Multivariate Analysis of Breast Cancer Prediction Parameters

Benjamin Schleich, Hema Sudarsanam, and Chanchal Saha

Department of Systems Science and Industrial Engineering
State University of New York at Binghamton
Binghamton, New York 13902

Corresponding author's Email: bschlei1@binghamton.edu

Abstract: This study explores breast cancer data with multivariate techniques. Breast cancer is the second leading disease that causes death, 100 deaths per day, of women living in the USA. Therefore, this study aims in identifying the independent decision parameters given in the data-set which are accounted for correct prediction of breast cancer. Logistic Regression Analysis (LRA) is performed to assess the correct diagnosis probability. A Principal Component Analysis (PCA) is also performed to determine the number of decision parameters can be reduced. The experimental results indicate that the LRA has a very high correct validation while the PCA suggested that the principal components can be significantly reduced. In future, Factor Analysis (FA) can be conducted to find the latent relations among the variables, which might give a better understanding of the correlation among the variables.