



Hazte cliente de BBVA y... ahórrate 6 meses de suscripción

WUOLAH
+ BBVA

NETFLIX

Spotify

HBO max

Disney+

PlayStation Plus

DAZN

Ahora, si te abres una Cuenta Online en BBVA, te reembolsamos una de estas suscripciones durante 6 meses (hasta 9,99€/mes) al pagarla con tu tarjeta Aqua Débito

Promoción solo para nuevos clientes de BBVA. Válida hasta el 30/06/2023.
Estas empresas no colaboran en la promoción.

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

Aprendizaje no-supervisado

Aprendizaje supervisado: Hay que dar al sistema ejemplos ya clasificados. El objetivo del sistema es descubrir reglas de clasificación.

A la hora de representar los individuos tendremos un número de individuos n descrito por un conjunto de variables m . Los datos se representan en forma de matriz $n \times m$ donde las filas son los individuos y las columnas las variables. Las variables m son las dimensiones en las que representamos los individuos. Suele ser de ayuda poder inspeccionar los datos visualmente, para ello se utiliza el diagrama de dispersión (scatter plot). Los algoritmos de aprendizaje automático son sensibles a la forma en que los individuos están representados, por eso es importante representar los individuos y sus variables adecuadamente.

Aprendizaje no supervisado: No se le da ninguna información al sistema, tiene que descubrir patrones en el conjunto de entrenamiento que permita agrupar unos ejemplos de otros.

El objetivo del aprendizaje no supervisado es encontrar estructura en los datos proporcionados sin atender a ninguna categoría prefijada. Se busca estructura en los individuos (agrupamiento/ clustering) o en las variables (reducción de la dimensionalidad). La estructura nos permite ganar comprensión sobre los datos.

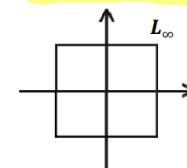
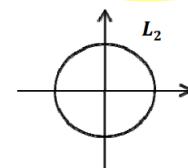
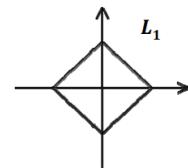
Las técnicas de reducción de la dimensionalidad se aplican cuando contamos con muchas variables en un problema. Se parte de un conjunto de variables m que se busca reducirlo a un conjunto p mucho menor de factores que conservan la máxima información inicial. Se obtienen como una combinación de las variables originales.

El objetivo de las técnicas de agrupamiento o clustering es agrupar los n individuos de nuestro conjunto de datos en una serie de grupos de forma que los individuos del mismo grupo sean lo más parecidos entre sí y de forma que los individuos de grupos diferentes sean lo más diferentes entre sí. De esta forma los grupos revelan cierta estructura de los individuos de nuestro conjunto de datos. Existen dos grandes grupos de algoritmos de agrupamiento o clustering:

Algoritmos de clustering jerárquico:

Cada individuo empieza siendo un cluster \rightarrow hay n clusters. Se repite el algoritmo hasta que todos los individuos formen un único cluster, agrupando los clusters más próximos en un único cluster. Necesitamos definir la distancia que se usa para medir la proximidad.

$$L_1(A, B) = \sum_{i=1}^m |x_{Ai} - x_{Bi}| \quad L_2(A, B) = (\sum_{i=1}^m (x_{Ai} - x_{Bi})^2)^{1/2} \quad L_\infty(A, B) = \max_i |x_{Ai} - x_{Bi}|$$



En estas imágenes se muestran los puntos que se encuentran a la misma distancia en un espacio bidimensional con cada una de estas distancias

- La importancia que tienen las (o la) variables con mayor distancia aumenta en $L_1 < L_2 < L_\infty$

- En L_1 damos igual valor a todas las diferencias, en L_2 el cuadrado hace que pesen más las diferencias grandes y en L_∞ solamente se tiene en cuenta la variable donde la diferencia es mayor

WUOLAH
+ BBVA

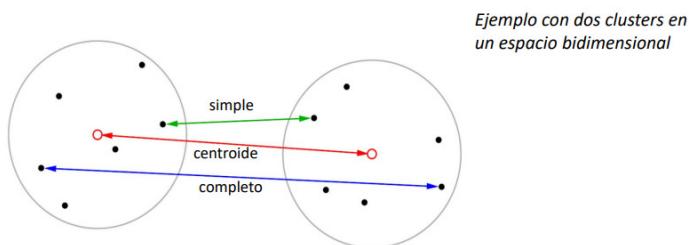


WUOLAH

Centroide: Se toma la distancia entre los puntos medios (el vector medio) de los dos *clusters*

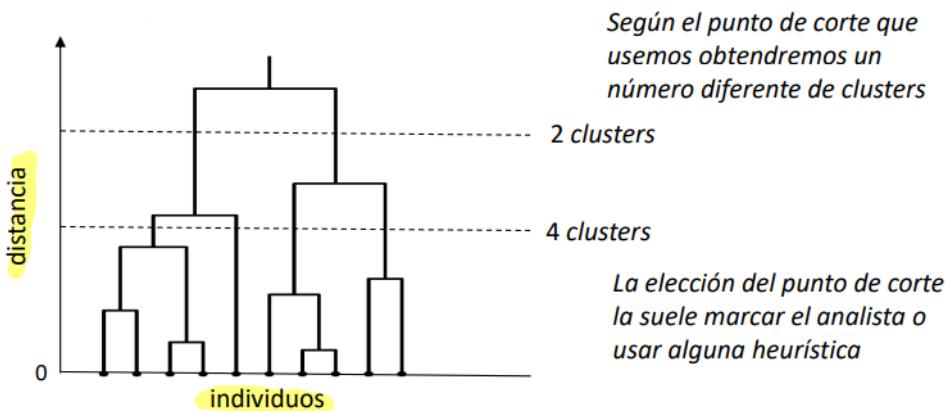
Enlace simple (*single linkage*): Se toma la distancia entre los puntos más próximos de los dos *clusters*

Enlace completo (*complete linkage*): Se toma la distancia entre los puntos más alejados de los dos *clusters*

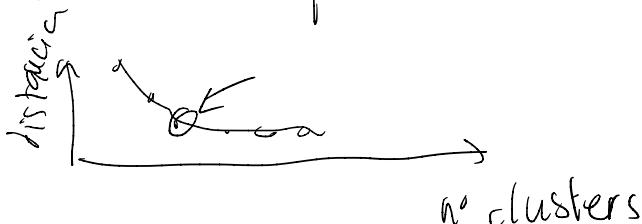


1. **Crear la matriz de distancias D.**
2. **Agrupar individuos.**
 - a. Partición inicial -> **cada individuo es un cluster.**
 - b. **Calcular la siguiente partición usando la matriz D.**
 - i. **Elegir los dos clusters más cercanos y agrupar los dos en un único cluster. Actualizar la matriz D.**
 - c. **Repetir paso b hasta tener un único cluster con todos los individuos. Representar el dendograma.**

El dendrograma es una representación bidimensional de la jerarquía inferida por el algoritmo de clustering jerárquico. En un eje ponemos los individuos y los clusters, el otro eje representa la distancia.



Para el corte podemos usar diagrama de codo



Algoritmos de clustering basados en particiones:

Tienen como objetivo dividir los n individuos en un número de clusters k . Divide el espacio de representación m dimensional en k regiones. El algoritmo de k -medias es el más común y funciona de la siguiente manera:

1. Inicializar centroides a puntos aleatorios del espacio.
2. Asignar cada individuo al centroide más cercano
3. Actualizar la posición de los centroides al valor medio de las posiciones de los individuos asignados.

Se cicla hasta que la posición de los centroides no cambia.

A la hora de decidir el número de clusters k , que vamos a utilizar podemos ir probando diferentes valores pero existen funciones para valorar la dispersión de los centroides.

$$\text{Índice Dunn} = \frac{\min(\text{distancia interclúster})}{\max(\text{distancia intraclúster})}$$

Cuanto mayor sea el número de clusters k , mejores serán los valores de Dunn y Davies-Bouldin. Para saber exactamente cuál es el número ideal se utiliza el diagrama del codo.

Aprendizaje no supervisado

Possible ejercicios

- Construir un dendrograma usando un tipo de distancia
- Dan dendogramas, decidir en cuantos cluster separar usando diagrama del codo



Hazte cliente de BBVA y... ahórrate 6 meses de suscripción

WUOLAH
+ BBVA

NETFLIX

Spotify

HBO max



PlayStation Plus

DAZN

Ahora, si te abres una Cuenta Online en BBVA, te reembolsamos una de estas suscripciones durante 6 meses (hasta 9,99€/mes) al pagarla con tu tarjeta Aqua Débito

Promoción solo para nuevos clientes de BBVA. Válida hasta el 30/06/2023. Estas empresas no colaboran en la promoción.

