

Entregable WEKA

Laura Rodríguez Navas

Marzo 2020

no más de 4 pàgines!!!!!!!!!!!!!!

12600 variables predictoras (todas numéricas) variable clase binaria tumor,
normal número de instancias disponibles (136)

Preparación de Datos

remove nan order for class variable

Uso de los clasificadores

cross-validation (5cv)

NaiveBayes

J48 (C4.5)

Técnicas de preprocesamiento

Selección de variables

AttributeSelectedClassifier

FilteredClassifier

Visualitzation

Confusion matrix ROC

1. inicialmente lanzaremos los clasificadores sobre la base de datos inicial, obteniendo así una medida inicial que intentaremos mejorar.

2. mejorar con la selección de variables

de variables es no lanzar una selección de tipo multivariada (filter o wrapper) sobre todas las variables, puesto que podría llevar mucho tiempo. Se recomienda primero hacer una selección univariada tipo ranker y quedarse con los mejores 200-500 atributos.