

Ejercicios-Resueltos-KNN-KNNI.pdf



user_2335920



Especialidad: Sistemas de Información



4º Grado en Ingeniería Informática



**Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación
Universidad de Granada**



Accede al documento original

antes



**Descarga sin publi
con 1 coin**



Después

WUOLAH



Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato
→ Planes pro: más coins

perdo
espacio



Necesito
concentración

ali ali ooh
esto con 1 coin me
lo quito yo...

WUOLAH



Ejercicio inventado,

1) Del ej anterior, hazmelo ahora con kNN y sacame las medidas para TPR, TNR y G-MEAN.

Ya tenemos de antes los valores de

$$TPR = \frac{TP}{TP+FN} = 2/2$$

$$TNR = \frac{TN}{TN+FP} = 3/4$$

$$\text{Ahora, G-Mean} = \sqrt{TPR \cdot TNR} = \sqrt{1 \cdot 0.75} = 0.866$$

Aplicamos el algoritmo kNN (K=3). Este algoritmo compara las distancias del conjunto de prueba con las del conjunto de entrenamiento y asigna la clase más frecuente entre los k vecinos más cercanos. Usamos una métrica simple: cada atributo tendrá el mismo peso y la distancia será la cantidad de atributos diferentes.

Paso 1: definir la distancia entre instancias

Para cada atributo (color, altura, zagas, textura)

- Si los valores son iguales: distancia = 0
- Si los valores son diferentes: distancia = 1

Distancia total = Suma de las diferencias de los atributos,

o Instancia 11 (Azul, Bajo, Si, Rugoso)

id	color	tamaño	zagas	textura	distancia	clase
1	Púrpura	Alto	Si	Rugoso	2	Venenoza
2	Azul	Alto	No	Peludo	3	Comestible
3	Púrpura	Alto	Si	Suave	3	Comestible
4	Rojos	Bajo	Si	Peludo	2	Comestible
5	Azul	Bajo	No	Suave	2	Comestible
6	Púrpura	Alto	Si	Peludo	3	Venenoza
7	Púrpura	Bajo	No	Peludo	3	Venenoza
8	Azul	Bajo	Si	Peludo	1	Venenoza
9	Rojos	Alto	No	Peludo	4	Comestible
10	Azul	Alto	Si	Suave	2	Venenoza

Vecinos más cercanos ($K=3$)

1. S8, Venenosa, distancia = 1
2. S1, Venenosa, distancia = 2
3. S4, Comestible, distancia = 2

Clase predicha: Venenosa

• Instancia 12 (Rojo, Bajo, No, Suave)

- | | |
|---------------------|----------------------|
| id 1: distancia = 4 | id 6: distancia = 4 |
| id 2: distancia = 3 | id 7: distancia = 2 |
| id 3: distancia = 3 | id 8: distancia = 3 |
| id 4: distancia = 2 | id 9: distancia = 2 |
| id 5: distancia = 4 | id 10: distancia = 3 |

Vecinos más cercanos ($K=3$)

1. S5, Comestible, distancia = 1
2. S4, Comestible, distancia = 2
3. S7, Venenosa, distancia = 2

Clase predicha = Comestible

• Instancia 13 (Azul, Bajo, Si, Suave)

- | | |
|---------------------|----------------------|
| id 1: distancia = 3 | id 6: distancia = 3 |
| id 2: distancia = 3 | id 7: distancia = 3 |
| id 3: distancia = 2 | id 8: distancia = 1 |
| id 4: distancia = 2 | id 9: distancia = 4 |
| id 5: distancia = 1 | id 10: distancia = 1 |

Vecinos más cercanos ($K=3$)

1. S5, Venenosa, distancia = 1
2. S8, Venenosa, distancia = 1
3. S10, Venenosa, distancia = 1

Clase predicha = Venenosa

• Instancia 14 (Rojo, Alto, Sí, Peludo)

id1: distancia = 2
id2: distancia = 2
id3: distancia = 2
id4: distancia = 1
id5: distancia = 4

id6: distancia = 1
id7: distancia = 3
id8: distancia = 2
id9: distancia = 1
id10: distancia = 2

