Lesson 9 Assignment

Submit Assignment

Due Sunday by 11:59pm **Points** 100 **Submitting** a file upload
Available Oct 8 at 12am - Oct 28 at 11:59pm 21 days

For this assignment, you are going to use **Parkinsons Telemonitoring Data Set (https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring)** .

Data Set Information

This dataset is composed of a range of biomedical voice measurements from 42 people with early-stage Parkinson's disease recruited to a six-month trial of a telemonitoring device for remote symptom progression monitoring. The recordings were automatically captured in the patient's homes.

Columns in the table contain subject number, subject age, subject gender, time interval from baseline recruitment date, motor UPDRS, total UPDRS, and 16 biomedical voice measures.

Each row corresponds to one of 5,875 voice recording from these individuals.

The main aim of the data is to predict the motor and total UPDRS scores ('motor_UPDRS' and 'total_UPDRS') from the 16 voice measures.

1. Perform an exploratory analysis on the data and Remove motor_UPDRS column (10 points)
2. Use cross-validation to build a linear regression model to predict total_UPDRS (25 points)
3. Use cross-validation to build a regression tree model to predict total_UPDRS (25 points)
4. Use cross-validation to build a neural network model to predict total_UPDRS (25 points)
5. Which model has better performance? Is there any way to improve the model? (5 points)
6. Try to optimize the tree model or neural network model (Choose one). (10 points)

Once completed, submit a Word or Pdf file to this assignment.