

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



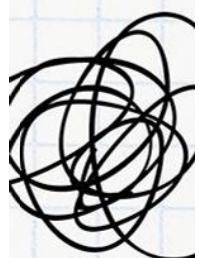
WUOLAH

Importante

Puedo eliminar la publi de este documento con 1 coin

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

pierdo
espacio



Necesito
concentración

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

wuolah

Ejercicio inventado

1) Del ej anterior, hazmelo ahora con k=3 y sacame las medias para TPR, TNR y G-Mean.

Ya tenemos de antemano los valores de

$$TPR = \frac{TP}{TP+FN} = 2/2 \quad TNR = \frac{TN}{TN+FP} = 3/4$$

$$\text{Ahora, } G\text{-Mean} = \sqrt{TPR \cdot TNR} = \sqrt{2 \cdot 0.75} = 0.866$$

Aplicando el algoritmo kNN ($K=3$). Este algoritmo compara las distancias del conjunto de prueba con los del conjunto de entrenamiento y asigna la clase más frecuente entre los k vecinos más cercanos. Usaremos 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

Paso cada atributo (color, altura, rayas, textura)

- Si los valores son iguales: distancia = 0

- Si los valores son diferentes: distancia = 1

Distancia total = Suma de las diferencias de los atributos

• Instancia #1 (Azul, Baja, Sí, Rugoso)

id	color	tamaño	rayas	textura	distancia	close
1	Púrpura	Alto	Sí	Rugoso	2	Venenoosa
2	Azul	Alto	No	Peludo	3	Comestible
3	Púrpura	Alto	Sí	Suave	3	Comestible
4	Rojo	Bajo	Sí	Peludo	2	Comestible
5	Azul	Bajo	No	Suave	2	Comestible
6	Púrpura	Alto	Sí	Peludo	3	Venenoosa
7	Púrpura	Bajo	No	Peludo	3	Venenoosa
8	Azul	Bajo	Sí	Peludo	1	Venenoosa
9	Rojo	Alto	No	Peludo	4	Comestible
10	Azul	Alto	Sí	Suave	2	Venenoosa

Vecinos más cercanos ($K=3$)

1. S8, Venenosa, distancia = 15
2. S1, Venenosa, distancia = 25
3. S4, Comestible, distancia = 25

Clase predicha: Venenosa

• Instancia 12 (Rojo, Bajo, ño, 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 = 15
2. S4, Comestible, distancia = 25
3. S7, Venenosa, distancia = 25

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

Vecino más cercano ($K=3$)

1. S5, Venenosa, distancia = 15
2. S8, Venenosa, distancia = 15
3. S10, Venenosa, distancia = 15

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. S 4, Comestible, distancia = 1
2. S 6, Venenosa, distancia = 1
3. S 9, Comestible, distancia = 1

Clase predicha: Comestible.

• Instancia 15 (Púrpura, Alto, Ro, 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. S 1, Venenosa, distancia = 1
2. S 2, Comestible, distancia = 2
3. S 3, Comestible, distancia = 2

Clase predicha: Comestible.

• Instancia 16 (Púrpura, Alto, Ro, Suave)

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

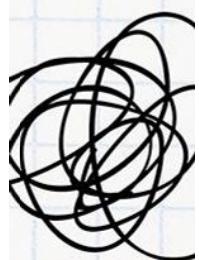
id6: distancia = 2
id7: distancia = 2
id8: distancia = 4
id9: distancia = 2
id10: distancia = 2

Importante

Puedo eliminar la publi de este documento con 1 coin

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

pierdo
espacio



Necesito
concentración

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

wuolah

Veal Veinos más cercanos ($K=3$)

1. S3, Comestible, distancia = 1S

2. S4, Venenosa, distancia = 2S

3. S2, Comestible, distancia = 2S

Planteamiento: 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.

id	Edad	Ingreso	Estado Civil	Clase
1	25	50000	Soltero	do
2	35	80000	Casado	Si
3	45	120000	Soltero	Si
4	20	20000	Casado	do
5	30	70000	Soltero	do
6	40	90000	Casado	Si

Conjunto
de entrenamiento

id	Edad	Ingreso	Estado Civil	Clase
7	28	60000	Casado	do
8	32	40000	Casado	do

Conjunto
de prueba

Paso 1: definir la medida de distancia para datos mixtos.

$$\text{Distancia total} = \text{Distancia numérica} + \text{Distancia categórica}$$

1 Distancia numérica: normalizamos las características numéricas para enteros que deninen.

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

Luego usamos la distancia euclídea

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

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+
diamond Descargas sin publi al mes	10 🟡	40 🟡	80 🟡
clock Elimina el video entre descargas	✓	✓	✓
folder Descarga carpetas	✗	✓	✓
download Descarga archivos grandes	✗	✓	✓
circle Visualiza apuntes online sin publi	✗	✓	✓
glasses 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 muestra con los de entrenamiento.

