

K- VECINOS MÁS PRÓXIMOS (K-NN)

APRENDIZAJE AUTOMÁTICO SUPERVISADO



**FACULTAD DE
PSICOLOGÍA**
UNIVERSIDAD COMPLUTENSE DE MADRID

Guillermo de Jorge Botana

Dpto. Psicobiología y Metodología en Ciencias del Comportamiento

Facultad de Psicología.

Universidad Complutense de Madrid.

Tabla de contenido

Introducción	4
Disposición de los datos	4
Aspectos formales de la técnica	6
Distancia entre ejemplares en un espacio	6
Vectores y distancias	8
Procedimiento	10
Fase de preparación del modelo	11
Fase de asignación de nuevos ejemplares	11
Ventajas e Inconvenientes	13
Código de muestra	14

NOTA:

El contenido de este texto corresponde a uno de los temas de una asignatura del Máster de Metodología de las Ciencias del Comportamiento y de la Salud. Está elaborado para tener un texto base de lo que es explicado en clase. Aunque el texto es seguido y coherente, puede ser susceptible de algunas mejoras y ampliaciones. No obstante, es lo suficientemente autocontenido para llevar a cabo un estudio independiente sobre él.

He decidido publicar este texto fuera del ámbito de la asignatura por si puede resultar de utilidad para otros estudiantes o por si a otros docentes les puede facilitar la tarea.

Introducción

La técnica de K- vecinos más próximos (K-NN) es una de las más sencilla en el arsenal de los métodos supervisados. De hecho, se suele emplear como técnica auxiliar en procedimientos de mayor alcance. Se podría resumir el procedimiento diciendo que un nuevo ejemplar sin categoría asignada es comparado con un conjunto de ejemplares que sí tienen categoría. El nuevo ejemplar se asignará a la categoría más repetida entre los ejemplares que más se parecen a él.

Hecho este resumen, en este texto daremos forma a cómo se representan los ejemplares y cómo se calcula su parecido. Por eso merece la pena adentrarse en su formalismo. Porque el concepto de espacio y la métrica de similitud mediante la distancia también se emplean en otras técnicas, por ejemplo, en K-medias. Es más, el concepto de distancia espacial clásica también está sujeta a críticas. Todo esto nos ayudará a reflexionar sobre diferentes temas.

Disposición de los datos

Las técnicas supervisadas se caracterizan por **tener previamente valorados los ejemplares** que forman parte de la muestra con la que se genera el modelo. Cada ejemplar está valorado (con un escalar) o etiquetado (con una categoría o clase) en una propiedad que interesa predecir en un futuro. Una vez que el modelo está generado, la tarea será predecir esa propiedad en ejemplares no vistos.

La muestra para generar el modelo en este tipo de técnicas suele representarse en una tabla en el que **las filas son casos o ejemplares y las columnas características**. Además, una de esas características, habitualmente colocada en la última columna, es la que se ha de predecir en casos novedosos. Para la técnica de K-NN vamos a partir de la misma disposición de los datos. Las propiedades de los ejemplares pueden llamarse también variables. Serán variables independientes o **predictoras** las que sirven para predecir y variable dependiente o **criterio** la que se tendrá que predecir. En la disposición de K-NN, las variables predictoras suelen ser continuas y la variable dependiente categórica (aunque puede también ser continua a modo de

regresión). El hecho de que la variable criterio sea categórica se debe a que lo que se quiere predecir es una categoría o clase.

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.0	2.0	3.5	1.0	versicolor
6.0	2.2	4.0	1.0	versicolor
6.0	2.2	5.0	1.5	virginica
6.2	2.2	4.5	1.5	versicolor
4.5	2.3	1.3	0.3	setosa
5.0	2.3	3.3	1.0	versicolor
5.5	2.3	4.0	1.3	versicolor
6.3	2.3	4.4	1.3	versicolor
4.9	2.4	3.3	1.0	versicolor
5.5	2.4	3.8	1.1	versicolor
5.5	2.4	3.7	1.0	versicolor
4.9	2.5	4.5	1.7	virginica
5.1	2.5	3.0	1.1	versicolor
5.5	2.5	4.0	1.3	versicolor
...
...
...

