frac{p(x_{2i},\mathrm{c}_j)}{p(x_{2,i})\,p(\mathrm{c}_j)}$$

X1	X2	C
0	0	1
0	0	1
0	0	1
1	0	1
1	0	1
1	1	1
0	1	0
0	1	0
0	1	0
1	1	0
1	1	0
1	0	0

$$I(X_2,C)=P_{(x_2,C)}(0,0)\log_2\frac{P_{(x_2,C)}(0,0)}{P_{x_2}(0)P_C(0)}+P_{(x_2,C)}(0,1)\log_2\frac{P_{(x_2,C)}(0,1)}{P_{x_2}(0)P_C(1)}+$$

$$+P_{(x_2,C)}(1,0)\log_2\frac{P_{(x_2,C)}(1,0)}{P_{x_2}(1)P_C(0)}+P_{(x_2,C)}(1,1)\log_2\frac{P_{(x_2,C)}(1,1)}{P_{x_2}(1)P_C(1)}=$$

$$=\frac{1}{12}\log_2\frac{\frac{1}{12}}{\frac{6}{12}\cdot\frac{6}{12}}+\frac{5}{12}\log_2\frac{\frac{5}{12}}{\frac{6}{12}\cdot\frac{6}{12}}+\frac{5}{12}\log_2\frac{\frac{5}{12}}{\frac{6}{12}\cdot\frac{6}{12}}+\frac{1}{12}\log_2\frac{\frac{1}{12}}{\frac{6}{12}\cdot\frac{6}{12}}=$$

$$=-\frac{1}{6}\log_2 3+\frac{5}{6}\log_2 5-\frac{5}{6}\log_2 3=-\log_2 3+\frac{5}{6}\log_2 5$$

mientras que la variables X2 si lo es

SELECCIÓN DE PROTOTIPOS

EDICIÓN DE WILSON

- Someter a prueba a cada uno de los elementos del fichero de casos inicial.

- Para cada caso se compara su clase verdadera con la que propone un clasificador K-NN obtenido con todos los casos excepto el mismo.

- Si ambas clases no coinciden, el caso es eliminado.

- Edición de Wilson repetitiva parando el procedimiento cuando en 2 selecciones sucesivas no se produzcan cambios.

SELECCIÓN DE PROTOTIPOS

CONDENSACIÓN DE HART

Para cada caso, y siguiendo el orden en el que se encuentran almacenados los casos en el fichero, se construye un clasificador K-NN con tan sólo los casos anteriores al caso en cuestión

• Si el caso tiene un valor de la clase distinto al que le asignaría el clasificador K-NN, el caso es seleccionado

• Si por el contrario la clase verdadera del caso coincide con la propuesta por el clasificador K-NN, el caso no se selecciona.

• El método es dependiente del orden en que se encuentren almacenados los casos en el fichero

Clasificación: K-vecinos más cercanos:K-NN

César Hervás-Martínez
Grupo de Investigación AYRNA

Departamento de Informática y Análisis Numérico
Universidad de Córdoba
Campus de Rabanales. Edificio Einstein.
Email: chervas@uco.es

2019-2020

OneR

OneR: clasificador con reglas que solo tienen una variable en el antecedente.

- **Para atributos discretos (en Weka se discretizan los numéricos).**
- **Se genera un conjunto de reglas del tipo:**
- **Si variable=valor entonces Clase = categoría.**

ALGORITMO OneR
Para cada atributo *a:*
    Para cada valor del atributo *v:*
        Contar cuantas veces aparece cada etiqueta en ejemplos en los que *a=v.*
        Encontrar la etiqueta más frecuente (*cmax).*
        Crear una regla Si *a=v entonces Clase=cmax.*
    Fin Para
    Calcular el error de aplicar todas las reglas de *a.*
Fin Para
Devolver las reglas del atributo *a con menor error.*