# Práctica 8

# Aprendizaje Automático

Javier Pellejero Ortega & Zhaoyan Ni

Inteligencia Artificial

Grupo 11

Doble grado Matemáticas e ingeniería informática

1. Agrupamiento.

Para el agrupamiento hemos optado por una base de datos de semillas que podemos encontrar en el link [**http://archive.ics.uci.edu/ml/datasets/seeds**](http://archive.ics.uci.edu/ml/datasets/seeds), y mide una serie de atributos de hasta 3 tipos distintos de semillas de trigo (*Kama*, *Rosa* y *Canadian*) que comentaremos a continuación. En total hay datos de 70 de cada uno de los subtipos de semilla.

|  |  |  |
| --- | --- | --- |
| Nombre del Atributo | Descripción | Tipo |
| **area** | área de la semilla | Real |
| **perimeter** | perímetro de la semilla | Real |
| **compact** | compacidad que viene dada por la fórmula *C=4·pi·Area/perimeter2* | Real |
| ***lengthKernel*** | *longitud de la semilla* | Real |
| ***width*** | *ancho de la semilla* | Real |
| ***a*ssymmetryCoefficient** | coeficiente de asimetría | Real |
| **lengthKernelGrove** | longitud de la curvatura central de la semilla | Real |
| **varietiesOfWheat** | variedad de la semilla: 1-Kama, 2-Rosa y 3-Canadian | Categórico |

Comentemos ahora la preparación que hemos realizado sobre los datos. Para empezar, hemos optado por una normalización de los datos. Esto es debido a que vamos a utilizar el algoritmo jerárquico que maneja distancias.