

Propuesta de TFG

Amparo Bañlo

2017-2018

Clasificación supervisada

La *clasificación supervisada*, también denominada *análisis discriminante* o *reconocimiento de patrones*, tiene como objetivo estudiar los rasgos diferenciales de un vector de variables de interés en varias poblaciones conocidas. El término “supervisada” se refiere a la existencia de una muestra del vector (*training sample*) que está correctamente clasificada en la población correspondiente. En base a esta muestra, se intenta encontrar “discriminantes” (funciones del vector) cuyos valores numéricos estén lo más separados posible en poblaciones diferentes. Estos discriminantes servirán posteriormente para clasificar una nueva observación en alguna de dichas poblaciones.

El objetivo del trabajo es presentar el problema matemático de la clasificación supervisada y hacer una revisión de las técnicas más habituales: discriminación lineal, k -vecinos más próximos, árboles de clasificación, ... Al mismo tiempo se ilustrará el uso de cada una de estas técnicas en la práctica mediante el análisis de un conjunto de datos reales con el programa estadístico R.

Referencias

Devroye, L., Györfi, L. and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer.