

FACULTAD DE CIENCIAS ACTUARIALES

Cuevas, C., & Riccomagno, E. (2017). Reduction of Dimensionality for Classification. Archives of Data Science (On line first) KIT Scientific Publishing, 2(1).

Resumen. En el presente artículo se presenta un algoritmo para la reducción de la dimensionalidad que resulta útil en problemas de clasificación estadística donde se discriminan observaciones para dos distribuciones normales multivariadas. Se basa en el Análisis de Componentes Principales y consiste en una diagonalización simultánea de dos matrices de covarianza. El criterio de reducción de la dimensionalidad resulta de la contribución de cada componente principal al área bajo la curva ROC de una función discriminante. Se consideran las puntuaciones lineal y cuadrática, centrándose en el caso cuadrático.

Abstract. We present an algorithm for the reduction of dimensionality useful in statistical classification problems where observations for two multivariate normal distributions are discriminated. It is based on Principal Components Analysis and consists of a simultaneous diagonalization of two covariance matrices. The criterion for reduction of dimensionality is given by the contribution of each principal component to the area under the ROC curve of a discriminant function. Linear and quadratic scores are considered, the focus being on the quadratic case.