Figura 1.

La [figura 1](#) muestra la disposición de los datos en el ya famoso conjunto de las especies de flores. A partir de un conjunto de variables predictoras, ancho del pétalo, largo del pétalo, ancho del sépalo y largo del sépalo se desea predecir a cuál de las tres especies pertenece: Versicolor, Virginica, Setosa. Dicho con genericidad; de un conjunto de predictores continuos se predecirá una clase.

En K-NN **no se lleva a cabo estrictamente un entrenamiento**, sino que el modelo es prácticamente la tabla sin apenas elaboración. Como su nombre ya nos indica -vecinos más próximos- lo que se hará es asumir que cualquier ejemplar puede o no ser vecino de otro. Los ejemplares se agrupan en vecindarios según sean o no parecidos. Y para que dos ejemplares sean vecinos mandatoriamente hay que aludir a la **noción de espacio**. La propiedad de ser vecino entraña un espacio sobre el que constatarlo. El espacio como referencia de la similitud. Tal es la homologación.

Para llevar a cabo esa homologación cada ejemplar de esa tabla original (como la de la [figura 1](#)), representado en una fila, se representa como **un punto en un espacio**. La posición en ese

espacio está determinada por las propiedades de cada ejemplar, es decir, por los valores en las variables predictoras. Y asumiendo esto, podemos considerar **las distancias entre ejemplares como medida de similitud entre ellos**. Entonces la cosa será sencilla. Si queremos clasificar un nuevo ejemplar del que no sabemos la clase, colocaremos también el ejemplar en una ubicación en ese espacio de referencia y calcularemos las distancias con todos los ejemplares de la tabla (también ubicados en ese espacio). Se asume que los que más cerca estén serán también los que tengan la misma clase del ejemplar novedoso. Por tanto, la clase mayoritaria de esos ejemplares más cercanos (sus vecinos) será también asignada a este nuevo ejemplar.

Pero para hacer operativo este sencillo mecanismo tenemos que estudiar las formas en que se formalizan los ejemplares como un punto en un espacio y cómo se calculan las distancias (y se puede predecir de un ejemplar que es vecino). La homologación de los ejemplares a puntos o coordenadas y el cálculo de las distancias nos darán la llave para hacerlo.

Aspectos formales de la técnica

Distancia entre ejemplares en un espacio

El hilo conductor desplegado hasta ahora ha sido bastante instigador. Nos ha llevado inequívocamente a la **necesidad de formalizar un espacio y unas métricas de distancia**. No habrá otra forma de identificar vecinos. Bajo la asunción de un espacio, ser vecino es ser próximo. No hay más. Pudiéndose representar las características de un ejemplar en un vector, es decir, con los números de las variables predictoras de cada ejemplar, la asunción espacial consistirá en:

- 1) **cada ejemplar es ubicado en una coordenada** en base a las componentes del vector. El vector identificará al ejemplar en una ubicación en un espacio.
- 2) En ese espacio se acepta **una métrica de distancia como medida de similitud** entre ejemplares.

Pongamos algunos formalismos que determinarán la métrica de la distancia.

Siendo \mathbf{U} una matriz cuyas filas son los vectores con las n variables predictoras de cada ejemplar, dicha matriz podrá ser expresada de esta forma:

$$\mathbf{U} \in \mathbf{R}^n \quad (\text{Fórmula 1})$$

Podemos estar hablando entonces de un espacio-fila cuyas coordenadas son vectores de n componentes. Para que una función cumpla ser una medida de distancia en ese espacio ha de cumplir que para cualquier par de vectores fila $\mathbf{x}, \mathbf{y} \in \mathbf{U}$:

$$f(\mathbf{x}, \mathbf{x}) = D_0 \quad (\text{Fórmula 2})$$

$$f(\mathbf{x}, \mathbf{y}) \geq D_0 \quad (\text{Fórmula 3})$$

$$f(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}, \mathbf{x}) \quad (\text{Fórmula 4})$$

Donde D_0 es un número real.

