

El algoritmo general para construir un bosque aleatorio de clasificación es el siguiente:

1. Desde $b=1$ hasta B :
 - a. Obtener una muestra con bootstrap Z^* de tamaño N del conjunto de datos de entrenamiento.
 - b. Ajustar un árbol T_b a los datos muestreados repitiendo recursivamente los siguientes pasos en cada nodo terminal hasta llegar al tamaño mínimo n_{min} del nodo:
 - i. Seleccionar aleatoriamente m de las p variables disponibles.
 - ii. Seleccionar la variable que mejor corte/partea de entre las m .
 - iii. Partir el nodo terminal utilizando dicha variable.
2. Construir el conjunto de árboles $\{T_b\}_1^B$.
3. Sea $\hat{C}_b(x)$ la predicción sobre clase del b -ésimo árbol. Entonces:

$$\hat{C}_{bosque}^B(x) = \text{voto mayoritario} \{ \hat{C}_b(x) \}_1^B$$

En un bosque aleatorio lo que se hace es generar una cantidad grande de árboles, cada uno de estos árboles se calcula sobre una muestra diferente de datos, de igual manera la selección de las variables de selección en cada nodo se hace sobre un conjunto aleatorio de estas. Esto se llama bagging, y su fin es lograr plasmar las interacciones no lineales que ocurren entre las variables, además de que con esto se disminuye la varianza del modelo colectivo.

El estimador obtenido (la probabilidad) se calcula utilizando bagging sobre los votos emitidos de cada árbol.

Para este ejercicio se ajustó un bosque aleatorio con 10,000 árboles de clasificación, ajustados sobre 10,000 muestras independientes de 10,000 elementos. Las variables predictoras fueron los saldos promedio en captación de los últimos 6 meses. La variable objetivo fue el tener o no un crédito.

La matriz de confusión del modelo queda como sigue:

Real \ Predicción	no tiene	tiene	error de la categoría
no tiene	77.11%	22.89%	23.28%
tiene	6.65%	93.35%	6.81%

Es decir, casi 77% de las veces el modelo predijo no tener crédito de forma correcta, y más del 93% de las veces predijo tener crédito de forma correcta.

El error OOB (out of bag) del modelo es del 12.33%.

Bibliografía

- G. Biau. Consistency of random forests and other averaging classifiers, *Journal of Machine Learning Research* 9. 2008
- L. Breiman. Clasification and Regression Trees, *Chapman & Hall*. 1983
- T. Hastie et al. The Elements of Statistical Learning, *Springer*. 2008