

Supervised Learning Project

Data Description:

The data consists of Biomechanical features which are to be used to classify patients according to their labels

Domain:

Healthcare/Medicine

Context:

The data have been organized in two different but related classification tasks.

3C_weka.csv (file with three class labels)

The first task consists in classifying patients as belonging to one out of three categories: Normal (100 patients), Disk Hernia (60 patients) or Spondylolisthesis (150 patients).

2C_weka.csv (file with two class labels)

For the second task, the categories Disk Hernia and Spondylolisthesis were merged into a single category labelled as 'abnormal'. Thus, the second task consists in classifying patients as belonging to one out of two categories: Normal (100 patients) or Abnormal (210 patients).

Attribute Information:

Each patient is represented in the data set by six biomechanical attributes derived from the shape and orientation of the pelvis and lumbar spine (each one is a column):

- pelvic incidence
- pelvic tilt
- lumbar lordosis angle
- sacral slope
- pelvic radius
- grade of spondylolisthesis

Learning Outcomes:

- Exploratory Data Analysis
- Application of Supervised Learning algorithm

Objective:

Demonstrate the ability to fetch, process and leverage data to generate useful predictions by training Supervised Learning algorithms.

Steps and tasks:

1. Import the necessary libraries (1 mark)
2. Import data 2C_weka.csv for 2 Class Classification. (1 mark)
3. Explore the dataset and mention the insights (4 marks)
4. Encode the Class variable to integer (2 marks)
5. Create the X(Feature-set) and Y(Target-set) sets for your Data. (2 marks)
6. Normalize your Data (X) to get values between 0 to 1 (2 marks)
7. Split the data as a train and test with a ratio of 70:30. (1 marks)
8. Train k-NN algorithm and optionally one other classification algorithm of your choice using the train data and measure their accuracy scores on the test data (5 marks)
9. What is the best K value for your classification? Find at which k value you get the best score (2 marks)
10. Import data 3C_weka.csv for 3 Class Classification and repeat all the steps which are given above (10 marks)