PR INITIAL COPY

1)Introduction To the Problem

2)Classification Architecture

3)Features Extracted

4)Classification Implementation

5)Results including confusion matrix,ROC and F1 score

6)main Inference from the study including comparision with other classifiers and previous methods in literature

7)Conclusions

Problem:

The goal of this work is to train a classifier to predict which activities users are engaging in based on sensor data collected from devices attached to all four limbs and the torso.

CLASSIFICATION ARCHITECTURE:

DRAW A FLOWCHART

Gathering DATA->FEATURE EXTRACTION->CHOOsing A MODEL->Training->Evaluation->PREDICtion

WE first gather data. Sometimes the data we collect needs other forms of adjusting and manipulation. Things like de-duping, normalization, error correction, and more.Hence,we need to consider all those factors and prepare the data.

From the Multiple Models available we need to pick a model which best suits our data.

 we then use our data to incrementally improve our model’s ability to predict and classify the data.

The training process involves initializing some random values for W and b and attempting to predict the output with those values. We must adjust the values in W and b such that we will have more correct predictions. This process then repeats till the Cost Function Converges to a global Minima.

Evaluation allows us to test our model against data that has never been used for training. This metric allows us to see how the model might perform against data that it has not yet seen and Finally Predict the output for a given problem.

Features Extracted:

8 users all participate in the same 19 activities. Each of the 5 devices (4 limbs and 1 torso) have 9 sensors (x,y,z accelerometers, x,y,z gyroscopes, and x,y,z magnetometers). The data is collected in 5 second segments with a frequency of 25 Hz for a total of 5 minutes for each activity for each user.

Hence in each text file(of one segment), there are 5 units x 9 sensors = 45 columns and 5 sec x 25 Hz = 125 rows. Additional 15 columns of data are added which give the magnitude of each sensor. (magnitude of data of particular sensor). Hence overall We obtain 125 rows and 60 columns for each segment.

**Mean, Variance, Skewness, and Kurtosis**

The distribution of each signal are approximately Normal. This means that we can take the first four statistical moments for each 5 second segment. By including the four moments, we are helping our models better learn the characteristic of each unique activity.

**Autocorrelation**

First Ten Values of the Autocorrelation are taken.

## Maximum five peaks of the Discrete Fourier Transform

After taking the DFT of each 5-s signal, the maximum five Fourier peaks are selected so that a total of 300 Fourier peaks are obtained for each segment.

Hence, For each column, Mean, Variance,Kurtosis,Skewness,Autocorrelation(First 10 values) anf First Five Peaks of Discrete Fourier Transform are calculated using Fast Fourier Transformation Algorithm. This gives 19 features for each column and hence for 60 columns we obtain,1140 featurers.(60x19).

CLASSIFICATION IMPLEMENTATION:

SUPPORT VECTOR MACHINE(SVM):

Quadratic SVM:

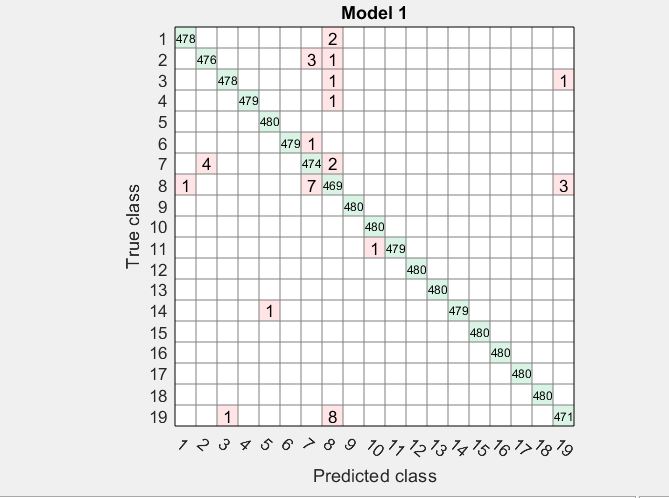
If the feature vectors in the original feature space are not linearly separable, SVMs pre-process and represent them in a higher-dimensional space where they can become linearly separable. The Support Vector Machine model performed substantially better than Logistic Regression. **The model is classifying activities at near 99% accuracy**.

The support Vector Machine Model gives a better accuracy when we apply feature Normalization. With Feature Normalization(Normalizing all features in the scale between 0-1),we get an accuracy of 99.7%. On Applying PCA(reducing the dimension to 30 from 1140),the accuracy is dropped to 99.6%. The drop in accuracy is insignificant when the training time is considered,which reduces immensely on applying PCA.