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

Aprendizaje supervisado

Contamos con una matriz de datos de n individuos, seguimos distinguiendo entre dos tipos de variables. Un conjunto de m variables de entrada x , y una variable de salida y . Si la variable de salida se trata de una **numérica** es un **problema de regresión**. Si se trata de una **variable categórica** es un **problema de clasificación**. El objetivo es ajustar un modelo que **relacione las variables de entrada con la de salida**, de forma que podamos predecir la respuesta lo mejor posible y entendamos la relación entre variables de entrada y salida.

Debemos comparar los resultados pronosticados por nuestro modelo con los resultados obtenidos. En **problemas de regresión** se suele utilizar el **Error Cuadrático Medio** o su raíz cuadrada.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Como medida de error a optimizar/reducir, el RMSE es equivalente al MSE, pero el RMSE es más inteligible ya que está en las mismas unidades de medida que la variable de salida

En **problemas de clasificación** podemos visualizar el error mediante la **matriz de confusión**, que para el caso de una clase binaria sería:

		Clase observada	
		1	0
Clase pronosticada	1	Verdaderos Positivos	Falsos Positivos
	0	Falsos Negativos	Verdaderos Negativos

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN}$$

Necesitamos medidas que reflejan otros aspectos de la clasificación (por clase):

- Tasa de Verdaderos Positivos (Recall/exhaustividad) = $VP / VP + FN$
- Tasa de Verdaderos Negativos (specificity) = $VN / VN + FP$
- Valor Predictivo Positivo (precisión) = $VP / VP + FP$

Podemos obtener clasificadores de muy distinto comportamiento, las medidas de rendimiento nos dicen cómo funcionan. En la fase de aprendizaje, podemos indicar qué medida queremos que optimice el clasificador en su aprendizaje.

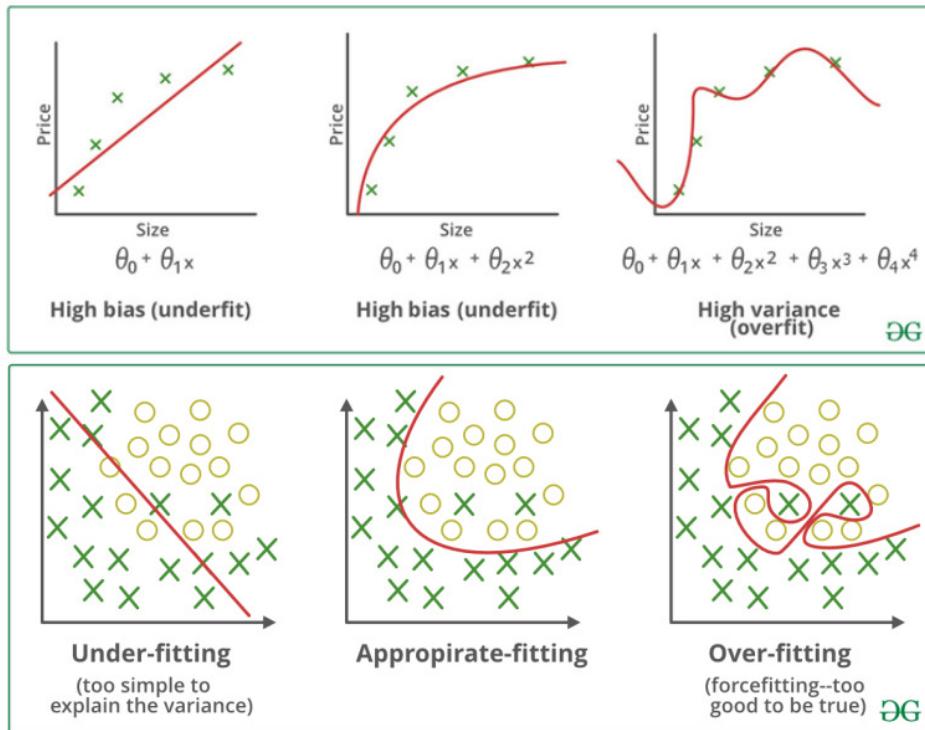
Abre tu cuenta



WUOLAH
+ BBVA

WUOLAH

Para medir la bondad de nuestra solución solamente contamos con nuestro conjunto de datos y debemos usarlo bien. Es posible que nuestro modelo aprenda nuestros datos perfectamente, minimizando el error a cero. En ese caso, nuestro modelo no será muy útil porque ha aprendido demasiado y seguramente ruido. Existen mecanismos para evitar este problema llamado sobreaprendizaje.



validación normal

La validación de nuestro modelo es una forma de estimar cómo de bueno es nuestro modelo ante datos nuevos. La validación simple consiste en partir nuestro conjunto de datos en dos (entrenamiento $\frac{3}{4}$ y test $\frac{1}{4}$). Si el modelo está entrenado correctamente su error en el conjunto de entrenamiento no debe ser muy diferente que en el conjunto test. Esto querrá decir que no ha sobre-aprendido y que el modelo generaliza bien con datos nuevos.

Si tenemos varios modelos y queremos elegir el mejor entre ellos, se opta por partir los datos en tres conjuntos (entrenamiento, validación y test). El tercer conjunto se hace para no sobreaprender ya que las dos primeras particiones las usamos para entrenar y elegir el mejor modelo.

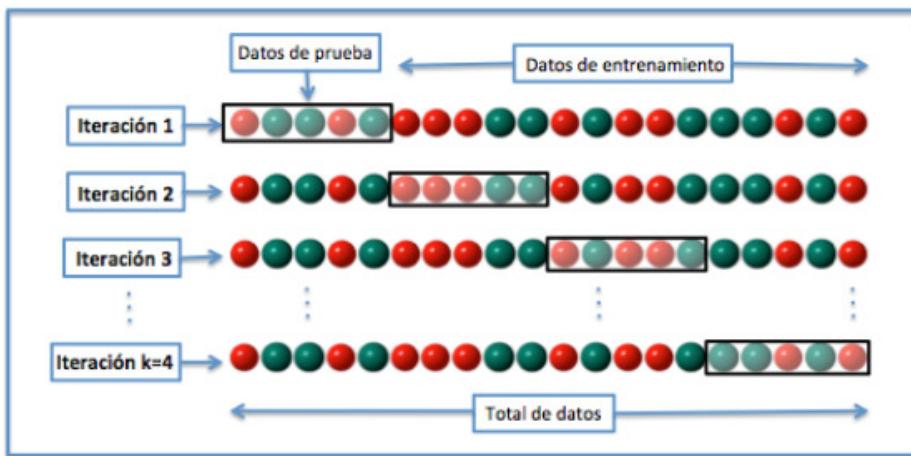
Existen estrategias más sofisticadas para realizar el entrenamiento y validación de un modelo:



-validación cruzada en k partes

Es una forma de estimar cómo de bueno es nuestro modelo ante datos nuevos. Consiste en los siguientes pasos:

1. Los datos se dividen aleatoriamente en k subconjuntos del mismo tamaño.
2. Se entrena el conjunto con $(k - 1)$ y se valida con el restante.
3. Se repite k veces el paso 2, cambiando el conjunto que se usa para validar.
4. Como medida de error final se presenta la media de las k medidas de error de validación.



Puede ser interesante estratificar las k partes. En un problema de regresión significa que la media de la variable respuesta sea similar en todas las partes. En un problema de clasificación significa que la proporción de clases en cada parte sea similar.

Si llevamos la validación cruzada al extremo y tomamos k como el número de individuos en nuestro conjunto de datos, estaríamos utilizando la validación Leave One Out validation, es adecuada cuando tenemos pocos datos y no tiene sentido dividirlo en partes mayores.

-los k vecinos más cercanos (k-NN):

Los k vecinos más cercanos es una técnica de aprendizaje supervisado muy sencilla. No calcula ningún modelo y demora todos los cálculos hasta el momento en que se le presenta un ejemplo nuevo. Usa todos los individuos disponibles ante un ejemplo nuevo y recupera los más relevantes para componer la solución.

Dado un ejemplo nuevo con variable de salida desconocida, recupera los k ejemplos que más se le parezcan de los n ejemplos disponibles. Devuelve la media de la solución de los k ejemplos recuperados. Se puede también devolver la media ponderada, donde se da más peso a los vecinos según su cercanía. Si la variable de salida es categórica, devuelve el valor más frecuente entre los valores solución de los k individuos recuperados. Aunque el k-NN no estime ningún modelo, debemos determinar el valor de k. Si es muy pequeño corre el riesgo de sobreaprender y si es muy grande corre el riesgo de generalizar demasiado. Para que no pase ni uno ni lo otro, podemos usar una estrategia de validación cruzada para encontrar el valor de k que minimiza el error de validación.



Hazte cliente de BBVA y... ahórrate 6 meses de suscripción

WUOLAH
+ BBVA

NETFLIX

Spotify

HBO max

Disney+

PlayStation Plus

DAZN

Ahora, si te abres una Cuenta Online en BBVA, te reembolsamos una de estas suscripciones durante 6 meses (hasta 9,99€/mes) al pagarla con tu tarjeta Aqua Débito

Promoción solo para nuevos clientes de BBVA. Válida hasta el 30/06/2023. Estas empresas no colaboran en la promoción.

