

*El objetivo de esta práctica es la implementación de reglas de clasificación teniendo en cuenta la selección de los parámetros de suviado.*

En un bosque de Bariloche hay dos variedades de hongos, que identificaremos como la variedad I y variedad II. En el archivo hongos\_clasificados.txt encontrará  $n = 500$  registros correspondientes a la altura y variedad de cada uno de los hongos examinados. A fin de clasificar un nuevo hongo de este bosque, implementaremos la regla de Bayes, pero sin suponer que las densidades condicionales involucradas en su cálculo pertenecen a una familia determinada. En lo que sigue, llamaremos  $f_1$  a la densidad de la altura de un hongo de la variedad I y  $f_0$  a la densidad de la altura de un hongo de la variedad II.

1. A partir de los alturas medidas en los hongos de variedad I estime la función de densidad  $f_1$ . Indique cómo determinó la ventana y que núcleo usó. Llamemos  $\hat{f}_{1,h_1}$  a la estimación resultante de la función de densidad  $f_1$ .
2. A partir de los alturas medidas en los hongos de variedad II estime la función de densidad  $f_0$ . Indique cómo determinó la ventana y que núcleo usó. Llamemos  $\hat{f}_{0,h_0}$  a la estimación resultante de la función de densidad  $f_0$ .
3. Implemente una función **class.est.variedad** que determine la variedad de un hongo mediante la regla plug-in de Bayes  $\hat{g}$  basada en las estimaciones  $\hat{f}_{1,h_1}$  y  $\hat{f}_{0,h_0}$  ya obtenidas en los dos ítems anteriores y las proporciones de cada variedad en los datos registrados en el archivo hongos\_clasificados.txt.
4. Calcule el Error de Clasificación Empírico de  $\hat{g}$  utilizando los datos del archivo hongos\_clasificados.txt.
5. ¿Le parece que las ventanas halladas en a) y b) con las que implementó la regla de clasificación son las más adecuadas a los fines de la clasificación?

Implemente una función **class.optim.est.variedad** que determine la variedad de un hongo mediante la regla plug-in de Bayes  $\hat{g}$  basada en las estimaciones  $\hat{f}_{1,h_1}$  y  $\hat{f}_{0,h_0}$  y las proporciones de cada variedad en los datos registrados en el archivo hongos\_clasificados.txt en la que las ventanas se determinan simultáneamente por Convalidación Cruzada.

6. Estime el Error de Clasificación de la regla de plug-in Bayes  $\hat{g}$  mediante el Error de Clasificación Empírico utilizando los datos del archivo hongos\_clasificados.txt, pero ahora implemente la regla asumiendo que las densidades  $f_1$  y  $f_0$  son normales y que desconoce sus parámetros. Compare con los resultados anteriores.

7. **Para entregar:** Implemente una función **class.nopar** que dado un punto  $x_{new}$  determine la clase a la que pertenece el nuevo individuo que tiene este valor en la covariable mediante la regla plug-in de Bayes  $\hat{g}$  basada en las estimaciones no paramétricas de las densidades  $f_1$  y  $f_0$  usando núcleo gaussiano. Para ello entrar como input de la función implementada el punto  $x_{new}$ , los vectores de datos  $X_{datos}$  e  $Y_{datos}$  y las ventanas  $h_1$  y  $h_0$ : `class.nopar( $x_{new}$ ,  $X_{datos}$ ,  $Y_{datos}$ ,  $h_1$ ,  $h_0$ )`.