Machine Learning

Unit I

Factor analysis

Canul Cal Jorge

Herrera Tuz Alan

González González Julio

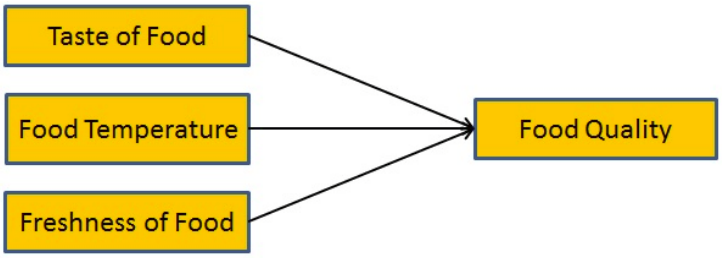
Tzuc Zih Ricardo

Merida, Yucatan

September 15th, 2023

Factor Analysis

The factor analysis technique extracts the maximum common variance from all the variables and puts them into a common score. It is a theory that is used in training the machine learning model and so it is quite related to data mining. The belief behind factor analytic techniques is that the information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset.



Characteristics.

- Reduce a large number of variables into fewer numbers of factors.
- Puts maximum common variance into a common score.
- Associates multiple observed variables with a latent variable.
- As the same numbers of factors and variables,where each factor contains a certain amount of overall variance.
- It is also used to find the fundamental structure of a huge set of variables.
- The main advantage of using FA instead of PCA is that the outputs are much easier to interpret.

Benefits.

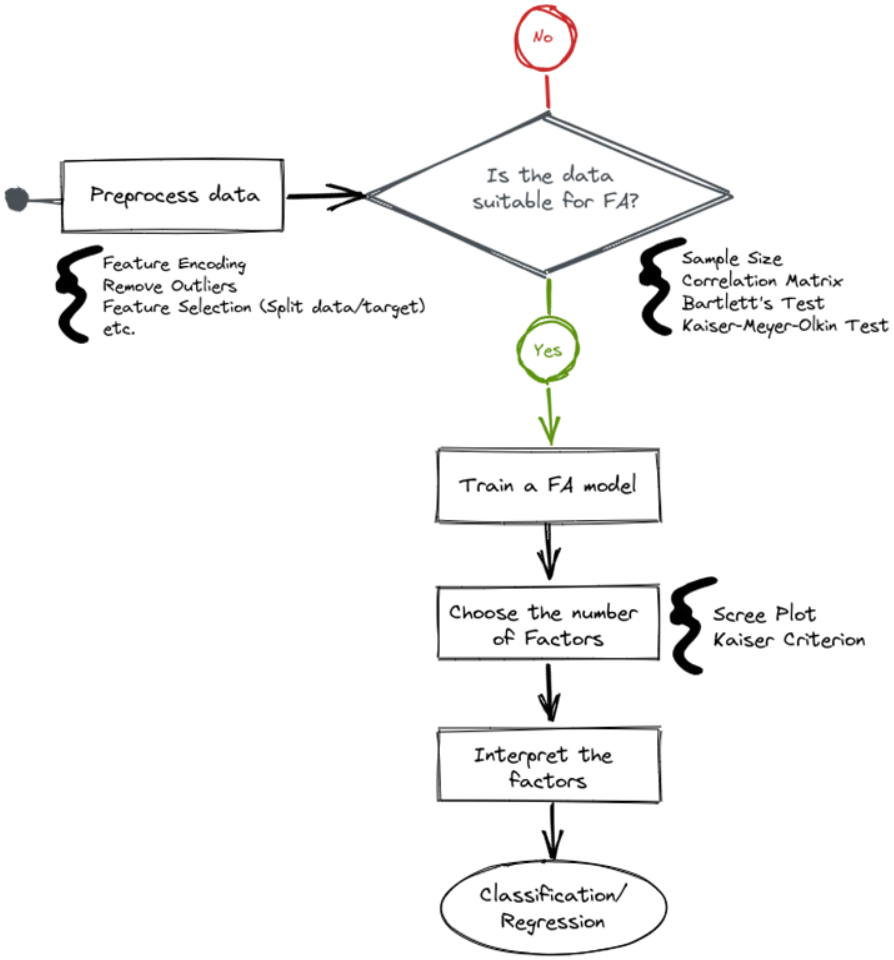
- The process provides a more concise and easily understandable representation of the data, removing unnecessary complexity. Factor analysis enhances interpretability and enables clearer insights into the underlying structure and relationship within the dataset.
- Capture the most impactful factors, thus enhancing the predictive capabilities of machine learning models.

- It can be utilized to determine hidden dimensions or limitations, which may or may not be obvious via direct analysis.

Example:

As an example, the process described is about a Student Performance Dataset, interpreting the output factors and using them in a classification task to predict student grades.

- The dataset originally has 33 variables and about 395 students from two Portuguese schools in their Math class. The features include student grades, demographic, social, and school-related information. We have information about the first (G1), second (G2), and final grade (G3), but we'll try to predict G3 without using G1 and G2 since these variables are highly correlated with G3, and it is not useful to use G1 and G2 as we want to grasp exclusively the relationship of the other variables with G3.(Henrique, 2022)



References

Henrique, A. (2022). Dimensionality reduction with factor analysis on student performance data. Medium. https://medium.com/geekculture/dimensionality-reduction-with-factor-analysis-on-student-performance-data-fd4ca7082f63

Factor Analysis in Machine Learning. (s/f). Www.javatpoint.com. Retrieved from: https://www.javatpoint.com/factor-analysis-in-machine-learning

Follow, D. (2020, enero 10). Introduction to factor analytics. GeeksforGeeks. https://www.geeksforgeeks.org/introduction-to-factor-analytics/

Upgrad Team (2023). What's Factor Analysis and its Types? Upgrad. Retrieved from: https://www.upgrad.com/us/blog/what-is-factor-analysis-and-its-types/