PROJECT PHASE 3
CSE 572
DATA MINING
SPRING 2018

SUBMITTED TO:

Professor Ayan Banerjee
Ira A. Fulton School of Engineering
Arizona State University

A Report by:

Athithyaa Selvam (aselvam@asu.edu)
Hari Siddarth Velaudampalayam Kesavan (hvelauda@asu.edu)
Mohan Vasantrao Yadav (mvasantr@asu.edu)
Raam Prashanth Namakkal Sudhakar (rnamakka@asu.edu)
Sangeetha Swaminathan (sswami11@asu.edu)
(Team 18)

TABLE OF CONTENTS

1. DATA PREPARATION 3
1.1 Accuracy Metrics 3
1.1.1 Precision 3
1.1.2 Recall 3
1.1.3 F-1 Score 3
1.1.4 Result 3

2. SUPPORT VECTOR MACHINES 4
2.1 Steps to classify data using SVM 4
2.2 Accuracy Metrics for each class using SVM 4

3. DECISION TREE 5
3.1. Steps to classify data using Decision Tree 5
3.2 Accuracy Metrics for each class using Decision Tree 5

4. NEURAL NETWORK 6
4.1. Steps to classify data using Neural Network 6
4.2 Accuracy Metrics for each class using Neural Network 7

5. CONCLUSION 7

1. DATA PREPARATION:

We worked on data collected from multiple users. The feature matrix from phase 2 is used as input for the three classification techniques. Training and Testing is performed on data collected from multiple users.

Number of Features used: 10

1.1 ACCURACY METRICS:

1.1.1 Precision:

Precision is the fraction of positive records that are actually positive in the positive class declared by the classifier. Mathematically, it is represented as,

$$\text{Precision} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})} \qquad (1.1.1.1)$$

1.1.2. Recall:

Recall is the fraction of positive records that are correctly predicted by the classifier. Mathematically, it is represented as,

$$\text{Recall} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})} \qquad (1.1.2.1)$$

1.1.3. F-1 Score:

F-1 score is the harmonic mean of precision and recall. Harmonic mean of any two numbers tends to be towards the smaller number and hence F-1 score ensures that recall and precision are high.

$$\text{F-1} = \frac{2 * \text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} \qquad (1.1.3.1)$$

1.1.4. Result:

Table 1.1.4.1 - Accuracy metrics

	Precision	Recall	F-1 Score
SVM	0.6939400848	0.8144666385	0.7316880487
Decision Tree	0.6902978066	0.7811492812	0.7308156287
Neural Network	0.7835200531	0.871499202	0.823479934

- **Average Precision:** 0.7225859815
- **Average Recall:** 0.8223717072
- **Average F-1 Score:** 0.7619945371

2. SUPPORT VECTOR MACHINES (SVM)

Support Vector Machine is a supervised learning model used for classifying a data point to one of the two available classes. A training dataset with two different classes is used to build the SVM model and this model is used to further classify the given data point to one of the two classes accurately. The algorithm generates an optimal hyperplane separating the two classes from the training data.

File: Code/Task3SVM.m

2.1. Steps to classify data using SVM

1. 60% of the data is used for training and 40% of the data is used for testing.
2. SVM model is built based on the training data and training labels using the matlab inbuilt function fitcsvm().
3. The inbuilt predict() function is called which takes in the SVM model constructed and the test data and it returns the predicted label.
4. The predicted labels are compared with actual test labels using the inbuilt function confusionmat() which generates the confusion matrix.
5. Values from the confusion matrix are used to calculate the precision, recall and F-1 Score.

2.2. Accuracy Metrics for each class using SVM

Table 2.2.1 - SVM Accuracy metrics

Class	Precision	Recall	F-1 Score
About	0.5417439703	0.7891891892	0.6424642464
And	0.7146464646	0.8761609907	0.7872044506
Can	0.6025641026	0.7532051282	0.6695156695
Cop	0.641221374	0.8208469055	0.72
Deaf	0.6357308585	0.904290429	0.7465940054
Decide	0.7222222222	0.9161073826	0.8076923077
Father	0.7339449541	0.8571428571	0.7907742998
Find	0.7795698925	0.9539473684	0.8579881657
Go Out	1	0.4528301887	0.6233766234
Hearing	0.5677570093	0.8209459459	0.6712707182

3. DECISION TREE

Decision tree is a greedy classification approach in which a tree like model is developed where each node is split based on a condition and the leaves represent the possible outcome classes of that model. Using this model the user can understand the target classes just by observation.

File Name: Code/Task3dt.m

3.1. Steps to classify data using Decision Tree

1. 60% of the data is used for training and 40% of the data is used for testing.
2. Decision Tree is built based on the training data and training labels using the matlab inbuilt function fitctree().
3. The inbuilt predict() function is called which takes in the tree generated and the test data and it returns the predicted label.
4. The predicted labels are compared with actual test labels using the inbuilt function confusionmat() which generates the confusion matrix.
5. Values from the confusion matrix are used to calculate the precision, recall and F-1 Score.

3.2. Accuracy Metrics for each class using Decision Tree

Table 3.2.1 - Decision tree accuracy metrics

Class	Precision	Recall	F-1 Score
About	0.5179704017	0.6621621622	0.581257414
And	0.7485875706	0.8204334365	0.7828655835
Can	0.6202898551	0.6858974359	0.6514459665
Cop	0.678125	0.7068403909	0.692185008
Deaf	0.712	0.8811881188	0.7876106195
Decide	0.7631578947	0.9731543624	0.8554572271
Father	0.6843853821	0.7357142857	0.7091222031
Find	0.6732394366	0.7861842105	0.7253414264
Go Out	0.9166666667	0.8301886792	0.8712871287
Hearing	0.5885558583	0.7297297297	0.6515837104

4. NEURAL NETWORK

Neural Networks are used to find patterns in data by creating a complex model relationship between inputs and outputs. Neural Networks are prominent for being versatile, i.e., they change themselves as they progress from starting, preparing and ensuing runs to give more data about the world.