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

árboles de decisión

Los árboles de decisión son una técnica de aprendizaje supervisado, pueden usarse en problemas de clasificación o de regresión. Permite el aprendizaje de conceptos de forma inductiva.

El árbol de decisión se construye recursivamente, a partir de un conjunto de ejemplos de entrenamiento. En cada nodo se pregunta por el valor de una variable que busca disminuir la entropía de los nodos hijos. Se selecciona aquel atributo que mayor disminución de entropía genera. La ganancia de información o disminución de la entropía es la diferencia entre la entropía del conjunto original y la de los subconjuntos obtenidos. El atributo usado en el nodo no se volverá a usar en otros niveles, se elimina del conjunto de atributos.

La terminación de un árbol de decisión ocurre cuando todos sus ejemplos pertenecen a la misma clase (la entropía es nula) o cuando no quedan más atributos para expandir nodos. Siempre existe el riesgo de sobreaprendizaje, para evitar esto, hay alternativas que no generan árboles completos como usar mecanismos de poda o utilizar un umbral de entropía por debajo de la cual no ramificar. Además siempre conviene evaluar el sobreaprendizaje de un árbol mediante una estrategia de validación.

El árbol construido sirve para clasificar ejemplos nuevos del conjunto de validación. Para ello se coloca el ejemplo nuevo en la raíz y respondiendo las preguntas desciende por las ramas hasta llegar a un nodo hoja.

El árbol aprendido es muy interpretable y debe ser analizado por un experto para entender la clasificación. A veces es tan grande que es difícil de interpretar.

En los árboles de decisión la entropía juega un papel muy importante, vamos a hablar de ella. La entropía mide la ausencia de homogeneidad de un conjunto de ejemplos con respecto a su clase. La entropía inicial (antes de clasificar) de un nodo se define como:

$$E(N) = - \sum_{i=1}^p P(s_i) \log_2 P(s_i),$$

siendo $P(s_i)$ la probabilidad de que un ejemplo tenga la clase s_i en el nodo N , y calculándose únicamente para las clases observadas en el nodo N (para evitar el cero en el logaritmo)

La entropía final de un nodo N tras usar el atributo A que tiene q valores (a_j) se define como:

$$E(N|A) = \sum_{j=1}^q P(a_j) \left(- \sum_{i=1}^p P(s_i|a_j) \log_2 P(s_i|a_j) \right),$$

siendo $P(s_i|a_j)$ la probabilidad de que un ejemplo del nodo hijo con $A = a_j$ tenga la clase s_i

Abre tu cuenta



WUOLAH
+ BBVA

WUOLAH

Ahora vamos a hablar de los diferentes **mecanismos de poda**. Hemos comentado anteriormente que la poda se realiza para simplificar el árbol y ganar en capacidad de interpretación y legibilidad, tanto como para evitar el sobreaprendizaje y mejorar la capacidad de generalización.

-Poda mediante reducción del error:

Este tipo de poda requiere un conjunto de ejemplos no usados para generar el árbol.

1. Clasifica dichos ejemplos y calcula el error que cometan todas las hojas
2. El número de errores de un subárbol es la suma de los errores de todas sus hojas
3. Calcula el número de errores que cometería el nodo que origina el subárbol si no se expandiera
4. Si los errores sin expandir son iguales o menores que expandido, se poda el subárbol
 - Se recalcula el número de errores de los subárboles que contenían dicho subárbol
 - **Ventaja:** Evita el sobreajuste de forma efectiva
 - **Inconveniente:** Requiere de nuevos ejemplos

-Poda pesimista:

1. Clasifica un conjunto de ejemplos (usado o no para generar el árbol)
2. Para cada nodo hoja, calcula los errores que comete y añade un valor prefijado r
3. El número de errores de un subárbol es la suma de los errores de todas sus hojas
4. Calcula el número de errores que cometería el nodo que origina el subárbol si no se expandiera e increméntalo también con r
5. Si los errores sin expandir son iguales o menores que expandido, se poda el subárbol
 - Se recalcula el número de errores de los subárboles que contenían dicho subárbol
 - **Ventaja:** No necesita ejemplos extra
 - **Inconveniente:** El número de elementos en la hoja no afecta

-redes neuronales:

Una red neuronal es una función capaz de representar complejas relaciones no lineales entre los datos de entrada y la variable a predecir. Las redes neuronales se basan en perceptrones multicapa.

Un perceptrón multicapa (MLP) es una red neuronal de tipo perceptrón con al menos tres capas. Tiene una capa de entrada con tantas neuronas como variables de entrada, una o más capas ocultas con un número variable de neuronas, cada una con su función de activación, y una capa de salida con una o más neuronas con su función de activación. Las capas ocultas son capaces de modelar relaciones no lineales entre las variables de entrada y salida.

Las funciones de activación son funciones no lineales que permiten modelar funciones complejas, calculan combinaciones lineales de funciones no lineales. Las funciones de activación pueden ser de distintos tipos: logística, tangente hiperbólica, rectificada lineal, identidad y softmax. Usaremos un tipo u otro depende de los rangos de las variables que estamos modelando.

- **Logística:** Se usa cuando la variable de salida es una probabilidad. Es una función sigmoidal que varía entre 0 y 1.
- **Tangente hiperbólica:** Se usa cuando la variable de salida es un número entre el -1 y el 1. Es una función sigmoidal que varía entre -1 y 1.
- **Identidad:** Se usa en la capa de salida para problemas de regresión no acotados.
- **Rectificada linear (ReLU):** Se usa cuando la variable de salida es un número positivo. Es una función con dos tramos, uno lineal que varía entre el 0 y el infinito (muy usada en Deep learning).
- **Softmax:** Se usa en la capa de salida en problemas de clasificación multiclas. Interpreta la probabilidad de que el ejemplo pertenezca a cada una de las n clases. Valores entre el 0 y 1 que suman 1.

A la hora de entrenar una red neuronal, el objetivo es configurar la red de forma que se minimice el error total cometido. Existe la función de error (función de pérdida) que calcula el error cometido por la red neuronal. Esta función de error depende de los pesos y bias de las neuronas. Dada una función de error diferenciable, podemos encontrar uno de sus mínimos locales mediante el algoritmo de descenso de gradiente.

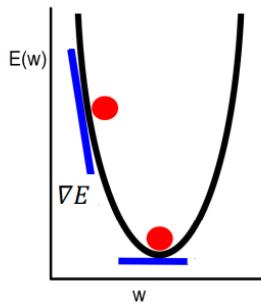
Partimos de un punto aleatorio p con un cierto error $E(p)$

Calculamos el gradiente de la función $\nabla E = (\frac{\partial E}{\partial w_1}, \dots, \frac{\partial E}{\partial w_s})$ en el punto p

- El vector gradiente indica la dirección en la que la función de error crece más rápido

Calculamos un nuevo punto $p = p - \alpha \nabla E$ y volvemos al paso anterior.

- α indica cuánto nos desplazamos y se conoce como tasa de aprendizaje (*learning rate*)



En cada paso del algoritmo, disminuimos el error hasta llegar a un mínimo local donde el gradiente será 0.

Si la función no es convexa no se garantiza encontrar un mínimo global.

El algoritmo de retropropagación disminuye el error de la red mediante el algoritmo de descenso de gradiente. Funciona de la siguiente manera, se inicializan los pesos de forma aleatoria y con valores pequeños y para cada ejemplo k del conjunto de entrenamiento se hacen dos fases:

- Hacia delante: Calculamos la salida de la red para un ejemplo y calculamos el error entre la salida real de la red y la salida deseada para ese ejemplo.
- Hacia atrás: Calculamos el gradiente de la función de error con respecto a los pesos de la red para el ejemplo k y ajustamos los pesos para que reduzcan el error respecto al ejemplo k .

De esta forma se reduce el error con respecto al ejemplo k , pero en realidad nos interesa reducir el error total con respecto a todos los ejemplos. Así que calculamos el gradiente medio y modificamos los pesos de la red en dirección contraria. La actualización de la red para todo el conjunto de entrenamiento recibe el nombre de época.

Para evitar el sobreaprendizaje usamos técnicas de regularización. La que usaremos nosotros es la regularización L2. La idea de esta técnica es penalizar los pesos demasiado grandes para que la red generalice mejor. Usa una variable que permite controlar cuánto queremos regularizar. $L1$ (penalización) y dropout

A la hora de configurar un perceptrón multicapa debemos tener varias cosas en cuenta:

- Los datos de entrada deben ser numéricos, pueden ser valores sin normalizar o sin estandarizar. No pasa nada por utilizar variables irrelevantes, ya que se les acabará asignando peso cero.
- Determinar el número de capas y de neuronas puede hacerse probando varias configuraciones y determinando la mejor mediante validación cruzada. Debemos entender que un MLP con una única capa oculta suficientes neuronas puede



Hazte cliente de BBVA y... ahórrate 6 meses de suscripción

WUOLAH
+ BBVA



Ahora, si te abres una Cuenta Online en BBVA, te reembolsamos una de estas suscripciones durante 6 meses (hasta 9,99€/mes) al pagarla con tu tarjeta Aqua Débito

Promoción solo para nuevos clientes de BBVA. Válida hasta el 30/06/2023. Estas empresas no colaboran en la promoción.

