Credit Task 1

Submission requirements: Submit a report (pdf format) and the code file (.ipynb), which is testable by the marker.

Datafile: SCADI.csv

Data Description: This dataset contains 206 attributes of 70 children with physical and motor disability based on ICF-CY. For more information click this link.

1. Determine the number of subgroups from the dataset using attributes 3 to 205 i.e., exclude attributes 1, 2 and 206. Is this number the same as the number of classes presented by attribute 206? Explain and justify your findings.

2. Is this data facing the curse of dimensionality? If so, then how to solve this problem. Explain with a two-dimensional plot and report relevant loss of information.

3. After applying principal component analysis (PCA) on a given dataset, it was found that the percentage of variance for the first N components is X%. How is this percentage of variance computed?

***Dataset filename*:** obesity_levels.csv

Dataset description: This dataset includes data for the estimation of obesity levels in individuals based on their eating habits and physical condition. The data contains 17 attributes and 2111 records.

Features and labels: The attribute names are listed below. The description of the attributes can be found in this article (web-link).

4. Create a machine learning (ML) model for predicting "weight" using all features except "NObeyesdad" and report observed performance. Explain your results based on following criteria:

a. What model have you selected for solving this problem and why?
b. Have you made any assumption for the target variable? If so, then why?
c. What have you done with text variables? Explain.
d. Have you optimized any model parameters? What is the benefit of this action?
e. Have you applied any steps for handling overfitting or underfitting issues? What is that?

Assessment feedback

The results with comments will be released within 5 business days from the due date.

Referencing

You must correctly use the Harvard method in this assessment. See the Deakin referencing guide.

Academic integrity, plagiarism and collusion

Plagiarism and collusion constitute extremely serious breaches of academic integrity. They are forms of cheating, and severe penalties are associated with them, including cancellation of marks for a specific assignment, for a specific unit or even exclusion from the course. If you are ever in doubt about how to properly use and cite a source of information refer to the referencing site above.

Plagiarism occurs when a student passes off as the student's own work, or copies without acknowledgement as to its authorship, the work of any other person or resubmits their own work from a previous assessment task.

Collusion occurs when a student obtains the agreement of another person for a fraudulent purpose, with the intent of obtaining an advantage in submitting an assignment or other work.

Work submitted may be reproduced and/or communicated by the university for the purpose of assuring academic integrity of submissions: https://www.deakin.edu.au/students/study-support/referencing/academic-integrity.