File: Code/Task3nn.m

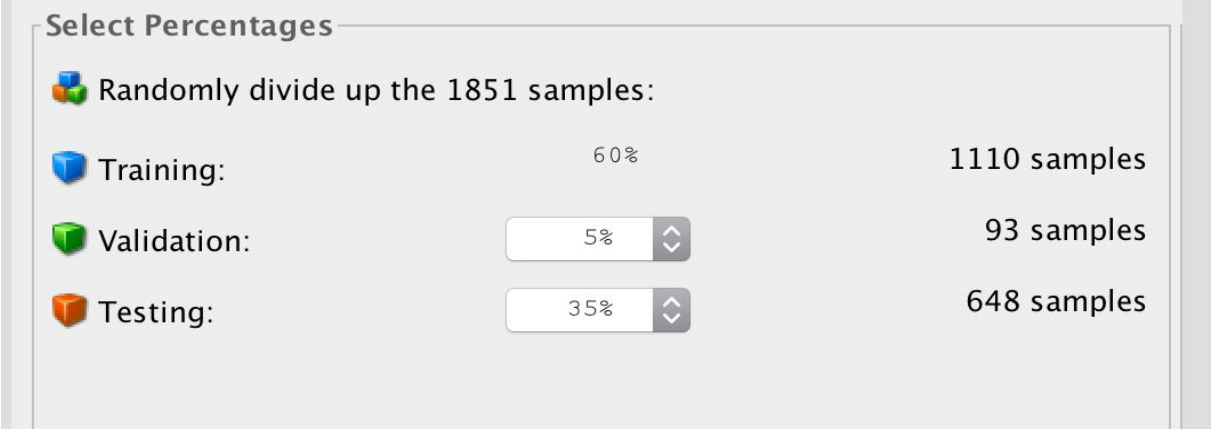

Fig 4.1 - Splitting of data

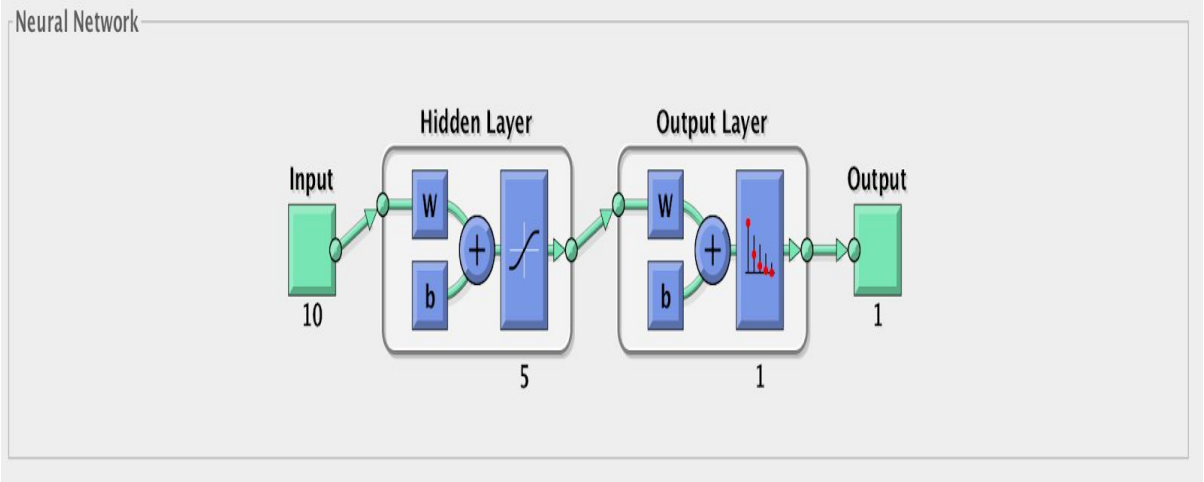

Fig 4.2 - Constructed neural network

4.1. Steps to classify data using Neural Network:

1. The data is divided into training, test and validation based on the percentage indicated as mentioned in Fig 4.1 using Neural Network Toolbox.
2. Neural Network is trained using scaled conjugate gradient back-propagation.
3. The Confusion matrix is plotted and the Precision, Recall and F-1 Scores are calculated.

4.2. Accuracy Metrics for each class using Neural Network:

Table 4.2.1 - Neural network accuracy metrics

Class	Precision	Recall	F-1 Score
About	0.7832817337	0.8006329114	0.7918622848
And	0.70609319	0.8454935622	0.76953125
Can	0.7527272727	0.9118942731	0.8247011952
Cop	0.7148148148	0.9146919431	0.8024948025
Deaf	0.8007968127	0.741697417	0.7701149425
Decide	0.7945736434	0.92760181	0.8559498956
Father	0.7717842324	0.7848101266	0.7782426778
Find	0.7635658915	0.8834080717	0.8191268191
Go Out	0.914893617	1	0.9555555556
Hearing	0.8326693227	0.9047619048	0.867219917

5. CONCLUSION

From our observation of the results produced from the three techniques, we find that Neural Network gives better accuracy for most of the gestures when compared to Support Vector Machine and Decision Tree.