

Descripciones vectoriales (**word vectors**)



¿Que es un vector de palabras Word Vector?

Estos son simplemente representaciones vectoriales numéricas que tratan de capturar el significado de una palabra. Se podría decir que estos vectores son las características que tiene una palabra y que serán de utilidad para su clasificación.

Haciendo un vector de palabras

Supongamos que queremos obtener una descripción (caracterización de las siguientes palabras) de las palabras

Hombre

Principe

Mujer

Princesa

Rey

Reina

Haciendo un vector de palabras **Eligiendo los descriptores**

Es necesario identificar qué conceptos son necesarios para realizar la descripción de las palabras, estas serán las dimensiones de nuestro vector

Hombre

Principe

Mujer

Princesa

Rey

Reina

Masculinidad	nobleza	riqueza	edad	Feminidad

Haciendo un vector de palabras **Eligiendo los descriptores**

Palabra	Masculinidad	nobleza	riqueza	edad	Feminidad
Hombre	1	0	0.5	0.7	0
Mujer	0	0	0.5	0.7	1
Princesa					
Principe					
Rey					
Reina					

Haciendo un vector de palabras **Eligiendo los descriptores**

Palabra	Masculinidad	nobleza	riqueza	edad	Feminidad
Hombre	1	0	0.3	0.7	0
Mujer	0	0	0.3	0.7	1
Princesa	0	0.8	0.7	0.3	1
Principe	1	0.8	0.7	0.3	0
Rey					
Reina					

Haciendo un vector de palabras **Eligiendo los descriptores**

Palabra	Masculinidad	nobleza	riqueza	edad	Feminidad
Hombre	1	0	0.3	0.7	0
Mujer	0	0	0.3	0.7	1
Princesa	0	0.8	0.7	0.3	1
Principe	1	0.8	0.7	0.3	0
Rey	1	1	1	0.7	0
Reina	0	1	1	0.7	1

Utilidad de las representaciones vectoriales

Ya que tenemos una representación numérica vectorial de una palabra podemos hacer operaciones entre los diferentes conceptos lo cual permite describir conceptos que incluso no definimos.

Palabra	Masculinidad	nobleza	riqueza	edad	Feminidad
Hombre	1	0	0.3	0.7	0
Rey	1	1	1	0.7	0

Rey - Hombre = Título nobiliario de rey

Título de Rey	0	1	.7	0	0
---------------	---	---	----	---	---

Utilidad de las representaciones vectoriales

También podemos sumar conceptos y construir el significado

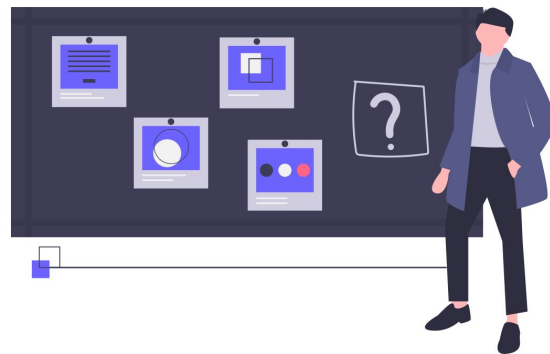
Palabra	Masculinidad	nobleza	riqueza	edad	Feminidad
Título de Rey	0	1	.7	0	0
Mujer	0	0	0.3	0.7	1

