



Reconocimiento de patrones

Tarea 5

Clustering 2

Alumno:

Pérez Rodríguez Raúl Francisco

Octubre 2017

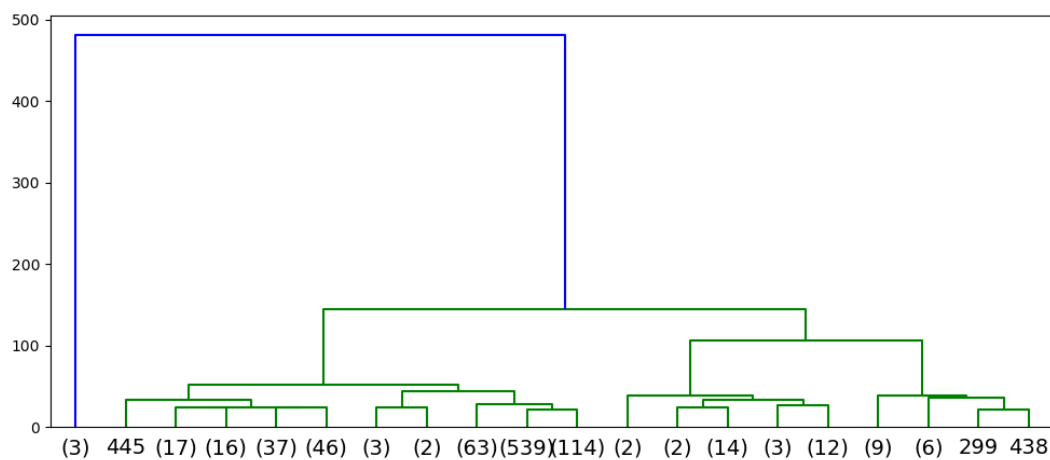
Obtenga los datos del ejemplo de sobrevivientes del Titanic elimine las columnas no numéricas y realice un análisis mediante dendrogramas de los datos.

El conjunto de datos iniciales del Titanic, contienen 10 variables con 891 instancias. Eliminando las variables no numéricas, quedan las variables Survived ,Pclass, Age, SibSp, Parch, Fare.

Data Dictionary

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

Utilizando la técnica de dendrograma con el método de centroid y truncando para obtener los últimos 20 clusters obtenemos la siguiente gráfica.



Analizando los últimos 20 clusters del dendrograma se pueden observar que puede haber entre 3 y 5 grupos.