Marked Tutorial 1 (5 marks)
Soft Computing (7168) & Soft Computing PG (7197)
Semester 2- 2024

Overview

This assessment needs to be done individually. GenAI or opensource code can only be used to assist with secondary tasks (e.g., reading file, logging & saving output) and need to be referenced as mentioned in Lecture 1. The datasets might need pre-processing. Inspecting the data or plotting graphs for it can help with this.

Task 1 Credit Card Fraud Detection

It is important that credit card companies can recognise fraudulent credit card transactions so that customers are not charged for items that they did not purchase. The "card.csv" dataset contains credit card transactions made by European cardholders over two days. For privacy reasons, the dataset contains only input variables which are the result of a PCA transformation. Features V1, V2,… V28 are the principal components obtained; the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount.
Dataset columns:
- A: time
- B to AC: V1 to V28
- AD: amount
- AE: target class (fraud vs normal)

Using this dataset, train a neural network by splitting the data into 70% for training, 10% for validation, and 20% for testing. Try different network architectures (e.g., number of neurons & number of hidden layers) and training parameters (e.g., learning rate & batch size). Fill in the table based on the results of your experiments.

	Network architecture	Training parameters	Training accuracy	Test accuracy
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
Key findings (In a maximum of 150 words, summarise the key findings from this experiment.)				

Task 2 – Abalone age estimation

The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope -- a boring and time-consuming task. Other measurements, which are easier to obtain, are used to predict the age. Further information, such as weather patterns and location (hence food availability) may be required to solve the problem. Use the Kaggle abalone dataset (Abalone Dataset (kaggle.com)) to train a neural network by splitting the data into 70% for training, 10% for validation, and 20% for testing. Try different network architectures (e.g., number of neurons & number of hidden layers) training parameters (e.g., learning rate & batch size), and input features. Fill in the table based on the results of your experiments.

	Network architecture	Training parameters	Features used	Training MSE	Test MSE
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
Key findings (In a maximum of 150 words, summarise the key findings from this experiment.)					

Deliverables & Submission

This marked tutorial needs to be submitted by the deadline indicated on Canvas. You need to:
- Upload a .docx or .pdf file which contains the results of the two tasks. Use the tables provided above to fill in the results. Don't include extra data
- Upload the source code used for generating the results
- Attend the tutorial session on Week 4 to discuss your submission with the tutor