La primera y segunda propiedad hacen alusión a que la medida de distancia tiene un rango cuyo mínimo se da en la situación en que los vectores son los mismos (los mismos ejemplares). La segunda propiedad se refiere a que, si los ejemplares no son los mismos, la distancia habrá de ser mayor a ese mínimo. La tercera propiedad es la **simetría** que dirá que sea cual sea el orden de los argumentos, la distancia será igual. La misma distancia hay de Segovia a Madrid que de Madrid a Segovia. Además, la medida de la distancia también tiene que cumplir la llamada **desigualdad triangular**, que se expresa de esta forma:

$$f(\mathbf{x}, \mathbf{z}) \leq f(\mathbf{x}, \mathbf{y}) + f(\mathbf{y}, \mathbf{z}) \quad (\text{Fórmula 5})$$

Esto tiene mucho sentido si observamos un espacio con los tres puntos $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$:

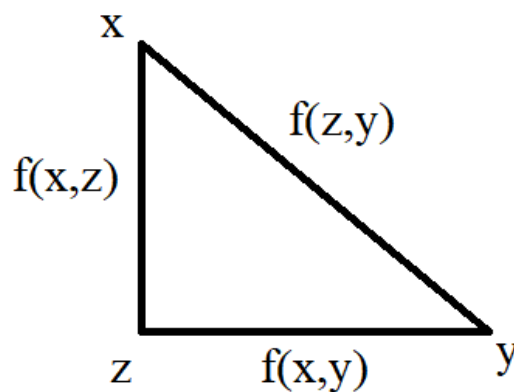


Figura 2

Si nos movemos en un espacio con esas tres coordenadas $\{x, y, z\}$, es mandatorio que un solo lado del triángulo sea menor que la suma de los dos restantes (figura 2). ¡a la fuerza! Un espacio impone a la medida de distancia ésta y las tres propiedades anteriores. Si decimos estar jugando en un espacio euclídeo con puntos y distancias, como es el caso al usar K-NN, éstas serán las reglas tanto de coordenadas como de la admisión de la métrica de la distancia. En ese sentido K-NN comparte estas asunciones con la técnica de K-medias.

Teniendo ya la métrica de distancia admitida, es razonable pensar que, si las ubicaciones de los ejemplares se deben a sus características, la lejanía o cercanía de ejemplares **es una medida de similitud entre ellos**. Los ejemplares vecinos serán similares, y, por ende, se podrá inferir a unos la propiedad ausente de sus vecinos. Esta es la clave de la técnica de K- vecinos más próximos (K-NN).

Vectores y distancias

Hemos mostrado en el apartado anterior algunos formalismos sobre el significado de asumir la similitud como una distancia en un espacio vectorial. Nos hemos entretenido en explicitar las propiedades que debe cumplir una medida de distancia que operase en un espacio. Sin embargo, no queremos dejarlo ahí. Queremos aterrizar esos conceptos y hacerlos operativos. Para ello vamos a expandir el concepto de coordenada a partir de las componentes de un vector y vamos a instanciar la medida de distancia en una función concreta.