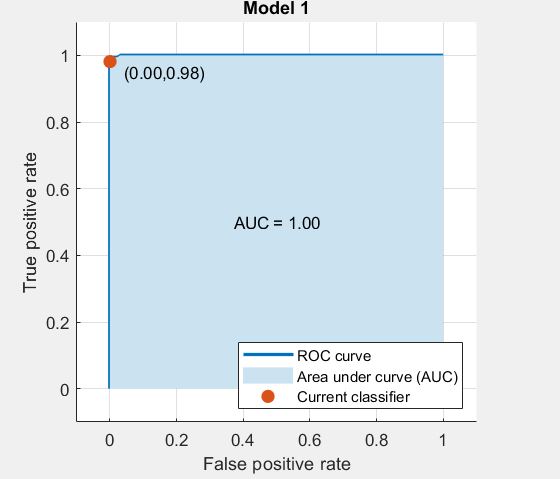
Without Feature Normalization, we get an accuracy of 99.7 % when we train the entire features. But the accuracy drops to 99% after applying PCA.

Hence,for Support Vector Machine, the ideal way is to Normalise the Feature and then apply PCA(reducing it to 30 features from 1140) as it gives the ideal output in terms of both Accuracy and Training Time.

CONFUSION MATRIX FOR SVM WITH MEAN NORMALIZATION AND PCA WHICH GIVES AN ACCURACY OF 99.6%

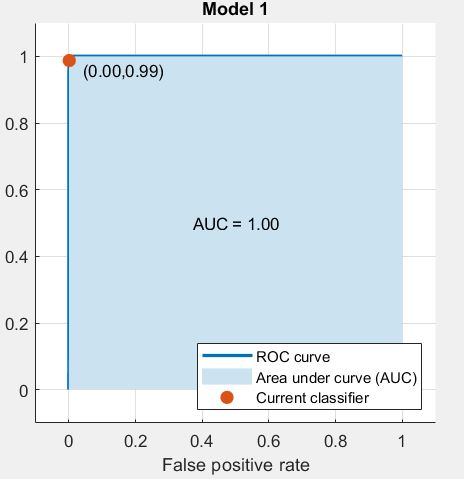


ROC Curve for classes 8,19:

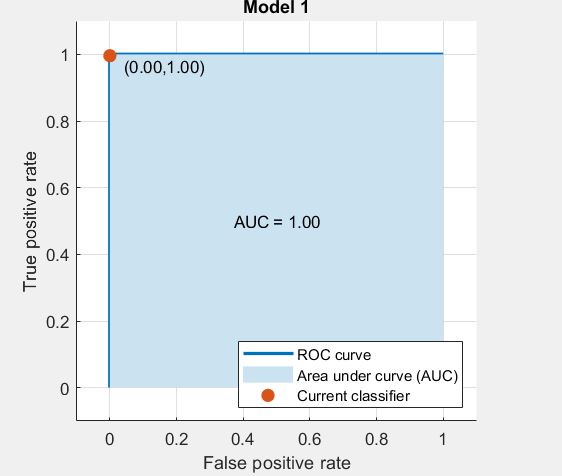


\

ROC CURVE For CLASSES 2,7:



ROC CURVE FOR CLASSES 1,3,4,5,6,9,10,11,12,13,14,15,16,17,18:



F-1 score:0.992

ARTIFICIAL NEURAL NETWORKS:

Multi-layer ANNs consist of an input layer, one or more hidden layers to extract progressively more meaningful features, and a single output layer, each composed of a number of units called neurons. The model of each neuron includes a smooth nonlinearity, called the activation function.

For ANN, Mean Normalization of the features give less accuracy compared to Non-Scaled data and hence we consider the original Scaling itself.

