**Group Id-3**

**Siva Teja Yadav - rr4899**

**Roja Balakrishnan - Pi7973**

**CS697A – Topic in Computer Science – Machine Learning**

**Summer 2021**

**Assignment 3 (10 points)**

**Q1 [1pts]** Create the following **training datasets** from the optdigits.tra set:

X25: Randomly chosen N=25 instances from each class.

X100: Randomly chosen N=100 instances from each class.

**Q2 [2pts]:** Decision Trees, classification:

Use library sklearn.tree.**DecisionTreeClassifier**

algorithm. For the DecisionTreeClassifier determine the value of the **tree depth** parameter (experiment with depth=2, 3, 5, 10) that results in the best test error. Report the training and test errors for each depth value and the training set. How does the best depth value change as the number of instances change?

**For X25:**

Training Error for depth-2: 0.628

Test Error for depth-2: 0.6549805230940456

Training Error for depth-3: 0.384

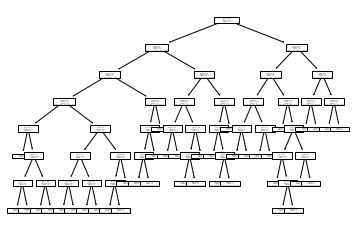
Test Error for depth-3: 0.4852531997774068

Training Error for depth-5: 0.09199999999999997

Test Error for depth-5: 0.34668892598775736

Training Error for depth-10: 0.0

Test Error for depth-10: 0.320534223706177



**For X100:**

Training Error for depth-2: 0.637

Test Error for depth-2: 0.6638842515303283

Training Error for depth-3: 0.389

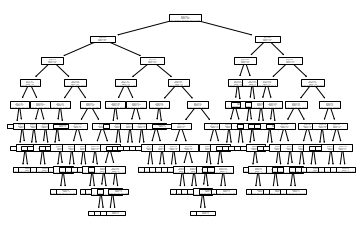
Test Error for depth-3: 0.4718976071229828

Training Error for depth-5: 0.118

Test Error for depth-5: 0.23483583750695602

Training Error for depth-10: 0.0

Test Error for depth-10: 0.21368948247078468

****

### X100 test set with max\_depth 10 has the best depth value as the number of instances change from 25 to 100, the error rate gets minimal

**Q2 [2pts]:** Repeat Q1 using XGBoost and LightGBM algorithms. Experiment with at least 5 different parameter settings to see their effect on training and test errors. How do best parameters change as the #training instances change?

**XGB for X25**

Training Error for depth-2: 0.06399999999999995

Test Error for depth-2: 0.25375626043405675

Training Error for depth-3: 0.0

Test Error for depth-3: 0.16304952698942687

Training Error for depth-5: 0.0

Test Error for depth-5: 0.12687813021702843

Training Error for depth-7: 0.0

Test Error for depth-7: 0.13244296048970505

Training Error for depth-9: 0.0

Test Error for depth-9: 0.13522537562604342

**XGB for X100**

Training Error for depth-2: 0.08599999999999997

Test Error for depth-2: 0.18475236505286585

Training Error for depth-3: 0.0010000000000000009

Test Error for depth-3: 0.09515859766277124

Training Error for depth-5: 0.0

Test Error for depth-5: 0.07178631051752926

Training Error for depth-10: 0.0

Test Error for depth-10: 0.07289927657206452

## Light GBM for X25

Training Error for depth-2: 0.03600000000000003

Test Error for depth-2: 0.19254312743461321

Training Error for depth-3: 0.0

Test Error for depth-3: 0.1279910962715637

Training Error for depth-5: 0.0

Test Error for depth-5: 0.1096271563717307

Training Error for depth-10: 0.0

Test Error for depth-10: 0.1096271563717307

## Light GBM for X100

Training Error for depth-2: 0.07199999999999995

Test Error for depth-2: 0.16917084028937113

Training Error for depth-3: 0.0

Test Error for depth-3: 0.06956037840845852

Training Error for depth-5: 0.0

Test Error for depth-5: 0.05954368391764053

Training Error for depth-10: 0.0

Test Error for depth-10: 0.06121313299944353

**Q4 [2pts]:** Multilayer Perceptrons: classification

Use library sklearn.neural\_network.**MLPClassifier**.

For the neural network (MLP) determine the value of the best **hidden\_layer\_sizes** (experiment with 1 and 2 hidden layers and 2, 5, 10 hidden units in each layer) that results in the best test error for each of the training data sets you created.

## MLP for X25

Training Error for hidden\_layers-2: 0.732

Test Error for depth-2: 0.7607122982749026

Training Error for hidden\_layers-3: 0.18400000000000005

Test Error for depth-3: 0.38341680578742343

Training Error for hidden\_layers-5: 0.02400000000000002

Test Error for depth-5: 0.3116304952698943

Training Error for hidden\_layers-10: 0.0

Test Error for depth-10: 0.1886477462437396

## MLP for X100

Training Error for hidden\_layers-2: 0.897

Test Error for depth-2: 0.899276572064552

Training Error for hidden\_layers-3: 0.30300000000000005

Test Error for depth-3: 0.3528102392877017

Training Error for hidden\_layers-5: 0.0050000000000000044

Test Error for depth-5: 0.12353923205342232

Training Error for hidden\_layers-10: 0.0010000000000000009

Test Error for depth-10: 0.10183639398998334

**Q4 [3pts]:** **Regression for digit completion**: regression

Use library sklearn.neural\_network.**MLPRegressor**

Using only the data for class 6 and class 9 in X100 for training, use the first 48 features as inputs and predict the next 16 features, i.e. create a neural network with 16 outputs.

Report the test error (use only the instances from classes 6 and 9).

Which pixels are easier to predict?

Error for each of the 16 features:

The test Error for feature 48: 0.5281633448406688

it is a easy to predict

The test Error for feature 49: 0.8170340310412312

it is a easy to predict

The test Error for feature 50: 2.305377215285779

It is not easy to predict

The test Error for feature 51: 2.532534723671279

It is not easy to predict

The test Error for feature 52: 3.715140149651046

It is not easy to predict

The test Error for feature 53: 3.138932931011379

It is not easy to predict

The test Error for feature 54: 3.065485791944507

It is not easy to predict

The test Error for feature 55: 1.3048045039849963

it is a easy to predict

The test Error for feature 56: 0.5281633448406688

it is a easy to predict

The test Error for feature 57: 0.5327671820989032

it is a easy to predict

The test Error for feature 58: 1.1780368880771195

it is a easy to predict

The test Error for feature 59: 2.067287527563326

It is not easy to predict

The test Error for feature 60: 2.3935227996872266

It is not easy to predict

The test Error for feature 61: 2.9158082112678616

It is not easy to predict

The test Error for feature 62: 3.7984137618163807

It is not easy to predict

The test Error for feature 63: 0.8052693628536799

it is a easy to predict

### Pixels 48, 56 and 57 have got less error rate