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

resolver igual un problema que un MLP con más capas ocultas. Las capas ocultas añaden representación de los datos con un nivel de abstracción amyor y en principio es más fácil llegar a la solución a partir de ellas.

- A la hora de determinar qué función de activación utilizar tendremos en cuenta el tipo de problema.
- A la hora de dar un valor al parámetro de tasa de aprendizaje, debemos tener en cuenta que cuanto más pequeño sea su valor, mejores resultados podemos encontrar, pero más tiempo tarda el entrenamiento.
- Sobre qué tipo de técnica de regularización usar ya hemos determinado anteriormente que la única que vamos a utilizar es la regularización L2.
- A la hora de usar validación cruzada nos sirve cualquier estrategia, lo importante es no usar el mismo subconjunto de datos para entrenar, configurar y evaluar.



Abre tu cuenta



WUOLAH
+ BBVA

Aprendizaje supervisado

posibles ejercicios:

- Nos dan matriz de confusión y sacar exhaustividad...
- Nos dan error cuadrático y comentar teóricamente
- Preguntan por técnicas para que no sobreaprendan
 - validación cruzada
 - podas en árboles
 - alg. descenso de gradiente en redes neuronales
- Nos dan conjunto de datos y crear árbol
- Preguntas teóricas sobre un árbol dado
 - Variable más importante?
 - Podrias este nodo?
 - Entropía en este nodo?
 - precision, exhaustividad en nodo x?
- Nos dan valores de validación en k partes y comentar
- Diferencias entre K-mn y árbol de decisión.
- Cualquier pregunta teórica

El procesamiento del lenguaje natural es un área de intersección entre la Inteligencia Artificial y la lingüística, se basa en crear sistemas capaces de interpretar o generar documentos.

Los procesadores de lenguaje natural basados en técnicas probabilísticas aprenden a partir de datos prácticos. El uso de redes neuronales profundas permite encontrar patrones estadísticos complejos en los corpus de documentos (datos prácticos). **Permite entre otras cosas calcular la probabilidad de encontrar una frase o secuencia de palabras determinada o calcular la probabilidad de la siguiente palabra en una frase.** Como el texto es secuencial, podemos pensar que la probabilidad de una palabra depende de todas las anteriores, para ello necesitamos refrescar ciertas nociones de probabilidad.

La probabilidad de que una palabra sea (por ejemplo) ‘perro’, sabiendo que la primera palabra es ‘el’, es la fracción de veces que ‘el’ aparece seguido de ‘perro’ en nuestro corpus:

$$P(\text{perro} | \text{el}) = \text{numVeces}(\text{el}, \text{perro}) / \text{numVeces}(\text{el})$$

Si añadimos elementos a la probabilidad podemos afirmar que la probabilidad de cada palabra depende de todas las anteriores, esto se generaliza mediante la **regla de la cadena**.

$$P(w_1, \dots, w_n) = P(w_1) \cdot P(w_2 | w_1) \cdot P(w_3 | w_1, w_2) \cdots P(w_n | w_1, \dots, w_{n-1})$$

Esta hipótesis no es práctica, ya que no hay un corpus tan grande para asignar probabilidad a todas las posibles combinaciones de palabras. **En lugar de considerar la probabilidad de un elemento depende de todos los anteriores, supone que solamente los n-1 elementos anteriores tienen efecto sobre las probabilidades del siguiente elemento i-ésimo.**

Hipótesis de Markov: $P(w_i | w_1, \dots, w_{i-1}) \approx P(w_i | w_{i-(n-1)}, \dots, w_{i-1})$

El modelo **n-grama** es uno de los modelos estadísticos del lenguaje más simples pero más útiles. Se puede aplicar a fonemas, letras, sílabas, palabras... Utiliza la hipótesis de Markov para indicar que la dependencia es con los n-1 anteriores.

Modelo bigrama (n = 2): $P(w_n|w_{n-1})$

Modelo trígrama (n = 3): $P(w_n|w_{n-1}, w_{n-2})$

Supongamos este corpus

- <s> Yo quiero patatas </s>
- <s> Yo quiero patatas con carne </s>
- <s> No quiero carne </s>
- <s> Quiero carne con patatas </s>
- <s> Hoy quiero carne </s>
- <s> Quiero dormir </s>

Estimación óptima por máxima verosimilitud

$$P(w_n|w_{n-1}) = \frac{frec(w_{n-1}, w_n)}{frec(w_{n-1})}$$

<s> indica "inicio de oración"
</s> indica "final de oración"

Estos son las probabilidades de algunos bigramas del corpus

- $P(yo|<s>) = freq(<s>, yo) / freq(<s>) = 2/6 = 0,333$
- $P(no|<s>) = freq(<s>, no) / freq(<s>) = 1/6 = 0,167$
- $P(quiero|yo) = freq(yo, quiero) / freq(yo) = 2/2 = 1$
- $P(patatas|quiero) = freq(quiero, patatas) / freq(quiero) = 2/6 = 0,333$
- $P(carne|quiero) = freq(quiero, carne) / freq(quiero) = 3/6 = 0,5$
- $P(dormir|quiero) = freq(quiero, dormir) / freq(quiero) = 1/6 = 0,167$

Según este modelo, la continuación más segura de <s>Yo quiero...

- **$P(carne|quiero) > P(patatas|quiero) > P(dormir|quiero)$**
- La respuesta cambia si consideramos trigramas → $P(patatas | yo, quiero) = 1$

Probabilidades de bigramas obtenidos de un corpus supuesto

- $P(yo | <s>) = 0,25$ $P(quiero | <s>) = 0,75$
- $P(quiero | yo) = 0,5$ $P(tengo | yo) = 0,2$ $P(soy | yo) = 0,3$
- $P(ser | quiero) = 0,8$ $P(tomar | quiero) = 0,2$
- $P(café | tomar) = 0,6$ $P(leche | tomar) = 0,3$ $P(distancia | tomar) = 0,1$
- $P(artistas | ser) = 0,9$ $P(informático | ser) = 0,1$

$P(Yo quiero tomar café) =$

- $P(yo | <s>) P(quiero | yo) P(tomar | quiero) P(café | tomar) = 0,25 * 0,5 * 0,2 * 0,6 = 0,015$
- La frase entera puede no estar en el corpus. Lo normal es que no esté.
- Si un bigrama no está la probabilidad de la frase es cero!

La frase que se genera siguiendo la opción más probable es: "Quiero ser artista"

- $P(quiero | <s>) P(ser | quiero) P(quiero | artista) = 0,75 * 0,8 * 0,9 = 0,54$



Hazte cliente de BBVA y... ahórrate 6 meses de suscripción

WUOLAH
+ BBVA



Ahora, si te abres una Cuenta Online en BBVA, te reembolsamos una de estas suscripciones durante 6 meses (hasta 9,99€/mes) al pagarla con tu tarjeta Aqua Débito

Promoción solo para nuevos clientes de BBVA. Válida hasta el 30/06/2023. Estas empresas no colaboran en la promoción.

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

Hemos visto que cualquier n-grama que no esté en el corpus recibe probabilidad 0, podemos usar técnicas para aliviar los problemas de estimación de probabilidades en casos raros.

- **Alisado de Laplace:** calcula cualquier probabilidad condicional considerando que ha habido unas observaciones adicionales virtuales de todos y cada uno de los n-gramas posibles.

$$P(B|A) = \frac{\text{frec}(A, B) + t}{\text{frec}(A) + t * m}$$

Donde

- t es el número de observaciones virtuales adicionales
- m es el número de monogramas (palabras) existentes en el corpus

- **Alisado por interpolación lineal:** usa la probabilidad incondicional $P(w_2)$ calculada a partir de los datos para hacer que la probabilidad condicional $P(w_2|w_1)$ se parezca a ella.

$$P_{\text{Int}}(w_2|w_1) = \alpha P(w_2|w_1) + (1 - \alpha)P(w_2)$$

donde $\alpha \in [0,1]$ regula el peso que se le da a la probabilidad condicional y la probabilidad no condicionada

El valor de α se puede fijar empíricamente con el fin de ajustar el rendimiento

- También se puede hacer dependiente del "contexto"
 - Si existen muchos bigramas con la palabra w_1 entonces es mejor un valor alto
 - Si no existen muchos bigramas con la palabra w_1 es mejor un valor bajo

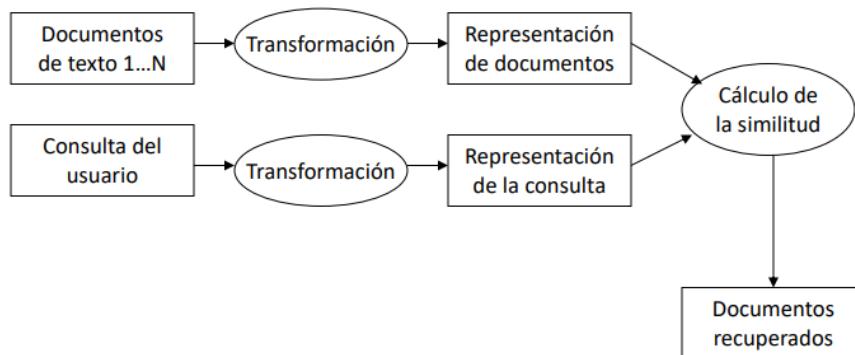


Abre tu cuenta



WUOLAH
+ BBVA

La recuperación de información consiste en encontrar los documentos relevantes asociados con una consulta.