• id 7

> Con id 1

$$d_{num} = \sqrt{(0.32 - 0.2)^2 + (0.4 - 0.3)^2} = 0.16$$

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

$$d_{total} = 1.16$$

> Con id 2

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

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

$$d_{total} = 0.344$$

> Con id 3

$$d_{num} = \sqrt{(0.32 - 1)^2 + (0.4 - 1)^2} = 0.81$$

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

$$d_{total} = 1.81$$

> Con id 4

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

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

$$d_{total} = 0.32$$

> Con id 5

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

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

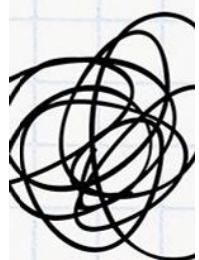
$$d_{total} = 1.13$$

Importante

Puedo eliminar la publi de este documento con 1 coin

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

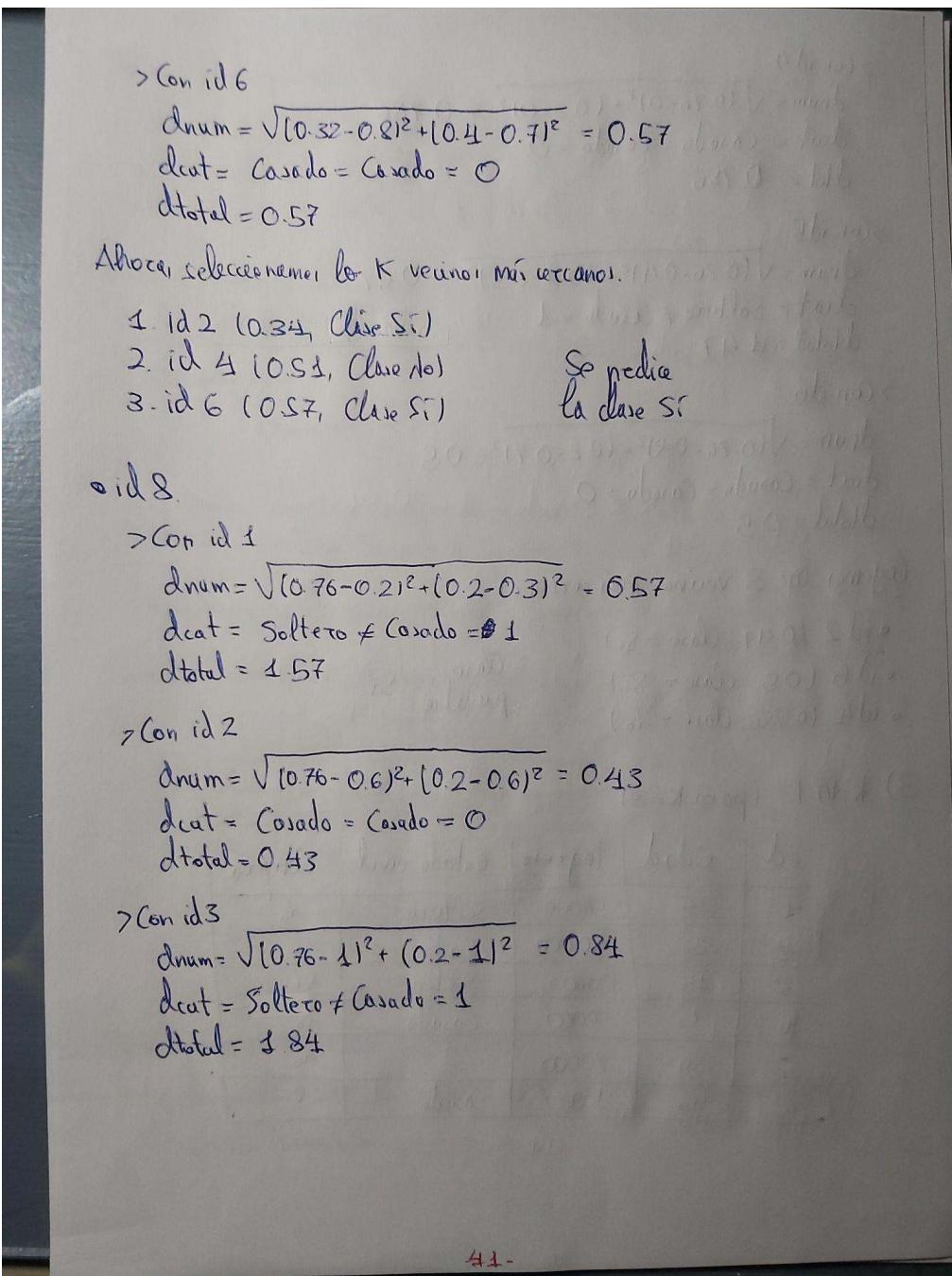
pierdo
espacio



Necesito
concentración

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

wuolah



> Con id 4

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

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

$$d_{total} = 0.786$$

> Con id 5

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

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

$$d_{total} = 1.47$$

> Con id 6

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

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

$$d_{total} = 0.5$$

Cogemos los 3 vecinos más cercanos.

• id 2 (0.47, clase = Sí)

• id 6 (0.5, clase = Sí)

• id 4 (0.786, clase = No)

Clase
predicha Sí.

