

### Práctica 3: Clasificación con datos desequilibrados

**Duración:** 3 sesiones

**Objetivo:** El objetivo de esta práctica es introducir los conceptos de clasificación en presencia de datos con desequilibrio de clases

En esta práctica se va a estudiar el funcionamiento de los diferentes métodos de gestión del desequilibrio de clases. Debido a su sencillez es fácil la implementación, aunque también es posible usar métodos ya incluidos en scikit-learn.

Seleccione un número suficiente de problemas de los repositorios usuales (UCI MLR, Weka, etc.). Se deben escoger varios problemas de 2 clases con diferente nivel de desequilibrio de clases. Elija al menos relaciones clase mayoritaria/minoritaria de 2:1.

Realice las siguientes tareas:

1. Elija tres métricas de las estudiadas en teoría para evaluar el rendimiento de los métodos en problemas con desequilibrio de clase. Elija, al menos, dos clasificadores diferentes. En primer lugar vamos a comparar el efecto del desequilibrio de clases en estos dos clasificadores. Estudie el error con las métricas escogidas y obtenga las conclusiones que observe sobre el efecto del desequilibrio.
2. Estudio de las dos técnicas básicas. Implemente, o use, las técnicas de *over-sampling* y *under-sampling*. Aplique estas dos técnicas a los conjuntos de datos seleccionados y estudie cómo se comportan para cada uno de los dos modelos de clasificación. Indique las ventajas e inconvenientes que observa en cada una de ellas.
3. (\*) Implemente el método SMOTE estudiado en teoría y compárelo con los dos anteriores. Evalúe su rendimiento con respecto a los dos métodos del apartado anterior.
4. Compare gráficamente los resultados más significativos de los ejercicios anteriores usando cualquiera de las representaciones gráficas que conozca.

#### 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.