DEPENDABLE ARTIFICIAL INTELLIGENCE

[CSL7370]

REPORT: ASSIGNMENT 1 (BIAS)

Tejas Gaikwad

MT19AI021
Indian Institute of Technology, Jodhpur

INTRODUCTION

The assignment focuses on bias in Artificial Intelligence. The sources of bias, detection, and mitigation are the key points covered under this assignment. The assignment is divided into 2 parts, the first one focuses on evaluation metrics and detection of bias in the system. This is performed on MNIST data, and bias detection is done for ‘1’ and ‘7’ in the dataset. The former one focuses on neural network performance for detecting the bias and mitigating the bias via Data addition and multitasking techniques.

MATERIALS

1. Data Sets
    a. http://yann.lecun.com/exdb/mnist/
    b. https://drive.google.com/file/d/1WYb3Xonb52ZPOpyN58t0WTQPXUnwDg0U/view?usp=sharing
1. Google Colab
2. Libraries
    a. Sklearn
    b. Numpy
    c. Matplotlib
    d. OpenCV

PROCEDURE

1. Question 1
    a. access MNIST Dataset and create Training-Testing Dataset(randomizing data)
    b. Segregate 1’s(6000) and 7’s(500) data and labels
    c. Take 3 different 7’s samples of 500
    d. Train SVM and Neural Network Model for 3 different training sets
    e. Test respective models and find confusion matrix and Prediction probability
    f. Find Mean Accuracy and standard deviation
    g. Plot ROC and Precision-Recall for different training sets

2. Question 2

    a. Access the images from provided folders, create training and testing dataset from folders (1) “.\specs_train”, (2) “.\specs_test”, (3) “.\nonSpecs_train”, (4) .\nonSpecs_test”, and (5) “.\data”
    b. Resizing images to 32*32
    c. Training neural network of architecture [128 -- 128 -- 128 -- 64 -- 1], activation function ‘sigmoid’ used