Vecinos más cercanos ($K=3$)

1. S4, Comestible, distancia = 1
2. S6, Venenosa, distancia = 1
3. S9, Comestible, distancia = 1

Clase predicha = Comestible

• Instancia 15 (Púrpura, Alto, No, Rugoso)

id1: distancia = 1
id2: distancia = 2
id3: distancia = 2
id4: distancia = 4
id5: distancia = 3

id6: distancia = 2
id7: distancia = 2
id8: distancia = 4
id9: distancia = 2
id10: distancia = 3

Vecinos más cercanos ($K=3$)

1. S1, Venenosa, distancia = 1
2. S2, Comestible, distancia = 2
3. S3, Comestible, distancia = 2

Clase predicha = Comestible

• Instancia 16 (Púrpura, Alto, No, Suave)

id1: distancia = 2
id2: distancia = 2
id3: distancia = 1
id4: distancia = 4
id5: distancia = 2

id6: distancia = 2
id7: distancia = 2
id8: distancia = 4
id9: distancia = 2
id10: distancia = 2

Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato
→ Planes pro: más coins

perdo
espacio



Necesito
concentración

ali ali ooh
esto con 1 coin me
lo quito yo...

WUOLAH

WUOLAH

Ver Venenos más cercanos ($K=3$)

1. S3, Comestible, distancia = 1 S
2. S1, Venenosa, distancia = 2 S
3. S2, Comestible, distancia = 2 S

Clase predicha: Comestible

Luego, tenemos para Knn ($K=3$)

id	Clase real	Clase predicha
11	Venenosa	Venenosa
12	Comestible	Comestible
13	Venenosa	Venenosa
14	Comestible	Comestible
15	Comestible	Comestible
16	Comestible	Comestible

2) Usar Knn ($K=3$) para este ej:

Conjunto
de entrenamiento

id	Edad	Ingreso	Estado Civil	Clase
1	25	50000	Soltero	No
2	35	80000	Casado	Si
3	45	120000	Soltero	Si
4	20	20000	Casado	No
5	30	70000	Soltero	No
6	40	90000	Casado	Si

Conjunto
de prueba

id	Edad	Ingreso	Estado Civil	Clase
7	28	60000	Casado	
8	39	40000	Casado	

Paso 1. Definir la métrica de distancia para datos mixtos

Distancia total = Distancia numérica + Distancia categórica

1. Distancia numérica: normalizamos los caracteres numéricos para enteros que dominen.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Luego usamos la distancia euclidiana

$$d_{numérica} = \sqrt{(x_1' - y_1')^2 + (x_2' - y_2')^2}$$

2. Distancia categórica: si los valores categóricos son diferentes, asignamos distancia = 1, si no distancia = 0.

Paso 2: Normalizar variables numéricas:

id	Edad	Ingreso	Estado	Clase
1	$\frac{25-20}{45-20} = 0.2$	0.3	Soltero	No
2	$\frac{35-20}{45-20} = 0.6$	0.6	Casado	Si
3	$\frac{45-20}{45-20} = 1$	1	Soltero	Si
4	$\frac{20-20}{45-20} = 0$	0	Casado	No
5	$\frac{30-20}{45-20} = 0.4$	0.5	Soltero	No
6	$\frac{40-20}{45-20} = 0.8$	0.7	Casado	Si

Ya que estamos normalizando los datos numéricos, del conjunto de prueba

id	Edad	Ingreso	Estado	Clase
7	0.32	0.4	Casado	.
8	0.76	0.2	Casado	.

Imagínate aprobando el examen

Necesitas tiempo y concentración

Planes	 PLAN TURBO	 PLAN PRO	 PLAN PRO+
 Descargas sin publi al mes	10 	40 	80 
 Elimina el video entre descargas			
 Descarga carpetas			
 Descarga archivos grandes			
 Visualiza apuntes online sin publi			
 Elimina toda la publi web			
 Precios Anual <input type="checkbox"/>	0,99 € / mes	3,99 € / mes	7,99 € / mes

Ahora que puedes conseguirlo,
¿Qué nota vas a sacar?