Si volvemos a la figura 1, podemos constatar que la parte de las filas de la tabla que corresponde a las variables predictoras son vectores que representan a cada ejemplar. Si existen n variables predictoras los vectores tendrán n dimensiones o componentes. Por ejemplo, el vector del ejemplar 3 de la tabla sería:

$$\mathbf{v}_3 = (6, 2.2, 5, 1.5,)$$

Y el vector del ejemplar 5 sería:

$$\mathbf{v}_5 = (4.5, 2.3, 1.3, 0.3)$$

Es reseñable que el espacio vectorial en el que serían expresados estos ejemplares cumpliría que $\mathbf{U} \in \mathbf{R}^4$, siendo \mathbf{U} la matriz cuyas filas son todos los vectores de los ejemplares. Por tanto, $\mathbf{v}_3, \mathbf{v}_5 \in \mathbf{U}$. Sin embargo, las coordenadas de los ejemplares de este espacio no serían representables en un gráfico. Trascienden al plano y al volumen. Sin embargo, podemos operar con las mismas funciones que si el espacio fuera plano o volumen. No es inconveniente. De

hecho, podemos computar un tipo de distancia que cumple las propiedades anteriores. No es la única métrica de distancia, pero es la más extendida y didáctica. Se trata de la Distancia Euclídea. Esta medida se define por ser **la longitud de la resta de dos vectores**.

La [figura 3](#) muestra el concepto de la distancia euclídea. Si tenemos dos coordenadas identificadas por el vector \mathbf{v}_3 y el vector \mathbf{v}_5 , la distancia será la longitud del vector resta sin importar el orden. Será la longitud del nuevo vector $(\mathbf{v}_3 - \mathbf{v}_5)$.

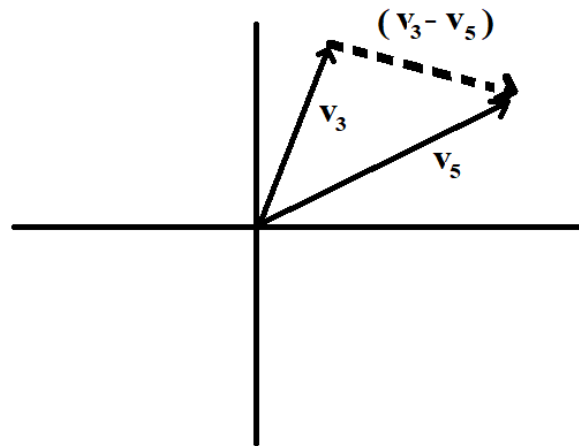


Figura 3

¿Y cómo se calculan la longitud de un vector? O, dicho de otra forma, ¿cómo se calcularía la longitud de ese vector resta? Pues por medio de la norma. La norma de un vector \mathbf{u} se define con la siguiente fórmula:

$$|\mathbf{u}| = \sqrt{u_1^2 + u_2^2 + \dots + u_n^2} \quad (\text{Fórmula 6})$$

La norma es la raíz cuadrada de la suma de los cuadrados de las componentes del vector \mathbf{u} . Es fácil ver que de la operación resulta un escalar, es decir, un solo número, una longitud. Por lo tanto, la longitud de un vector resta $(\mathbf{u} - \mathbf{v})$ se puede expresar como:

$$|\mathbf{u} - \mathbf{v}| = \sqrt{\sum_{i=1}^n (u_i - v_i)^2} \quad (\text{Fórmula 7})$$

Y tomando los dos ejemplares mencionados anteriormente:

$$\text{Distancia}(\mathbf{v}_3, \mathbf{v}_5) = |\mathbf{v}_3 - \mathbf{v}_5| = \sqrt{\sum_{i=1}^n (v_{3i} - v_{5i})^2} \text{ (Fórmula 8)}$$

Donde n son las dimensiones de los vectores, es decir, $n = 4$. Tomando los valores de los vectores y expandiendo el anterior sumatorio tendríamos la distancia entre el ejemplar 3 y el 5.

$$\text{Distancia}(\mathbf{v}_3, \mathbf{v}_5) = \sqrt{(6 - 4,5)^2 + (2,2 - 2,3)^2 + (5 - 1,3)^2 + (1,5 - 0,3)^2}$$

$$\text{Distancia}(\mathbf{v}_3, \mathbf{v}_5) = \sqrt{2,25 + 0,01 + 13,69 + 1,44} = 4,17$$

El ejemplar 3 y 5 están a una distancia de 4,17. Como se ve, este valor es arbitrario y no tiene una escala reconocible. Bien es verdad que tampoco lo necesitamos, porque nos vale con la propia prelación de los ejemplares. Nos vale simplemente con saber cuáles tienen menor distancia con el vector del ejemplar novedoso, es decir, cuáles son sus primeros k vecinos más próximos.