Título nobiliario de rey + Mujer = Reina

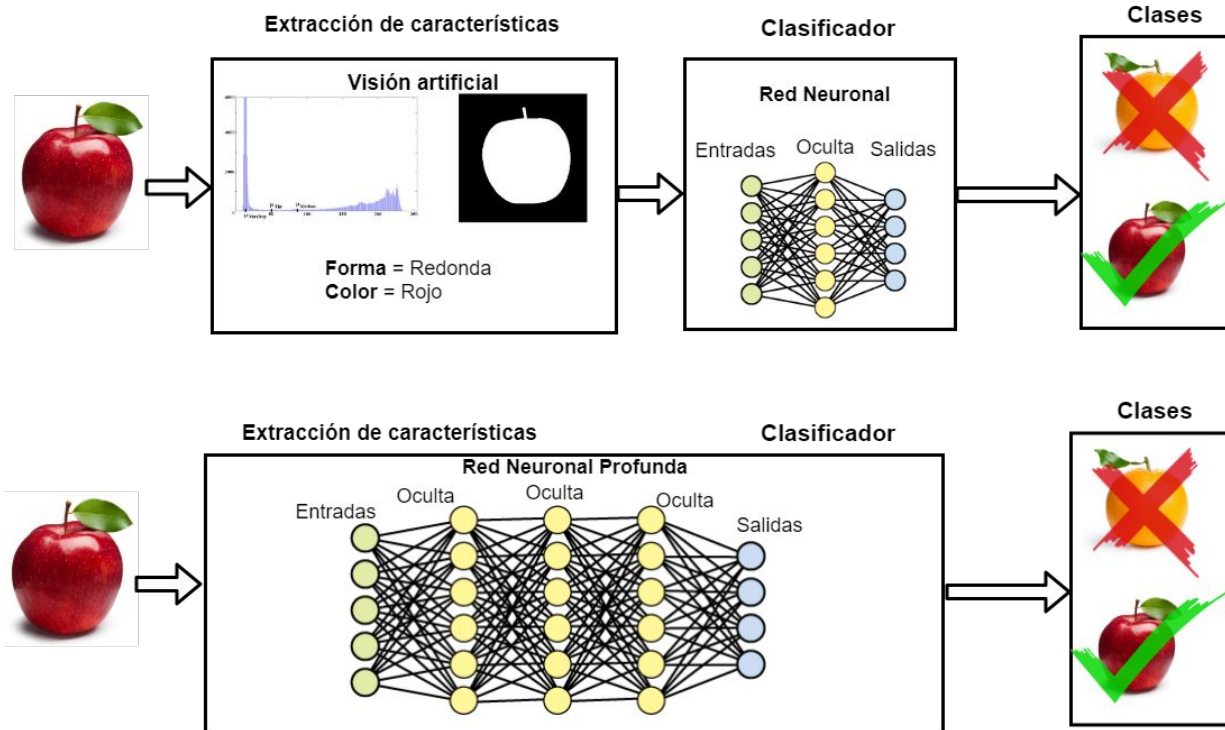
Reina	0	1	1	0.7	1
-------	---	---	---	-----	---

Pero cómo podemos llenar estos vectores sin tener que describir todos las palabras

El problema principal de estos vectores es que necesitamos describir todas las palabras a un nivel conceptual lo cual puede ser sencillo para unas cuantas pero no para todo el idioma español



Utilizando un extractor de características



Aprendiendo el significado del contexto



En múltiples ocasiones no tenemos la el conocimiento de una palabra específica en inglés. Pero podemos identificar más o menos su significado por el contexto del texto que estamos leyendo.

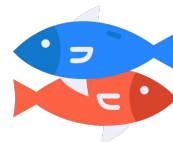
Ejemplo de contexto

Esto también aplica cuando el significado de una palabra es ambiguo por ejemplo **banco**

dinero, cheque, cajero



Mar, coral, arrecife



Bar, madera, asiento



Source Text

Training Samples

The quick brown fox jumps over the lazy dog. →

(the, quick)
(the, brown)

The quick brown fox jumps over the lazy dog. →

(quick, the)
(quick, brown)
(quick, fox)

The quick brown fox jumps over the lazy dog. →

(brown, the)
(brown, quick)
(brown, fox)
(brown, jumps)

The quick brown fox jumps over the lazy dog. →

(fox, quick)
(fox, brown)
(fox, jumps)
(fox, over)