WUOLAH

Paso 3. Calcular distancias. Calculamos la distancia de los puntos de prueba con los de entrenamiento.

• id 7

> Con id 1

$$d_{\text{num}} = \sqrt{(0.32 - 0.21)^2 + (0.4 - 0.3)^2} = 0.16$$

$$d_{\text{categorica}} = \text{Soltero} \neq \text{Casado} = 1$$

$$d_{\text{total}} = 1.16$$

> Con id 2

$$d_{\text{num}} = \sqrt{(0.32 - 0.6)^2 + (0.4 - 0.6)^2} = 0.344$$

$$d_{\text{categorica}} = \text{Casado} = \text{Casado} = 0$$

$$d_{\text{total}} = 0.344$$

> Con id 3

$$d_{\text{num}} = \sqrt{(0.32 - 1)^2 + (0.4 - 1)^2} = 0.91$$

$$d_{\text{categorica}} = \text{Soltero} \neq \text{Casado} = 1$$

$$d_{\text{total}} = 1.91$$

> Con id 4

$$d_{\text{num}} = \sqrt{(0.32 - 0)^2 + (0.4 - 0)^2} = 0.51$$

$$d_{\text{categorica}} = \text{Casado} = \text{Casado} = 0$$

$$d_{\text{total}} = 0.51$$

> Con id 5

$$d_{\text{num}} = \sqrt{(0.32 - 0.41)^2 + (0.4 - 0.5)^2} = 0.13$$

$$d_{\text{cat}} = \text{Soltero} \neq \text{Casado} = 1$$

$$d_{\text{total}} = 1.13$$

Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato
→ Planes pro: más coins

perdo
espacio



Necesito
concentración

ali ali ooh
esto con 1 coin me
lo quito yo...

WUOLAH



> Con id 6

$$d_{num} = \sqrt{(0.32-0.8)^2 + (0.4-0.7)^2} = 0.57$$

$$d_{cat} = \text{Casado} = \text{Casado} = 0$$

$$d_{total} = 0.57$$

Ahora, seleccionamos los K vecinos más cercanos.

1. id 2 (0.34, Clase SI)

2. id 4 (0.51, Clase No)

3. id 6 (0.57, Clase SI)

Se predice
la clase SI

• id 8

> Con id 1

$$d_{num} = \sqrt{(0.76-0.2)^2 + (0.2-0.3)^2} = 0.57$$

$$d_{cat} = \text{Soltero} \neq \text{Casado} = 1$$

$$d_{total} = 1.57$$

> Con id 2

$$d_{num} = \sqrt{(0.76-0.6)^2 + (0.2-0.6)^2} = 0.43$$

$$d_{cat} = \text{Casado} = \text{Casado} = 0$$

$$d_{total} = 0.43$$

> Con id 3

$$d_{num} = \sqrt{(0.76-1)^2 + (0.2-1)^2} = 0.84$$

$$d_{cat} = \text{Soltero} \neq \text{Casado} = 1$$

$$d_{total} = 1.84$$

> Con id 4

$$d_{num} = \sqrt{(0.76-0)^2 + (0.2-0)^2} = 0.786$$

$$d_{cat} = \text{Casado} = \text{Casado} = 0$$

$$d_{tot} = 0.786$$

> Con id 5

$$d_{num} = \sqrt{(0.76-0.4)^2 + (0.2-0.5)^2} = 0.47$$

$$d_{cat} = \text{Soltero} \neq \text{Casado} = 1$$

$$d_{total} = 1.47$$

> Con id 6

$$d_{num} = \sqrt{(0.76-0.8)^2 + (0.2-0.7)^2} = 0.5$$

$$d_{cat} = \text{Casado} = \text{Casado} = 0$$

$$d_{total} = 0.5$$

Cogemos los 3 vecinos más cercanos.

• id 2 (0.47, Clase = Si)

• id 6 (0.5, Clase = Si)

• id 4 (0.786, Clase = No)

Clase
predicha Si

3) K-NN (para K=3)

id	Edad	Ingreso	Estado civil	Compra
1	25	50000	Soltero	No
2	35	80000	Casado	Si
3	45	120000	Soltero	Si
4	?	20000	Casado	No
5	30	70000	?	No
6	40	90000	Casado	?

