

# Dataset Clasificación: Vehicle

Ricardo Ignacio Shepstone Aramburu

## Índice

<b>Análisis exploratorio de datos</b>	<b>1</b>
Definición del problema . . . . .	1
Preparación de los datos . . . . .	1
Descripción de los datos . . . . .	1

## Análisis exploratorio de datos

### Definición del problema

El dataset consta un conjunto de parámetros que describen la silueta bidimensional de un objeto tridimensional. Los datos se obtuvieron de las siluetas producidas por el HIPS (Sistema de Procesamiento de Imágen Jerárquico) en un experimento donde se usaron cuatro vehículos diferentes: un autobús, furgoneta Chevrolet, un Saab 9000 y un Opel Manta.

El objetivo por lo tanto de este problema de clasificación será determinar a partir de los parámetros de las siluetas de estos vehículos determinar de qué vehículo se trata.

En cuanto a la dependencia de las variables, tenemos que tener en cuenta que en la mayoría de las especies, un individuo crece en tamaño y aumenta en peso a lo largo de su vida, hasta llegar a cierto límite. Estas variables a su vez son dependientes entre ellas, ya que un individuo de mayor tamaño tendrá un peso mayor. Por otro lado, la edad no está directamente relacionada con el sexo, pero debido al dimorfismo sexual que existe en la mayoría de especies, el tamaño de un individuo puede verse afectado por su sexo en mayor o menor grado según la especie. En este caso particular, la variable de sexo tiene un tercer valor “Infant” que puede proporcionar cierta información sobre la edad del individuo, por lo que esta variable también tiene cierto grado de dependencia con el resto.

En base a esta información se procede a plantear ciertas cuestiones e **hipótesis**:

- El tamaño y el peso aumentan con la edad.
- Si es así, ¿Que relación hay con respecto al número de anillos?
- ¿Alcanzan estas especies un límite de tamaño y peso?
- ¿Cómo es la relación entre las variables de dimensión y las de peso?
- ¿Aumentan todas las variables de tamaño en la misma proporción (están correlacionadas)? ¿Y las de peso?
- La variable sexo influye sobre las medidas físicas del individuo.
- ¿Hay dimorfismo sexual? ¿En qué grado?
- ¿Qué relación hay entre la edad y que un individuo sea joven (“infant”) o adulto?

### Preparación de los datos

#### Descripción de los datos

Con el dataset de vehicle se construye un data frame con 846 observaciones y 19 variables. No se dispone de mucha información del dataset por lo que no se podrá proporcionar una descripción detallada sobre las

variables. Pero en la siguiente lista se ha recopilado la poca información sobre las variables que hay disponible en el repositorio de UCI (<https://archive.ics.uci.edu/ml/datasets/Statlog+%28Vehicle+Silhouettes%29>)

Nombre	Tipo de dato	Información
Compactness	Cuantitativo discreto	$(\text{Perimetro medio})^2/\text{area}$
Circularity	Cuantitativo discreto	$(\text{Radio medio})^2/\text{area}$
Distance circularity	Cuantitativo discreto	$\text{area}/(\text{distancia media desde frontera})^2$
Radius ratio	Cuantitativo discreto	$(\text{radio máx.}-\text{radio mín.})/\text{radio medio}$
Praxis aspect ratio	Cuantitativo discreto	eje menor/eje mayor
Max_length_aspect_ratio	Cuantitativo discreto	longitud perp. de máx. longitud/máx longitud
Scatter_ratio	Cuantitativo discreto	inercia del eje menor/inercia del eje mayor
Elongatedness	Cuantitativo discreto	$\text{area}/(\text{ancho encogido})^2$
Praxis_rectangular	Cuantitativo discreto	$\text{area}/(\text{longitud eje}*\text{anchura eje})$
Length_rectangular	Cuantitativo discreto	$\text{area}/(\text{longitud máx.}*\text{longitud perp.})$
Major_variance	Cuantitativo discreto	momento de 2º orden en eje mayor/area
Minor_variance	Cuantitativo discreto	momento de 2º orden en eje menor/area
Gyration_radius	Cuantitativo discreto	$(\text{mavar}+\text{mivar})/\text{area}$
Major_skewness	Cuantitativo discreto	$(\text{momento de 3º orden en eje mayor})/\sigma_{\text{min}}^3$
Minor_skewness	Cuantitativo discreto	$(\text{momento de 3º orden en eje menor})/\sigma_{\text{maj}}^3$
Minor_kurtosis	Cuantitativo discreto	$(\text{momento de 4º orden en eje mayor})/\sigma_{\text{min}}^4$
Major_kurtosis	Cuantitativo discreto	$(\text{momento de 4º orden en eje menor})/\sigma_{\text{maj}}^4$
Hollows_ratio	Cuantitativo discreto	$(\text{area de huecos})/(\text{area del polígono unido})$
Class	Categorico nominal	Clase del vehículo