Tener la medida de distancia en el espacio es tener ya la forma de operar con el algoritmo de K-NN. En el próximo apartado se explica ya el procedimiento concreto para asignar una clase a ejemplares novedosos.

Procedimiento

Ya anticipamos que el estudio de la técnica de K-NN iba a llevarnos poco. Ya contamos con parte del trabajo hecho al haber estudiado el concepto de espacio y sobre todo de distancia.

Para definir a un ejemplar como vecino hemos tenido que aludir a ello. En este apartado seremos ya exhaustivos y describiremos cómo funciona el algoritmo paso a paso.

Describiremos dos fases Una primera fase es de preparación de los datos y la otra de asignación de ejemplares. Pasemos a ver ambas:

Fase de preparación del modelo

Antes dijimos que en la técnica de K-NN no hay fase de entrenamiento propiamente dicha. Lo que hay es una fase en que se acopian los ejemplares etiquetados. Esto hace que en ocasiones se llame a esta técnica como de "**aprendizaje perezoso**" (al no haber propiamente aprendizaje) o de **aprendizaje basado en instancias**, pues el aprendizaje viene ya dado por las instancias existentes en memoria y sin ninguna elaboración. Se trata de comparar lo nuevo con lo residente en la memoria.

Pues bien, esta fase es justo la que prepara y formatea estas instancias en memoria. En ella se **preparan los datos de forma que los ejemplares de la muestra etiquetada** estén en disposición de ser comparados con ejemplares novedosos y sin etiquetar (o etiquetados, pero formando parte de la muestra de prueba). Ya sabemos lo que significa esto expresado en un nivel mayor de abstracción. **Cada ejemplar será una coordenada** identificada con el vector formado por sus valores en las variables predictoras. Por tanto, nuestro punto de partida será un espacio-fila (dónde estarán ubicados los vectores que representan a los ejemplares). A su vez, cada vector fila estará indizado por un índice y estará etiquetado con una clase. En otras palabras, tendremos ternas <índice, vector, clase>. Insistimos en que se consideren los vectores de los ejemplares como **puntos ubicados en un espacio de n dimensiones**. El algoritmo de la fase de asignación hará uso de esta abstracción.

Fase de asignación de nuevos ejemplares

Teniendo ya el espacio y los ejemplares de la muestra inicial insertos en él, podemos operar sobre él para asignar una clase un nuevo ejemplar. Simplemente tendremos que calcular la distancia de ese nuevo ejemplar con los ya ubicados en el espacio y asignar la clase mayoritaria de los más próximos. Los pasos que se siguen son cuatro:

1. **Se calcula la distancia entre el vector del nuevo ejemplar y cada uno de los vectores de los ejemplares** en la tabla mencionada en la fase de preparación. El procedimiento para el cómputo de la distancia es el mismo que el explicado antes: por medio de la distancia euclídea.
2. Lo calculado en el punto uno hará que cada uno de **los ejemplares tenidos en cuenta en la fase de preparación (los etiquetados de la tabla)** reciban un valor de

distancia con el ejemplar novedoso. Cada ejemplar tendrá una distancia con el novedoso.

3. **Se seleccionarán los k primeros vecinos**, los cuales serán los ejemplares con menor valor de distancia. Estos formarán el vecindario.
4. Se asignará el nuevo ejemplar a **la clase que más se repita** en esos k primeros vecinos.

