

Un problema muy común es intentar predecir el comportamiento de una variable dependiente Y a partir de un conjunto de p variables independientes $X \in R^p$

A partir de una muestra de observaciones $(\vec{x}_i, y_i)$, $i = 1, \dots, n$ se busca una regla de clasificación, RC mediante regresión logística, árboles de clasificación, etc.

Si se conociera la distribución conjunta de $(\vec{X}, Y)$, es decir,

$$P[(\vec{X}, Y) = (\vec{x}, y)]$$

el procedimiento más eficiente sería utilizar como regla de predicción la probabilidad a posteriori

$$RC(\vec{x}) = \operatorname{argmax}_y P(Y = y | \vec{x}) = \operatorname{argmax}_y P[(\vec{X}, Y) = (\vec{x}, y)]$$

El problema es que la distribución conjunta suele ser desconocida y es necesario estimarla.

Bagging. Bootstrap

El método **Bootstrap** se basa en sustituir la Función de distribución teórica $F_{(X,Y)}$ por la función de distribución empírica $\hat{F}(X, Y)$ definida por

$$\hat{F}(X, Y) = \begin{cases} \frac{1}{n} & \text{si } (\vec{x}, y) = (\vec{x}_i, y_i) \text{ para algún } i \\ 0 & \text{en otro caso} \end{cases}$$

(Teorma de Glivenko Catelli)

Una muestra bootstrap de tamaño n $(\vec{x}_i^*, y_i^*)$, se obtiene eligiendo al azar y con reemplazamiento n observaciones de la muestra original $(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n)$

Así se tiene una idea más clara de lo que ocurriría si se pudiera repetir varias veces el proceso de muestreo inicial

Bagging

A partir de una muestra $(\vec{x}_i, y_i)$, $i = 1, \dots, n$ el método **bagging** utiliza el siguiente algoritmo

- ▶ Se eligen b muestras bootstrap $(\vec{x}_i^{*b}, y_i^{*b}, b = 1, \dots, B)$
- ▶ Se calcula una regla de clasificación RC^b basada en la correspondiente muestra bootstrap b -ésima.
- ▶ Se aplica la regla de clasificación para una de esas B muestras, $RC^b(\vec{x}_i^{*b})$

La clasificación final de cada individuo $\vec{x}$ se hace como promedio de la B predicciones

$$RC_{bag}(\vec{x}) = \underset{k}{\operatorname{argmax}} \sum_{b=1}^B I(RC^b(\vec{x}) = k)$$

es decir, se asigna a la categoría más elegida.

La estimación de la probabilidad de pertenencia a un categoría de k , se debe basar en las probabilidades de clasificación obtenida en cada una de las B réplicas realizadas

$$\hat{P}_{bag}(\vec{x} \in k) = \frac{1}{B} \sum_{b=1}^B \hat{P}^b(\vec{x} \in k)$$

Esta estimación sugiere asignar las observaciones a clases empleando el siguiente criterio:

$$RC_{bag}(\vec{x}) = \operatorname{argmax}_k \{ \hat{P}_{bag}(\vec{x} \in k) \}$$

EL bagging se basa en la idea de disminuir la tasa de error promediando muchos criterios de clasificación.

DESVENTAJAS

- ▶ No se debe emplear cuando los clasificadores individuales funcionan mal
- ▶ La interpretación del criterio de clasificación suele ser mucho más difícil.
- ▶ Requiere muchos más cálculos