A la hora de pasar un documento a datos, se representa como el conjunto de palabras que contiene. Este conjunto de palabras es denominado como bolsa de palabras.

A la hora de **vectorizar** los documentos, primero se enumeran todas las palabras en todos los documentos, después se eliminan los valores duplicados y se ordenan. Finalmente se convierte cada palabra en un valor y se crea un vector cuyo valor i ésimo corresponde al término i ésimo. Son valores dispersos donde la mayoría de los valores son 0.

Tenemos diferentes tipos de vectorizaciones, el tipo que indica la presencia o ausencia de un término, el tipo que cuenta la frecuencia de aparición del término en el documento y el tipo que cuenta la frecuencia TF/IDF según los documentos.

TF-IDF *Term frequency – Inverse document frequency*

Asigna un peso a cada término t en un documento d

$$w_{t,d} = tf_{t,d} \times \log\left(\frac{N}{df_t + 1}\right)$$

Frecuencia del término t en el documento d

Inversa de la frecuencia de aparición del término en los documentos
(cuanto más frecuente, menos peso $w_{t,d}$)

N es el número total de documentos

df_t es el número de documentos del corpus donde aparece el término t y se le suma 1 para evitar dividir por cero si no existe en ninguno

En el documento d , un término tendrá más peso si es muy frecuente en dicho documento y no aparece casi en el resto

Esto hace que términos que aparecen en todos los documentos no tengan mucho peso.

Por tanto tenemos un espacio vectorial, donde los términos son las dimensiones y los documentos son vectores en este espacio. Una colección de documentos se puede representar como una matriz término-documento. M términos como columnas y N documentos como filas.

Se utiliza la similitud del coseno para ver como de iguales son dos vectores de términos. Se calcula como el coseno del ángulo que forman los dos vectores de términos.

$$\text{coseno}(d_j, d_k) = \frac{d_j \cdot d_k}{|d_j||d_k|} = \frac{\sum_{i=1}^M w_{i,j} \cdot w_{i,k}}{\sqrt{\sum_{i=1}^M w_{i,j}^2} \sqrt{\sum_{i=1}^M w_{i,k}^2}}$$

Al igual que en los clasificadores, en los sistemas de recuperación de información también se puede medir el rendimiento, y de la misma manera.

Exhaustividad (recall) o Tasa de Verdaderos Positivos: $TVP = \frac{VP}{VP+FN} = \frac{VP}{P}$

Precisión o Valor Predictivo Positivo: $VPP = \frac{VP}{VP+FP}$

Medida F1: $F1 = 2 \cdot \frac{VPP \cdot TVP}{VPP + TVP} = \frac{2 \cdot VP}{2 \cdot VP + FP + FN}$

- Es la media armónica de Precisión y Exhaustividad

		Documento relevante	
		1	0
Documento recuperado	1	VP	FP
	0	FN	VN

Ventajas

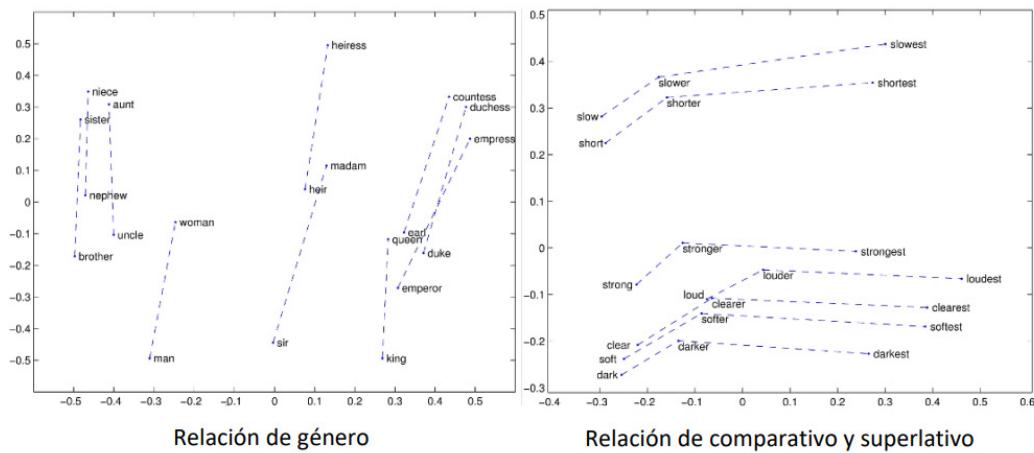
- Simplificación del problema que en muchos casos funciona
- Permite efectuar consultas de manera sencilla
- Permite trabajar textos con técnicas estadísticas y de aprendizaje automático

Inconvenientes

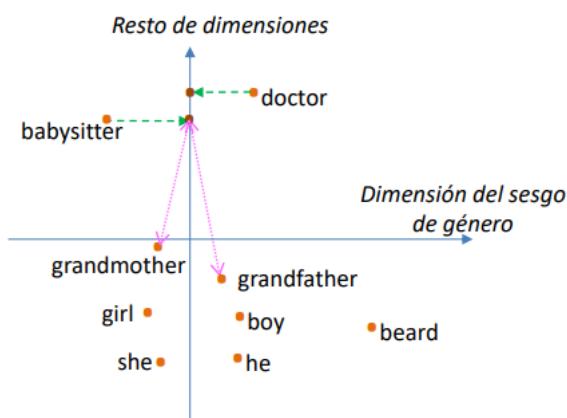
- No maneja la ambigüedad y la variabilidad léxica
 - Considera diferentes dos palabras sinónimas (casa y hogar)
 - Reconoce como iguales palabras polisémicas (ratón)
- Los vectores resultantes son dispersos (muchos ceros) lo cual es un problema para muchas técnicas de análisis de datos
- Requiere trabajar con corpus específicos si trabajamos en dominios especializados (con léxico propio), p.ej. medicina

Las **word-embedding** son representaciones de palabras en espacios de varias dimensiones que permiten superar algunas de las limitaciones de las representaciones de palabras que hemos visto. Puede reflejar similitudes semánticas y plasmar relaciones complejas entre palabras. Esto se logra representando cada palabra del diccionario mediante un vector de varias dimensiones que permite capturar su relación con otras palabras.

Si mostramos las palabras en un espacio denso donde las dimensiones tienen un significado, la posición de las palabras en dicho espacio nos indicaría la relación que tienen entre ellas. Las palabras “red”, “blue”, “yellow”, etc deberían estar relativamente cercanas, ya que suelen jugar un papel parecido en las frases. La palabra “tomato” debería estar más cerca de la palabra “red” que, por ejemplo, de “blue” ya que están asociadas con más frecuencia. Sería posible inferir que “man” es a “woman”, lo que “king” a “queen” si el espacio n-dimensional recoge las relaciones de género entre palabras.



Los word-embedding pueden capturar sesgos de los textos que se usaron para entrenarlos. En general aprenden asociaciones estereotipadas de género, raza, edad... Esto es un problema porque generan sistemas que reproducen esos sesgos. Existen mecanismos para eliminarlos.



Hazte cliente de BBVA y... ahórrate 6 meses de suscripción

WUOLAH
+ BBVA



Ahora, si te abres una Cuenta Online en BBVA, te reembolsamos una de estas suscripciones durante 6 meses (hasta 9,99€/mes) al pagarla con tu tarjeta Aqua Débito

Promoción solo para nuevos clientes de BBVA. Válida hasta el 30/06/2023. Estas empresas no colaboran en la promoción.

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

Abre tu cuenta



WUOLAH
+ BBVA

La **clasificación y agrupamiento de documentos** consiste en clasificar documentos en categorías predefinidas. Esto se puede lograr con aprendizaje supervisado mediante datos de entrenamiento y test, o con aprendizaje no supervisado, mediante clustering.

Como hemos visto con anterioridad se trabajan los datos en forma de matriz, donde las filas son documentos y las columnas las palabras, en las celdas tendremos el valor vectorizado.

