Abstract

In classification analysis, the dependent variable is frequently influenced not only by ratio scale variables, but also by qualitative (nominal scale) variables. Machine Learning algorithms accept only numerical inputs, hence, it is necessary to encode these categorical variables into numerical values using encoding techniques.

This paper presents a comparative study of seven categorical variable encoding techniques to be used for classification using Artificial Neural Networks on a categorical dataset. The Car Evaluation dataset provided by UCI is used for training. Results show that the data encoded with Sum Coding and Backward Difference Coding technique give highest accuracy as compared to the data pre-processed by rest of the techniques.

References

A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers


Index Terms
Computer Science Artificial Intelligence

Keywords
Machine Learning, Statistical Learning, Artificial Neural Networks, Data Preprocessing.