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

July 2017, from www.mymarketresearchmethods.com/types-of-data-
nominal-ordinal-interval-ratio/.
Forecasting Indian Index of Industrial Production”, Proceedings of the IEEE Tensymp 2017,
Kochi, India
whatis.techtarget.com/definition/machine-learning.
psych.colorado.edu/çarey/courses/psyc5741/handouts/Coding%
20Categorical%20Variables%202006-03-03.pdf.
978-1782162148.
8. Von Eye, Alexander, and Clifford C. Clogg, eds. “Categorical variables in developmental
research: Methods of analysis.” Elsevier, 1996.
2017, from stats.idre.ucla.edu/r/library/r-library-contrast-coding-
systems-for-categorical-variables/
10. Saravanan K and S. Sasithra, “REVIEW ON CLASSIFICATION BASED ON ARTIFICIAL
NEURAL NETWORKS”, International Journal of Ambient Systems and Applications (IJASA)
11. M. Bohanec and V. Rajkovic, “Knowledge acquisition and explanation for multi-attribute
decision making.” In 8th Intl. Workshop on Expert Systems and their Applications, Avignon,
Irvine, CA: University of California, School of Information and Computer Science.

Index Terms

Computer Science       Artificial Intelligence

Keywords

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