

Práctica 1: Clasificación usando ensembles de clasificadores

Duración: 3 sesiones

Objetivo: El objetivo de esta práctica es introducir los conceptos de clasificación usando agrupaciones (ensembles) de clasificadores.

Scikit-learn (<https://scikit-learn.org/stable/modules/ensemble.html>) provee diferentes métodos de clasificación basados en ensembles. Entre los más relevantes están:

- Bagging
- AdaBoost
- Gradient tree boosting
- Voting classifier
- Stacking
- Random forests

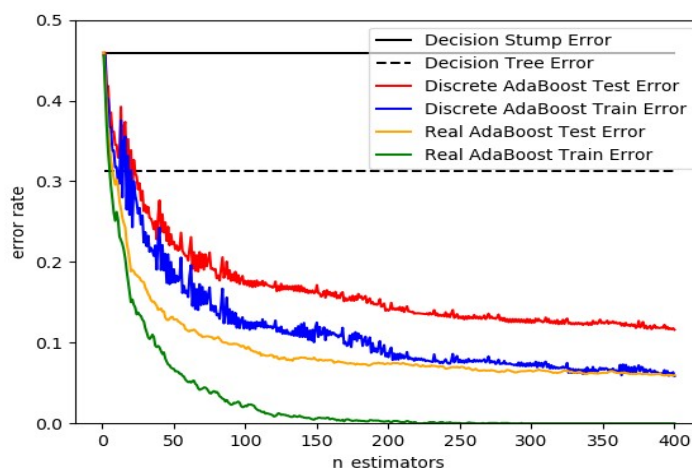
Además, usando los métodos de clasificación estándar se pueden construir métodos de ensembles como las diferentes variaciones de bagging y boosting de forma sencilla.

En esta práctica se va a estudiar el funcionamiento de los diferentes métodos de ensembles. En primer lugar se ha de seleccionar un método de clasificación “débil” que son los más apropiados para el uso con boosting. Métodos típicos son los árboles de decisión o el clasificador Naïve Bayes.

Seleccione un número suficiente de problemas de los repositorios usuales (UCI MLR, Weka, etc.). Se deben escoger varios problemas con diferente número de clases, variables e instancias.

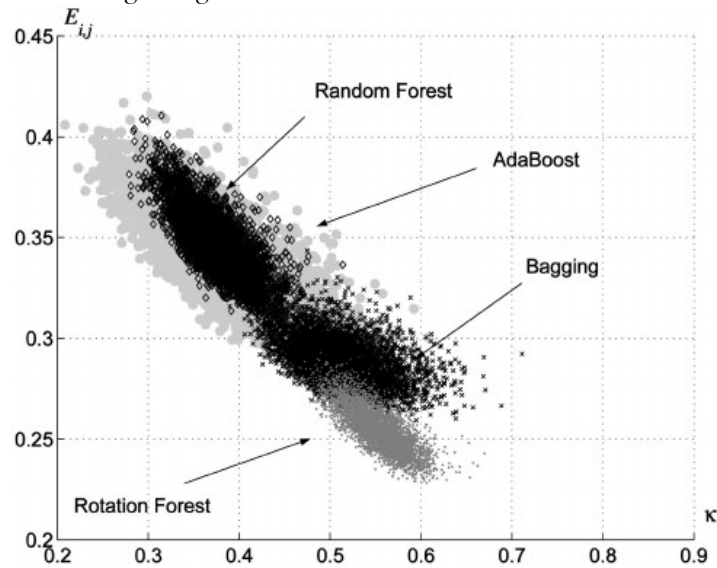
Realice las siguientes tareas:

1. Compare el rendimiento de al menos dos métodos no basados en boosting, como la mezcla por votación, bagging o random forests.
2. Compare el rendimiento de al menos dos métodos basados en boosting, esto incluye las diferentes configuraciones de AdaBoost. Estudie la evolución del error a medida que se añaden clasificadores para alguno de los problemas tal y como se muestra en la siguiente figura:



3. Compare el rendimiento de los métodos de stacking con el mejor de los métodos de cada uno de los apartados anteriores.

4. (*) Implemente un método adicional de ensembles de clasificadores de los muchos presentes en la bibliografía. Por ejemplo: MultiBoost, FeatureBoost, Random subspace methods, Rotation Forest, ReweightBoost, FloatBoost, MadaBoost, Gentle AdaBoost, etc. Compare su rendimiento con los métodos anteriores.
5. (*) Estudie la diversidad de los ensembles contruidos en un mismo problema usando los diagramas kappa-error tal y como se muestran en la figura siguiente:



NOTAS:

- Las comparaciones se pueden hacer globalmente usando tests estadísticos, para lo cual será necesario un mínimo de problemas cercano a 20 o discutiendo los resultados para cada problema elegido.
- Los ejercicios marcados con un * son opcionales.