Para poder categorizar un texto primero debemos construir un conjunto de entrenamiento clasificando cada documento, después crearemos un clasificador (Naive Bayes, k-NN...) y finalmente el clasificador clasifica documentos y deberemos evaluarlos.

El **clasificador Naive Bayes** simplifica la realidad asumiendo que todas las variables utilizadas para clasificar son independientes. Esta hipótesis será en muchos casos incorrecta pero obtiene resultados mejores que otras técnicas de clasificación más sofisticadas.

$$P(y = y_j | \mathbf{x} = \mathbf{x}_i) \propto P(y = y_j) \prod_{k=1}^d P(x_k = x_{ki} | y = y_j)$$

La probabilidad de observar cada clase y_i es la frecuencia relativa de casos observados en el conjunto de entrenamiento en la clase

$$P(y = y_j) = freq(y = y_j)$$

La probabilidad de observar el término x_k en el documento i condicionada que el documento sea de la clase y_i se calcula mediante frecuencias relativas como sigue

$$P(x_k = x_{ki} | y = y_j) = \frac{freq(x_k = x_{ki} \wedge y = y_j)}{freq(y = y_j)}$$

$$P(y = y_j | \mathbf{x} = \mathbf{x}_i) \propto P(y = y_j) \prod_{k=1}^d P(x_k = x_{ki} | y = y_j)$$

$$P(x_k = x_{ki} | y = y_j) = \frac{freq(x_k = x_{ki} \wedge y = y_j)}{freq(y = y_j)}$$

Emails de cada tipo

SPAM	No-SPAM
20	200

Prob. cond. de presencia de término

	SPAM	No-SPAM
millonario	0,25	0,025
Viagra	0,25	0,02
Nigeria	0,1	0,05
estimado	0,75	0,6
examen	0,1	0,25
ejercicio	0,2	0,2
clase	0,1	0,125

Prob. cond. de ausencia de término

	SPAM	No-SPAM
millonario	0,75	0,975
Viagra	0,75	0,98
Nigeria	0,9	0,95
estimado	0,25	0,4
examen	0,9	0,75
ejercicio	0,8	0,8
clase	0,9	0,875

Nuevo email a clasificar

$x_N = \{\text{millonario, Nigeria, estimado, clase}\}$

$$P(y = \text{SPAM} | \mathbf{x} = \mathbf{x}_N) = \frac{20}{220} (0,25 \cdot 0,75 \cdot 0,1 \cdot 0,75 \cdot 0,9 \cdot 0,8 \cdot 0,1) = 0,92 \cdot 10^{-4}$$

$$P(y = \text{No-SPAM} | \mathbf{x} = \mathbf{x}_N) = \frac{200}{220} (0,025 \cdot 0,98 \cdot 0,05 \cdot 0,6 \cdot 0,75 \cdot 0,8 \cdot 0,125) = 0,501 \cdot 10^{-4}$$

Es más verosímil que el email recibido sea SPAM

En los clasificadores podemos usar word-embedding en vez de tener un corpus de documentos.

PLN

posibles ejercicios:

- Nos dan corpus y calcular siguiente palabra usando n-gramas
- Vectorizar términos de diferentes maneras y mediante la similitud del coseno recuperar el documento más parecido a la consulta.
- Nos dan términos en categorías y determinar en qué categoría está una consulta usando Naïve Bayes

Las **redes semánticas** son un tipo de representación del conocimiento basado en grafos dirigidos, donde los conceptos son nodos y las aristas son relaciones específicas entre los conceptos. También tienen la inferencia, que es la herencia usada para una navegación organizada en el grafo.

Tipos de aristas/arcos:

A) Arcos estructurales: primitivas

- a) **InstanceOf:** instancia o ejemplar (Juan ejemplar Persona)
- b) **SubclassOf:** subclase (persona subclase vertebrado)
- c) **PartOf:** tiene parte (persona tiene parte boca)

B) Arcos descriptivos: semántica dependiente del dominio

- a) **Propiedades:** profesión, color de pelo...
- b) **Relaciones (no estructurales):** Amigo de, parente de...

Tipos de inferencia

A) Herencia de propiedades:

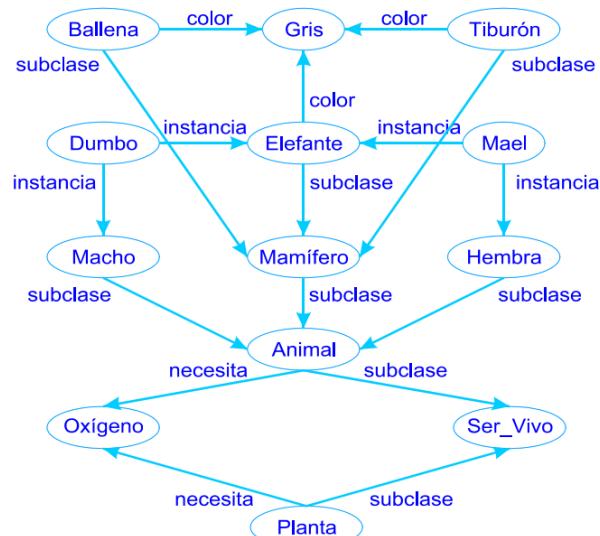
La notación de redes semánticas hace muy conveniente la utilización de razonamiento basado en herencia. Los nodos acceden a las propiedades definidas en otros nodos siguiendo los arcos InstanceOf SubclassOf.

¿De qué color es Dumbo?

- Gris

¿Qué puedo decir de Dumbo?

- Es un elefante
 - Es de color gris
- Es un macho
- Es un mamífero
- Es un animal
 - Necesita oxígeno
- Es un ser vivo



Este tipo de inferencia nos puede dar errores ya que, algunas propiedades de la clase en sí, no son heredables por los individuos de la clase. Para controlar excepciones tienen la regla de que se hereda el valor de la propiedad del nodo más cercano, y que si hay varios valores distintos a la misma distancia la respuesta será múltiple o no determinada.

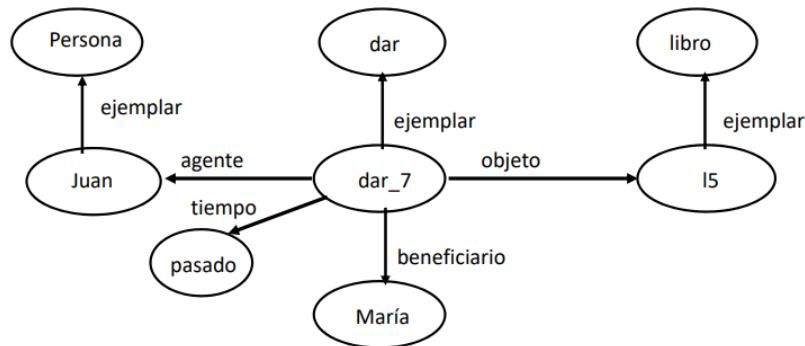
B) Intersección entre dos conceptos:

Para conocer la relación entre dos conceptos se utiliza un mecanismo de propagación de la activación. Este mecanismo se basa en activar ambos conceptos y propagar la activación a los nodos adyacentes hasta que se forma un camino entre los dos conceptos, si hay más de un camino entre los conceptos, se dice que tienen varias relaciones distintas.

Para buscar la intersección a veces es necesario generar la inversa de una relación, algunos sistemas lo hacen automáticamente con los arcos estructurales.

Una red semántica es la forma natural de representar relaciones correspondientes a predicados binarios en lógica, sin variables. Pero a la hora de representar **sucesos** necesitamos representar relaciones n-arias . Necesitamos convertirlas a relaciones binarias, para ello creamos un nuevo objeto que representa la relación concreta y se introducen predicados binarios para describir la relación de ese nuevo objeto con sus argumentos originales.

Juan dio el libro a María



- El **objeto** del suceso es un libro concreto que no está representado como tal en la frase dada por el usuario ➔ el sistema crea un objeto, ejemplar de libro y le da un nombre (*I5*)
- *Juan* sí es un individuo concreto al igual que *María*
- Este tipo de representación contesta preguntas de distinto tipo sobre el conocimiento que tenemos representado

Hazte cliente de BBVA y... ahórrate 6 meses de suscripción

WUOLAH
+ BBVA



Ahora, si te abres una Cuenta Online en BBVA, te reembolsamos una de estas suscripciones durante 6 meses (hasta 9,99€/mes) al pagarla con tu tarjeta Aqua Débito

Promoción solo para nuevos clientes de BBVA. Válida hasta el 30/06/2023. Estas empresas no colaboran en la promoción.