Draw a table:

5 hidden Layers- 88.87%

10 Hidden Layers-99.89%

20 Hidden Layers-99.97%

30 Hidden Layers-99.97%

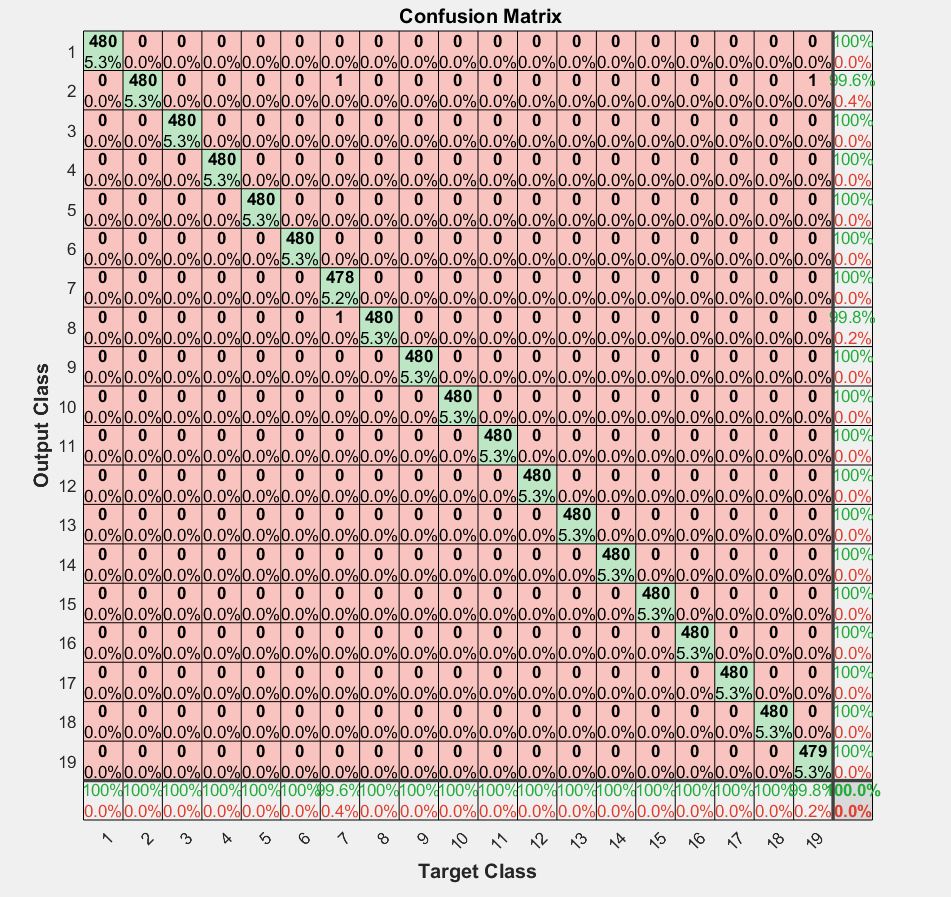
40 Hidden Layers-99.98%

100 Hidden Layers- 99.95%

An inordinately large number of neurons in the **hidden layers can increase** the **time** it takes to **train** the network. The amount of **training time can increase** to the point that it is impossible to adequately **train** the neural network.

The time taken for training increases with increase in the number of hidden Layers. Though the accuracy increases, it becomes insignificant when time taken to train is taken into account. So, 20 hidden layers are considered as it gives the best output in terms of Run-time and accuracy.

CONFUSION MATROIIX FOR 20 HIDDEN LAYERS



Precision=0.9996;

Recall=0.9996;

F-1 Score=0.9996;

COMPARING CLASSIFIERS:

Artificial Neural Networks gives the best output when both accuracy and Run-Time are considered. Though Quadratic SVM with Mean Normalization gives accuracy on the similar lines,the time taken for it to train is high compared to ANN. KNN without PCA gives 99.3% and with PCA gives 98%. Cubic SVM gives 98.9 with PCA and 99.6% without PCA.

Though Cubic And Quadratic SVM’s are giving accuracy on the similar lines ,the time taken to train a cubic SVM model is high compared to quadratic SVM and hence, Quadratic SVM is considered the second best.

DRAW A TABLE

ANN(20 Hidden Layers)-99.97%

Quadratic SVM(Mean Normalization and With PCA)-99.6%

Cubic SVM(without PCA)-99.6%

Cubic SVM(with PCA)-98.9%

KNN(without PCA)-99.3%

KNN(with PCA)- 98%

Conclusion-

Signal Processing and time series data can lead to engineering features and building machine learning models that predict which activity users are engaged in with 99% accuracy.

The model was able to learn which signals correspond to activities like walking or jumping for users.

We implemented and compared a number of different algorithms and ranked them in order of accuracy and training time.

Another aspect of activity recognition and classification that can be noticed is the normalization between the way different individuals perform the same activities. Each person does a particular activity differently due to differences in body size, style, and timing. When Mean Normalization

Is done(scaling all features between 0 and 1) we see a better accuracy in SVM’s. However it doesn’t have a significant effect in case of ANN.