

FEATURE SELECTION USING MATRIX CORRELATIONS AND ITS APPLICATIONS IN AGRICULTURE

B. K. Hooda¹ & Ekta Hooda²

¹Professor, Department of Mathematics & Statistics, CCS Haryana Agricultural University, Hisar, India

²Research Associate, Directorate of Extension Education, CCS Haryana Agricultural University, Hisar, India

ABSTRACT

Dimensionality reduction techniques are broadly categorized as feature extraction and feature selection. Feature extraction techniques select features in the transformed space while feature selection techniques consist of finding a subset of original features or variables that is optimal for a given criterion for adequate representation of the whole data. Principal Component Analysis (PCA) is often the most common choice for reducing dimensionality of multivariate data through feature extraction. However, dimensionality reduction using PCA does not provide a real reduction of dimensionality in terms of the original variables, since all of the original variables are used in projection to the lower dimensional. Several criteria have been proposed for selecting the best subset of features which can preserve the structure and variation of the original data. However, little is known about the applications feature selection techniques in agricultural and biological research where many measurements are taken on each individual. In the present study, applicability of matrix correlation based feature selection techniques has been examined for identification of informative and redundant features in wheat data. RV-coefficient (Robert and Escoffier, 1976) and Yanai's Generalized Coefficient of Determination (Ramsay et al., (1984) have been used to measure the similarity between two data matrices. Subsets selected using different criteria have been compared in terms of the measure of overall predictive efficiency. For identification of important features, secondary data of 67 wheat genotypes recorded for 14 characters have been used. Models built with subset of best features are expected not only to reduce the model complexity but also require less time and resources.

KEYWORDS: Feature Selection, Feature Extraction, Dimensionality, PCA, Matrix Correlation, Agriculture

Article History

Received: 12 Jan 2021 | Revised: 19 Jan 2021 | Accepted: 04 Feb 2021