En la [figura 4](#) se muestra visualmente el mecanismo. Tenemos ubicados en el espacio ejemplares previamente etiquetados en dos clases. Los cuadrados y triángulos nos sirven para representar esas dos clases. Se quiere clasificar un nuevo ejemplar representado en un círculo con una interrogación. El punto 1 del procedimiento es calcular las distancias de ese nuevo ejemplar a cada uno de los ya etiquetados. La propia cercanía de los ejemplares en el gráfico nos da cuenta intuitiva de las distancias. Con los cálculos del anterior punto, en el punto 2, cada ejemplar etiquetado tendría un valor de distancia. En el punto 3 se seleccionarían los k primeros vecinos. La línea que rodea de forma continua muestra esta selección con un $k = 3$. La línea que rodea de forma continua muestra un criterio de un $k = 5$. Si el criterio es el primero, el de $k = 3$, la clase mayoritaria de los vecinos de este nuevo ejemplar es el triángulo, y, por lo tanto, el nuevo ejemplar sería etiquetado como un triángulo. Ese sería ya el punto 4. Si fuese $k = 5$, la clase mayoritaria sería el cuadrado, y, por tanto, al nuevo se le asignaría cuadrado.

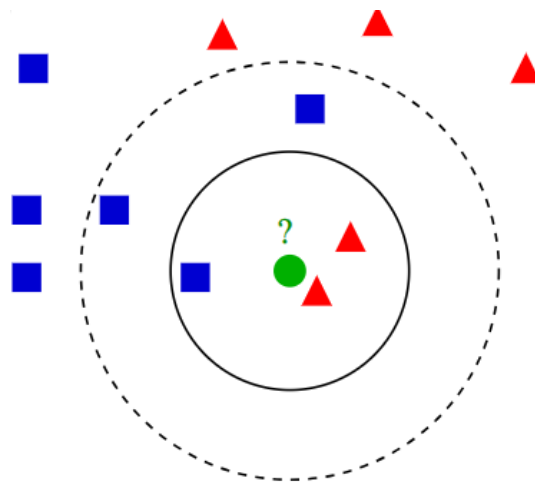


Figura 4. Tomada de https://es.wikipedia.org/wiki/K_vecinos_m%C3%A1s_pr%C3%B3ximos

Ventajas e Inconvenientes

La ventaja principal de esta técnica es su **facilidad** para implementarse. Únicamente se ha de contar con **una masa crítica de ejemplares en la muestra etiquetada** pero mucho menor que por ejemplo lo necesitado por un modelo de redes neuronales. Cuando no se dispone de mucha muestra, podemos empezar desplegando un sistema basado en K-NN hasta que tengamos suficiente muestra para un modelo más sofisticado. O incluso se pueden ir ampliando la muestra para K-NN en paralelo. Aumentar significa también afinar en la clasificación. Además, solo tiene el parámetro k , que es fácilmente optimizable de manera empírica.

Otra ventaja es que se puede hacer uso de esta técnica **después de haber aplicado una técnica de reducción de dimensionalidad**. Vectores con variables que en un principio eran observables pueden ser convertidos en vectores con variables latentes o factores que resumen de una manera efectiva la información. La ubicación de los ejemplares en el espacio sería identificada por estos nuevos vectores y la operativa sería la misma.

Existe también la posibilidad de que **los ejemplares estén representados por vectores con codificación muy sofisticada**. Este es el caso de las llamadas incrustaciones contextuales en los modelos de redes neuronales RNN-LSTM y Transformers. Si los ejemplares son frases, en una fase inicial se transforman esas frases en vectores contextuales gracias a esos modelos. Una vez se tienen esos nuevos vectores, se clasifican nuevos ejemplares en base a la distancia con ellos. Esto se haría en el supuesto de contar con poca muestra etiquetada. Si no fuese el caso, los propios modelos de redes podrían clasificar con más efectividad.

No obstante, algunos son los inconvenientes de esta técnica. Vamos a pasar a enumerarlos:

1. El algoritmo es bastante **inflexible** y depende mucho de las características de la muestra etiquetada.
2. Las **características irrelevantes a la comparación también forman parte del cálculo** de la distancia, por tanto, penalizan la decisión de los k primeros vecinos. No se formaliza un contexto en el momento del cómputo de la distancia.
3. El **coste computacional** de las comparaciones es a veces poco escalable.

