

1 Ejercicio Prácticas Clasificación

Fecha límite de entrega: 18 de junio de 2018 23:59.

Consideremos el dataset **abalone**, con información sobre características de diversas muestras de este gasterópodo. Queremos poder construir un buen clasificador que permita distinguir entre abalones infantiles (sexo = I) y adultos (sexo = M o F) en función del resto (todas o algunas) de las otras covariables explicativas, como por ejemplo longitud, diámetro, peso, etc. Para tal fin siga estas indicaciones:

1. Ensaye modelos de regresión logística, de k-vecinos más cercanos y de LDA y/o QDA. Los métodos a ser utilizados deben ser implementaciones propias, es decir, sin usar las librerías de R.
2. Para elegir algún modelo ganador, en el caso de regresión logística habrá de ensayar con diversas selecciones de covariables, considere las más razonables con algún criterio. Puede, y se recomienda que así se haga, probar con esquemas de training/testing o directamente de K-Fold o Leave-one-out-cross-validation (LOOCV) para elegir modelos ganadores.
3. En K-vecinos más cercanos se habrá de hacer la elección adecuada del k , nuevamente se recomienda un procedimiento como el que se utilice en el inciso anterior. A fin de poder tener comparabilidad, se recomienda usar el mismo esquema con la misma partición que en el inciso anterior. Para evitar complejidad, suponga ya definido un conjunto de covariables explicativas, como por ejemplo la que puede haberse elegido en el inciso anterior.
4. Caso de LDA o QDA es similar al caso de Regresión Logística en cuanto a selección de covariables.
5. Recuerde que los métodos de Regresión Logística y de K-vecinos más cercanos podrían requerir datos estandarizados.
6. Para obtener el resultado final del modelo ganador recuerde separar de entrada un conjunto de “validación” para poder evaluar finalmente, con el mejor modelo, la calidad de clasificación mediante una matriz de confusión y alguna otra métrica que considere adecuada (tasa de acierto, precisión, recall, etc).