Debemos de imputar

- La edad faltante para la fila 4
- El estado civil de la fila 5
- La compra faltante de la fila 6.

Paso 1: Métrica de distancia para datos mixtos

1. Distancia numérica: usamos la distancia euclidiana después de normalizar las var numéricas.

$$d_{num} = \sqrt{\sum \left(\frac{x-y}{\text{rang}_0} \right)^2}$$

2. Distancia categórica: asignamos distancia 1 si está casado y 0 si está soltero. La clase Si será 1 y clase 0 será 0

La distancia total será: $d_{total} = d_{num} + d_{cat}$

Paso 2: normalizar las var numéricas

id	Edad	Ingreso	Estado	Clase
1	0	0.3	Soltero	No
2	0.5	0.6	Casado	Si
3	1	1	Soltero	Si
4	?	0	Casado	No
5	0.25	0.5	?	No
6	0.75	0.7	Casado	?

Paso 3: imputar valores faltantes

- Imputamos la edad para id 4 a las filas con valores conocidos
- > Distancias (numéricas y categóricas).

Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato
→ Planes pro: más coins

perdo
espacio



Necesito
concentración

ali ali ooh
esto con 1 coin me
lo quito yo...

WUOLAH

WUOLAH

id 1

$$d_{num} = \sqrt{(10-0.3)^2} = 0.3$$

$$d_{cat} = \text{Soltero} \neq \text{Casado} = 1$$

$$d_{tot} = 1.3$$

id 2

$$d_{num} = \sqrt{(10-0.6)^2} = 0.6$$

$$d_{cat} = \text{Casado} = \text{Casado} = 0$$

$$d_{tot} = 0.6$$

id 3

$$d_{num} = \sqrt{(10-1)^2} = 1$$

$$d_{cat} = \text{Soltero} \neq \text{Casado} = 1$$

$$d_{tot} = 2$$

Los 3 vecinos más cercanos son

id 1 (Edad = 25), id 2 (Edad = 35), id 3 (Edad = 45)

Calculamos la edad imputada para id 4 = $\frac{25+35+45}{3} = 35$

• Imputamos el Estado para la fila 5 (id 5)

id 1

$$d_{num} = \sqrt{(10.25-0)^2 + (0.5-0.3)^2} = 0.32$$

$$d_{tot} = 0.32$$

id 2

$$d_{num} = \sqrt{(10.25-0.5)^2 + (0.5-0.6)^2} = 0.54$$

$$d_{tot} = 0.54$$

id 3

$$d_{num} = \sqrt{(10.25-1)^2 + (0.5-1)^2} = 1.09$$

$$d_{tot} = 1$$

Los 3 vecinos cercanos son

id 1 (Estado = Soltero), id 2 (Estado = Casado), id 3 (Estado = Soltero)
Se imputa como Soltero

• Para la clase de id6

id1

$$d_{num} = \sqrt{(0.75 - 0)^2 + (0.7 - 0.3)^2} = 0.85$$

$$d_{cat} = \text{Casado} \neq \text{Soltero} = 1$$

$$d_{tot} = 1.85$$

id2

$$d_{num} = \sqrt{(0.75 - 0.5)^2 + (0.7 - 0.6)^2} = 0.27$$

$$d_{cat} = \text{Casado} = \text{Casado} = 0$$

$$d_{tot} = 0.27$$

id3

$$d_{num} = \sqrt{(0.75 - 1)^2 + (0.7 - 1)^2} = 0.39$$

$$d_{cat} = \text{Soltero} \neq \text{Casado} = 1$$

$$d_{tot} = 1.39$$

Considerando los 3 vecinos más frecuentes:

id1 (Clase = No), id2 (Clase = Si), id3 (Clase = Si)

~~Por lo tanto~~ Imputamos la clase de id6 como Si.

id	Edad	Ingreso	Estado	Clase
1	25	50000	Soltero	No
2	35	80000	Casado	Si
3	45	120000	Soltero	Si
4	35	20000	Casado	No
5	30	70000	Soltero	No
6	40	90000	Casado	Si

45,