Covariance and PCA for Categorical Variables

Hirotaka Niitsuma and Takashi Okada

Abstract: Covariances from categorical variables are defined using a regular simplex expression for categories. The method follows the variance definition by Gini, and it gives the covariance as a solution of simultaneous equations using the Newton method. The calculated results give reasonable values for test data. A method of principal component analysis (RS-PCA) is also proposed using regular simplex expressions, which allows easy interpretation of the principal components.