Aim:- Loading dataset into keras/ pytorch;
creating training & testing split.

Requirements:- Dataset, jupyter Notebook, libraries.
Pandas, Numpy, scikit-learn.

Theory:-

- Deep learning
- Deep learning is a subset of Machine learning
that uses artificial neural networks to mimic
the human brain and solve complex problems.
widely used in areas like image recognition,
natural language processing, autonomous
vehicles.
- It consists of multiple layers
- Input layer- Accept raw data
- Hidden layer- Extract meaningful patterns from data.
- Output layer- Provides final prediction or result.

Dataset.
Dataset is a collection of data organized in
tabular form, typically consisting of rows &
columns where in each rows represents a
single data instance or observation, while each
column represents a different feature of the data.

Importance of splitting the dataset.

- In DL, it is essential to split the dataset into separate parts to ensure the model's performance is measured accurately. Without splitting the data, the model may memorize the entire dataset instead of learning general patterns, leading to overfitting.
- The data is generally split into.

(1) Training data - This is the portion of the dataset used to train the model. Where 80% of dataset is used to train the model.

(2) Testing data - This part of the dataset is kept separate and only used only after the training this is complete.
- It evaluates the model's performance or unseen data.
- Where 20% of dataset is used for testing.

Splitting the dataset allows the model to generalize well and accurately predict on new data.

Conclusion :-

Splitting data into training validation, and test sets is essential in deep learning to prevent overfitting and ensure that models generalize well to unseen data.

10/3/25