3) kNN 1 (para K=3)

id	Edad	Ingreso	Estado civil	Compra
1	25	50000	Soltero	No
2	35	80000	Casado	Sí
3	45	120000	Soltero	Sí
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 euclídea después de normalizar los datos numéricos.

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

2. Distancia categórica: asignamos distancia 1 si ésta Casado y 0 si ésta Soltero. La clase Si será 1 y clase No será 0

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

Paso 2: normalizar los vars numéricos

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

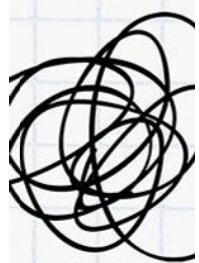
- Imputamos la edad para id 4 a los filas con valores conocidos
- > Distancia (numérica y categórica)

Importante

Puedo eliminar la publi de este documento con 1 coin

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

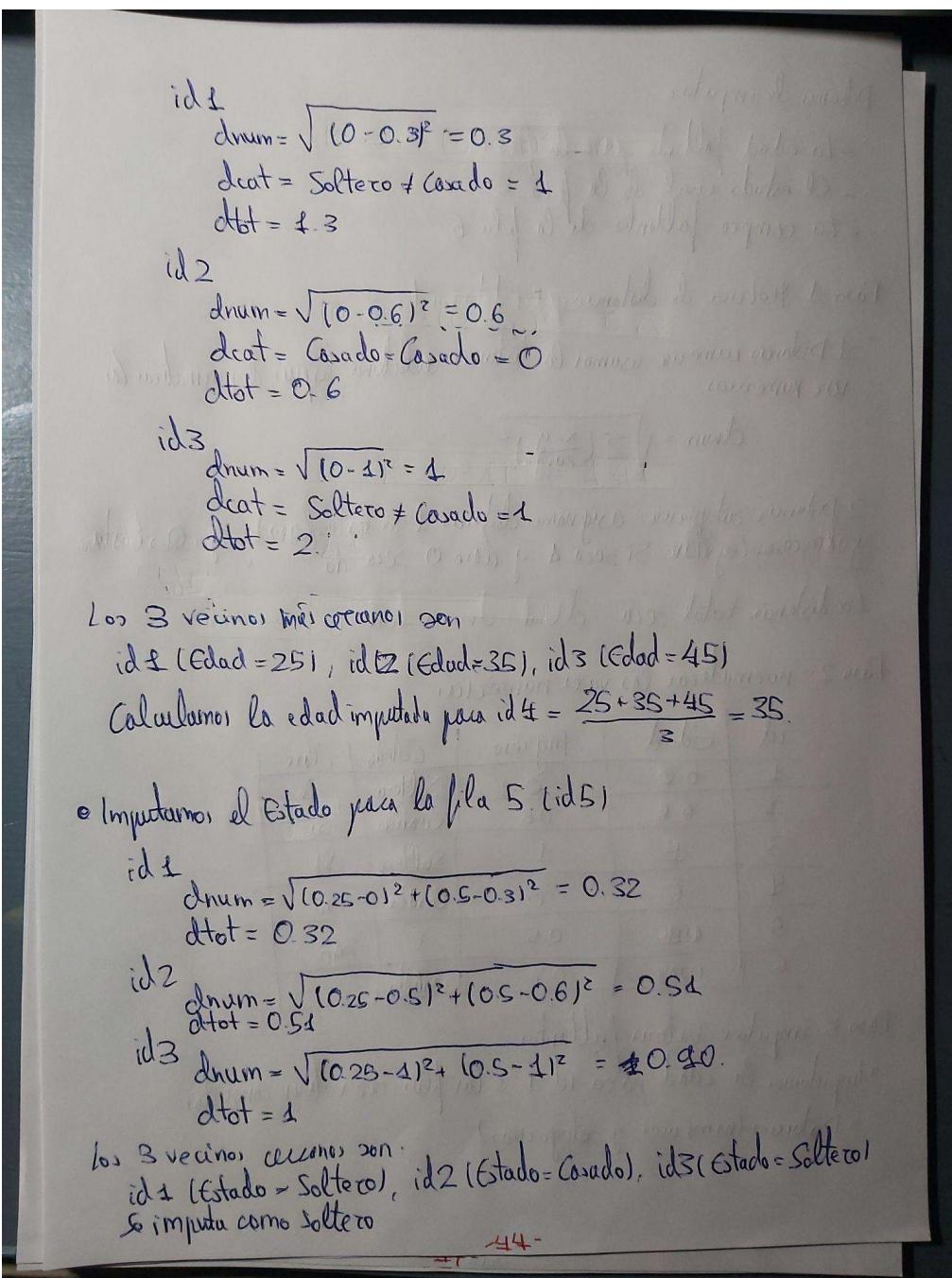
pierdo
espacio



Necesito
concentración

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

wuolah



• Para la clase de id6

id1

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

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

$$d_{tot} = 1.85$$

id2

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

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

$$d_{tot} = 0.27$$

id3

$$d_{num} = \sqrt{(0.75 - 41)^2 + (0.7 - 1)^2} = 0.39$$

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

$$d_{tot} = 4.39$$

Considerando los 3 vecinos más frecuentes:

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

~~Entonces~~ 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,