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

Abre tu cuenta

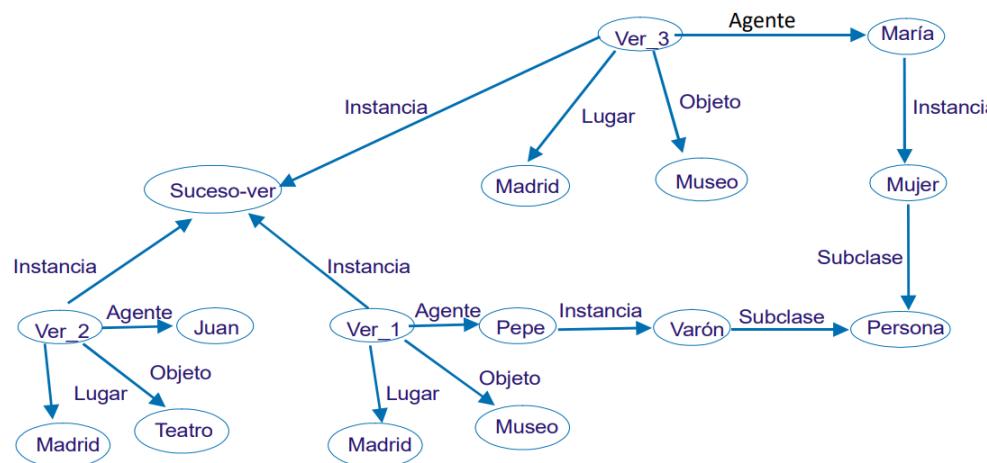


WUOLAH
+ BBVA

C) Contestar preguntas / recuperar información:

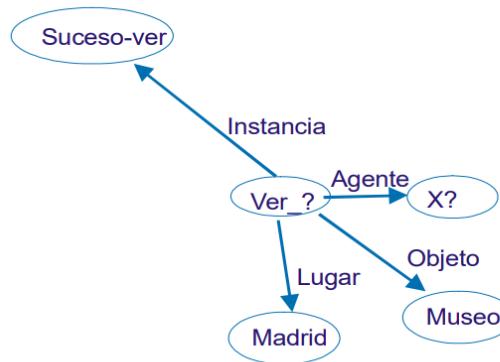
Para contestar una pregunta se representa la pregunta en forma de grafo consulta, se buscan fragmentos de la base de conocimiento que encajen con la consulta, donde cada posible encaje encontrado es una respuesta a la consulta.

- Base de conocimiento: tres instancias del Suceso ver



Transparencias de redes semánticas. Asunción Gómez Pérez. Facultad de Informática. UPM.

- Consulta: ¿quién vio un museo en Madrid?



Dos respuestas:

Equiparación 1:

$$\begin{aligned} \text{Ver}_? &\equiv \text{Ver}_1 \\ X? &\equiv \text{Pepe} \end{aligned}$$

Equiparación 2:

$$\begin{aligned} \text{Ver}_? &\equiv \text{Ver}_3 \\ X? &\equiv \text{María} \end{aligned}$$

Web semántica - Linked data - SPARQL

Aquí lo que nos interesa son SPARQL y ontologías.

SPARQL:

Lenguaje similar a SQL para hacer consultas sobre bases de conocimiento, en nuestro caso, wikidata. Estructura de una consulta:

SELECT (DISTINCT elimina duplicados) ?variables a recuperar,
para contarlas (COUNT(?x AS ?xx))
para valor max (MAX(?age) AS ?maxAge)
para valor min (MIN(?age) AS ?MinAge)
WHERE condiciones /tripletas (sujeto, predicado, objeto)
FILTER filtramos las variables recuperadas por condición
OPTIONAL {tripletas} para que recupere resultados sin valores para alguna variable
{consulta 1} UNION {consulta 2} para recuperar la unión de las dos consultas

GROUP BY ?node, para agrupar resultados con propiedades comunes
ORDER BY ASC/DES (?x)
LIMIT limita el número de resultados a recuperar
OFFSET indica el primer resultado que se debe mostrar

Lengua natural	ejemplo	SPARQL	ejemplo
oración	Julietta ama Romeo.	Punto	Julietta ama Romeo .
Conjunción (cláusula)	Romeo ama Julietta 'y' se mata .	Punto y coma	Romeo ama Julietta ; mata Romeo.
Conjunción (sustantivo)	Romeo mata Tybalt 'y' él mismo.	Coma	Romeo mata tybalt, Romeo.
Cláusula relativa	Julietta ama a alguien quien mata Tybalt.	Corchetes	Julietta ama [mata Tybalt].

Este por nodos es interesante:

```
%%wdsparql
##Isaac Asimov = Q34981
##Conyuge = P26
##Fecha inicio = P580
##fecha fin = P580

SELECT ?mujer ?mujerLabel ?fechaInicio ?fechaFin
WHERE
{
    wd:Q34981 p:P26 [
        ##apuntamos al nodo de conyuge de asimov
        ##ya que en el nodo de conyuge se guarda la fecha de inicio y de final de cada
        ps:P26 ?mujer;
        ##la declaración de la propiedad (nodo) es el mismo nombre del conyuge
        pq:P580 ?fechaInicio;
        pq:P582 ?fechaFin].
        ##los cualificadores de la propiedad (nodo) son la fecha de inicio y fin

    SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
```

Este con groupby es un ejemplo interesante:

```
%%wdsparql
## autor = P50
## Isaac Asimov = Q34981
## género literario = P136
## obras literarias de ciencia ficción = Q3238422
## subclase de = P279
## fecha de publicación = P577
SELECT ?obra ?obraLabel (MIN(?fecha) AS ?minFecha)
WHERE
{
    ?obra wdt:P50 wd:Q34981;
           wdt:P136 ?genero.
    ?genero wdt:P279* wd:Q3238422. ##hasta aqui es como la busqueda anterior
    OPTIONAL{?obra wdt:P577 ?fecha.} ##ponemos optional para que aparezcan aunque no tengan fecha de publicación asociada
    ##preguntamos por la fecha de publicacion de cada obra

    SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
GROUP BY ?minfecha ?obra ?obraLabel
##hacemos una agregacion ahora si una obra tiene más de una fecha de publicación se repite con sus diferentes fechas de publicacion
##pero como hemos puesto MIN() en select de las muchas fechas escoge la mas vieja
ORDER BY (?minfecha)
```

Este por filtrado de fecha:

```
##Autor = P50
##Isaac Asimov = Q34981
##Fecha de publicación = P577

%%wdsparql

SELECT ?obra ?obraLabel ?fecha
WHERE
{
    ?obra wdt:P50 wd:Q34981;
           wdt:P577 ?fecha.
    ##sacamos todas las obras que escribio Asimov y la fecha de publicacion
    FILTER("1970-01-01"^^xsd:dateTime <= ?fecha && ?fecha < "1980-01-01"^^xsd:dateTime).
    ##despues usamos filter para que solo nos de como resultado las que estan entre 1970-1980
    SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
ORDER BY ASC(?fecha)
##no se ppide pero las ordenamos de manera ascendente por fecha para ver que efectivamente da el reultado correcto
```

Y este por filtrado de existencia:

```
%wdsparql
## autor = P50
## Isaac Asimov = Q34981
## género literario = P136
## ciencia ficción = Q24925
## centro de estudios = P69
##fecha de nacimiento = P569
## fecha de defunción = P570

SELECT DISTINCT ?autor ?autorLabel ?fechaNacimiento ?fechaDefuncion ?colegioOtros ?colegioOtrosLabel ##usamos distinct para que no nos salgan valores duplicados
WHERE
{
    ?obra wdt:P136 wd:Q24925;
        wdt:P50 ?autor. ##consultamos las obras de ciencia ficcion y despues sus autores
    ?autor wdt:P69 ?colegioOtros.##consultamos donde estudiaron los autores
    OPTIONAL{ ?autor wdt:P569 ?fechaNacimiento. }##consultamos fecha de nacimiento de autores
    OPTIONAL {?autor wdt:P570 ?fechaDefuncion. } ##consultamos fecha de defuncion de autores

    FILTER EXISTS {
        wd:Q34981 wdt:P69 ?colegio.
        FILTER(?colegioOtros = ?colegio)
    }
##filtramos por existencia de autores que estudiaron en uno de los sitios donde fue educado asimov

SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
ORDER BY ?colegioOtrosLabel ?autor
#ordenados alfabéticamente por centro y luego por autor.
```

Ontologías

Componentes:

- Clases o conceptos
- Propiedades o roles o atributos
- Instancias o individuos

Diseño y desarrollo de ontologías:

Los conceptos en la ontología deben ser próximos a las entidades del dominio. Extraer las frases que describen el dominio, las clases serán nombres y las propiedades, verbos, por ejemplo.

- 1) Enumeramos los términos importantes
- 2) Definimos las clases y la jerarquía de clases
- 3) Definimos las propiedades de las clases
- 4) Definimos las facetas de las propiedades
- 5) Creamos instancias



Hazte cliente de BBVA y... ahórrate 6 meses de suscripción

WUOLAH
+ BBVA



Ahora, si te abres una Cuenta Online en BBVA, te reembolsamos una de estas suscripciones durante 6 meses (hasta 9,99€/mes) al pagarla con tu tarjeta Aqua Débito

Promoción solo para nuevos clientes de BBVA. Válida hasta el 30/06/2023. Estas empresas no colaboran en la promoción.

