

CST 476-2 Deep Learning

Lab Sheet 01

Activity: Exploring the Pima Indians Diabetes Dataset with Keras

Aim:

The aim of this lab sheet is to introduce building, training, and evaluating a simple neural network using the Keras library. In here, you will work with the Pima Indians Diabetes dataset to predict the onset of diabetes based on various health-related features.

Dataset:

Pima Indians Diabetes Dataset

Description of the Dataset:

The Pima Indians Diabetes dataset is a collection of health-related data points, primarily focusing on individuals of Pima Indian heritage. Comprising nine columns, the dataset includes essential attributes such as the number of pregnancies, plasma glucose concentration, diastolic blood pressure, triceps skinfold thickness, 2-hour serum insulin levels, body mass index (BMI), a diabetes pedigree function assessing familial diabetes likelihood, age, and the presence or absence of diabetes indicated by a binary outcome variable (0 or 1). These variables provide a comprehensive insight into the health profile of individuals, enabling the exploration of potential relationships and patterns related to the onset of diabetes.

Tasks:

1. Download the Pima Indians Diabetes dataset from the VLE.
2. Open your Jupyter Notebook environment and import the necessary libraries.
3. Load the Pima Indians Diabetes dataset into your Jupyter Notebook environment.
4. Review the dataset structure and characteristics.
5. Perform data preprocessing by splitting the dataset into training and testing sets.
6. Standardize the features using the StandardScaler from scikit-learn.
7. Build a neural network model using Keras.
8. Compile the neural network model by specifying the optimizer and loss function.
9. Choose appropriate metrics for model evaluation.
10. Train the compiled model using the training dataset.
11. Monitor training progress, including loss and accuracy metrics.
12. Evaluate the trained model's performance on the test set.
13. Make predictions for a given set of input features using the trained model.
14. Interpret the model's output and understand the significance of predictions.