K-NN se ha usado tradicionalmente en los sistemas de **recomendación** bajo la consigna de "las cosas que le gustan a mis vecinos me gustan a mí". También en la extracción de patrones parecidos y su subsiguiente asignación. No obstante, con la irrupción de nuevas técnicas de aprendizaje profundo y los inconvenientes que presenta la propia técnica de K-NN, su uso ha derivado en funciones auxiliares insertas en procedimientos de mayor alcance. Por eso es muy útil saber de su existencia y comprenderla. También es habitual verla implementada como complemento a modo de navaja suiza para resolver algún problema puntual que no tenga aspectos críticos.

Existe también una variación del algoritmo de K-NN en el que **la variable a predecir es continua**. Por ejemplo, cuando se trata de predecir el precio de una casa en base a sus características. En esta tarea se puede asignar un promedio del precio de sus vecinos. En el código siguiente se muestra justo un ejemplo de K-NN con esa tarea de regresión.

Código de muestra

El siguiente código aplica K-NN para llevar a cabo una regresión. Se trata de predecir la edad de un abalón (un molusco marino) a partir de un conjunto de propiedades valoradas previamente. La edad se mide a partir del número de anillos de su concha. La tabla tiene el siguiente aspecto:

Ejemplar	Sexo	largo	Diámetro	Ancho	Peso total	Peso carnosos	Peso de víscera	Peso de concha	Anillos(edad)
0	M	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.150	15
1	M	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070	7
2	F	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9
3	M	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.155	10
4	I	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055	7

Los valores de las variables predictoras (excluida la categórica sexo) serán los vectores para ubicar los ejemplares en el espacio. Este espacio será de 7 dimensiones. El código a ejecutar está tomado de <https://realpython.com/knn-python/> y comentado posteriormente. Se muestra en el recuadro siguiente:

```

#https://realpython.com/knn-python/
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

#la URL de los datos de la especie abalón
url = (
    "https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data"
)
#está en csv, así que se lee de ese formato
abalone = pd.read_csv(url, header=None)
#comprobamos que se ha cargado viendo los primeros registros
print(abalone.head())
#le ponemos nombre a las columnas
abalone.columns = [
    "Sex",
    "Length",
    "Diameter",
    "Height",
    "Whole weight",
    "Shucked weight",
    "Viscera weight",
    "Shell weight",
    "Rings",
]
#quitamos el sexo por no ser de interés para la regresión
abalone = abalone.drop("Sex", axis=1)
#los vectores en el espacio serán solo las variables
#independientes. Será la matriz X. Por eso se quita la variable criterio.
X = abalone.drop("Rings", axis=1)
X = X.values
#También en el vector y se colocan los valores de la variable a predecir
y = abalone["Rings"]
y = y.values
#Se instancia un nuevo ejemplar del que se ha de estimar el anillo
new_data_point = np.array([
    0.569552,
    0.446407,
    0.154437,
    1.016849,
    0.439051,
    0.222526,
    0.291208,
])

```

```
#se calculan las distancias entre el nuevo y todos los vectores de la matriz
#esto se hace calculando la norma a dichas restas
distances = np.linalg.norm(X - new_data_point, axis=1)
print(distances)

#calculadas las distancias, se ha de coger un vecindario de 3
#para ello se ordenan las distancias y se toman los índices de las distancias más cortas
k = 3
nearest_neighbor_ids = distances.argsort()[:k]
print(nearest_neighbor_ids)

#ahora tomamos los valores en y de esos índices. y contiene los valores en "rings"
nearest_neighbor_rings = y[nearest_neighbor_ids]
print(nearest_neighbor_rings)

#la predicción será el promedio de los valores de "rings" de esos vecinos.
prediction = nearest_neighbor_rings.mean()
print(prediction)
```