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.



Abre tu cuenta



WUOLAH
+ BBVA

El lenguaje que vamos a utilizar para representar ontologías será OWL. En OWL una base de conocimiento se compone de parte terminológica, donde se definen las clases y propiedades y parte assertiva donde se definen los individuos.

Las clases o conceptos son predicados unarios, los roles o propiedades son predicados binarios y los individuos son constantes.

Constructor	Sintaxis	Descripción
T (top)	Thing	Cualquier entidad
⊥ (bottom)	Nothing	Conjunto vacío (Clase insatisfacible)
Clases primitivas	Dog, Cat, Fiendly	$A \in N_C$
AND	Dog and Friendly	Perros que son amigables
OR	Dog or Cat	Tanto los perros como los gatos
NOT	not Friendly	Entidades no amigables
ONE OF	{monday, tuesday, ..., sunday}	Días de la semana (enumerado)
EXISTS	hasPet some Dog	Entidades que tienen perros como mascota
FOR ALL	hasPet only Dog	Entidades que sólo tienen perros como mascota
MIN CARD	hasPet min 2	Entidades que tienen al menos 2 mascotas
MAX CARD	hasPet max 3	Entidades que tienen como máximo 3 mascotas
CARDINALITY	hasPet exactly 1	Entidades que tienen exactamente una mascota

Axioma	Sintaxis
Subclass	Dog subclassof Animal
Equivalent	Friendly equivalent hasFriend some Animal
Disjoint	disjoint (Dog, Cat, Person)

Axioma	Sintaxis
Subproperty	hasElectronicDevice subpropertyof hasDevice
Equivalent	hasEletronicDevice equivalent hasGadget

Axioma	Sintaxis
Concept assertion	Dog(snoopy)
Role assertion	hasFriend(juan, snoopy)
Same	isabel = isa
Different	juan ≠ isabel

En OWL:

Usamos razonamiento en mundo abierto, no se asume nada de lo que no se conoce. Si no está en la base de conocimiento no lo sabe.

- Base de conocimiento:

Cat **subclassof** Animal
Dog **subclassof** Animal
Pet **subclassof** Cat **or** Dog
Person **subclassof** Animal
Friendly **equivalent** hasFriend **some** Animal
FriendlyPerson **equivalent** Person **and** hasFriend **some** Person
PetPerson **equivalent** Person **and** hasPet **some** Pet
PetPerson2 **equivalent** hasPet **some** Thing
hasPet: Person → Pet

- ¿Cuáles de las siguientes inferencias se pueden realizar?

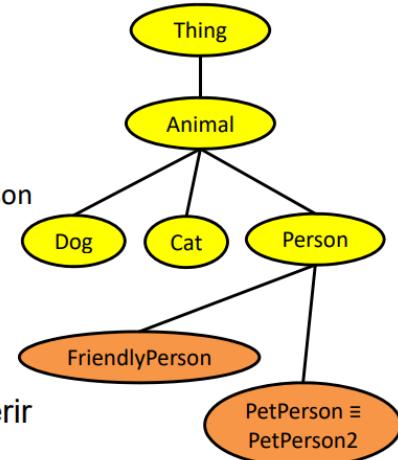
FriendlyPerson subclassof Animal	✓	FriendlyPerson subclassof Friendly	✓
Pet subclassof Cat	✗	Cat subclassof Pet	✗
Animal equivalent Cat or Dog or Person	✗	PetPerson equivalent PetPerson2	✓

:

- Base de conocimiento:

Cat **subclassof** Animal
Dog **subclassof** Animal
Pet **subclassof** Cat **or** Dog
Person **subclassof** Animal
Friendly **equivalent** hasFriend **some** Animal
FriendlyPerson **equivalent** Person **and** hasFriend **some** Person
PetPerson **equivalent** Person **and** hasPet **some** Pet
PetPerson2 **equivalent** hasPet **some** Thing
hasPet: Person → Pet

- La jerarquía de clases primitivas se aserta
- La jerarquía de clases definidas se puede inferir



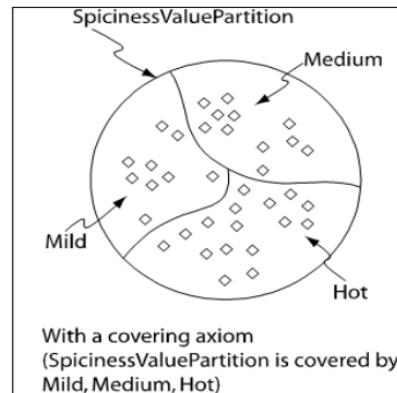
- Usamos particiones de valores para indicar que un conjunto es la unión de varios conjuntos disjuntos

Spiciness **equivalent** Mild **or** Medium **or** Hot
disjoint (Mild, Medium, Hot)

- Ahora podemos deducir que si algo es de tipo Spiciness pero no es Mild ni Medium entonces debe ser Hot

- ¡Ojo!, no es lo mismo que

Mild **subclassof** Spiciness
Medium **subclassof** Spiciness
Hot **subclassof** Spiciness
disjoint (Mild, Medium, Hot)



- Base de conocimiento:

Cat **subclassof** Animal
Dog **subclassof** Animal
Pet **subclassof** Cat **or** Dog
Person **subclassof** Animal
Friendly **equivalent** hasFriend **some** Animal
PetPerson **equivalent** Person **and** hasPet **some** Pet
VeryPetPerson **equivalent** hasPet **min** 2
hasPet: Person → Pet
disjoint(Cat, Dog, Person)

Cat(pelusa)
Dog(leo)
Pet(nieve)
Persona(alicia)
hasPet(alicia, pelusa)
hasPet(alicia, leo)
Persona(juan)
hasFriend(juan, alicia)
hasPet(juan, leo)
hasPet(juan, nieve)

- ¿Cuáles de las siguientes inferencias se pueden realizar?

Animal(nieve)	✓	Cat(nieve)	✗
PetPerson(juan)	✓	VeryPetPerson(juan)	✗
Friendly(alicia)	✗	VeryPetPerson(alicia)	✓

- La **similitud de Resnik** se basa en la teoría de la información y considera que dos conceptos son más similares cuanta más información comparten

$$sim(A, B) = \max_{C \subseteq A \cap B} [-\log p(C)]$$

- donde:
 - A y B son los conceptos a comparar
 - C es cada concepto subsumido por la intersección de A y B
 - p(C) es la probabilidad de que una instancia pertenezca a C
- Cuando más abstracto sea un concepto C, mayor será p(C) y por tanto menos información aportará (-log p(C))

La **similitud de Jaccard** mide la similitud entre dos conceptos como la razón entre el número de instancias comunes y totales

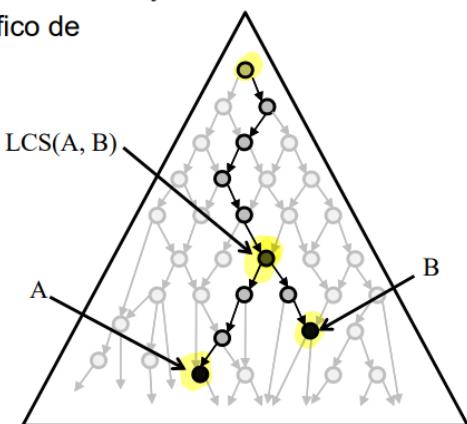
$$sim(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

- donde:
 - $|A \cap B|$ representa el número de instancias comunes
 - $|A \cup B|$ representa el número de instancias de la unión

La similitud basada en **distancias al LCS** mide la similitud entre dos conceptos como la razón entre las siguientes longitudes

$$sim(A, B) = \frac{\partial(\text{root}, C)}{\partial(\text{root}, C) + \partial(C, A) + \partial(B, C)}$$

- donde:
 - $\partial(A, B)$ es el mínimo número de aristas que conecta A y B
 - C = LCS(A, B) es el concepto más específico de la jerarquía que es más general que A y B (*least common subsumer*).
 - Puede haber varios



Hazte cliente de BBVA y...
ahórrate 6 meses de suscripción

WUOLAH
+ BBVA

NETFLIX

Spotify

HBO max

Disney+

PlayStation Plus

DAZN

Ahora, si te abres una Cuenta Online en BBVA, te reembolsamos una de estas suscripciones durante 6 meses (hasta 9,99€/mes) al pagarla con tu tarjeta Aqua Débito

Promoción solo para nuevos clientes de BBVA. Válida hasta el 30/06/2023. Estas empresas no colaboran en la promoción.

CONOCIMIENTO

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

posibles preguntas

- Consulta SPARQL
- Crear ontologías
- medidas de similitud.



Abre tu cuenta



WUOLAH
+ BBVA