CSE 535 Mobile Computing Assignment II

PROF. AYAN BANERJEE
I.R.A. FULTON SCHOOL OF ENGINEERING
ARIZONA STATE UNIVERSITY

Team Members

Srinivasula Reddy Kalluri

Nithiya Shree Uppara

Venu Gopinath Nukavarapu

Yasaswini Addada

Extraction:

Generated a merged csv file which contains the data of all input files with labels attached as book, car, movie, gift, sell, total. From the merged csv file, we dropped the over all scores. In the videos, we mainly focused on upper part of the body for extraction.

PCA Execution:

PCA can be used to reduce the dimensionality and to retain the most percentage from feature matrix. An orthogonal transformation is used to convert possible correlated variables to linearly uncorrelated variables. These variables are called as **Principal components**.

Training the models using given data:

We trained the models using the matrix obtained after PCA. Then we used Decision Tree Classifier, K-Nearest Neighbor Classifier, Random Forest and Multi-Layer Perceptron.

Decision Tree Classifier:

It is a supervised learning model. In this model, data is split which can be used for both classification and regression. The model selects the best attribute to split the records. Based on that attribute, it breaks the data into smaller subsets. Now tree is built using this subset. We set the parameters of Decision Tree Classifier like criterion as entropy and max_depth as 15 and min_samples_leaf as 5.

For this model, we got an accuracy of around 80%.

K- Nearest Neighbor Classifier:

It is a supervised machine learning algorithm. KNN believes that data is in the closer proximity when plotted as a graph. We need to choose the right value of k that fits our data. If we decrease the value of k to 1, our predictions become less stable. If we increase the value of k, our predictions become stable. The distance between all the k neighbors are calculated. The one which has the highest proximity will give as output class. We set the parameters of K-Neighbor Classifier like n-Neighbors as 30 and metric .as Euclidean.

For this model, we got an accuracy of around 79%.

Random Forest Classifier:

It is an ensemble classifier which is a mixture of 2 or more classifiers. It creates a set of decision trees. The decision trees are selected randomly from training data. It then decides the final class based on the votes from different decision trees. We set the parameters of Random Forest Classifier like n-estimators as 28 and max_depth as 8.

For this model, we got an accuracy of around 77%.

MLP Classifier:

It is a logistic regression classifier. For this classifier, the input is transformed using a non-linear transformation. It projects the input data into space where it becomes linearly separable. We call this as hidden layer. We set the parameters of MLP classifier like solver as lbfgs, hidden layer sizes as 5 and random states as 1.

For this model, we got an accuracy of around 70%.

Classifier	**Accuracy**
Decision Tree Classifier	80%
K-Nearest Neighbor Classifier	79%
Random Forest Classifier	77%
MLP Classifier	70%