

Pattern Recognition and Machine Learning

Minor project Report

Project Name : Detecting whether the person is wearing masks or not

Team Members:

Hari Bhutanadhu-B19EE017

Guvvala Sujitha-B19EE033

Importing Dataset and Splitting:

The task of this project is to build a model that detects whether the person is wearing a mask or not .

For this we uploaded a zip file that contains a required dataset in the Google colab. And we unzipped the dataset using an extractall function that contains Mask and without mask data images .

After extracting the zip file ,resized each image to the same size and convert the images into numpy arrays.

Assigned labels for these images .For masked images it is labelled as 1 and for the images where there is no mask is labelled as 0.

Splitted the data into training,validation and testing dataset of ratio 70-20-10.

SVM Model Building :

Performed GridSearchCV on SVM classifier.The hyperparameters taken for gridsearch on SVM are kernel and C values.

SVM is a linear model for classification. It solves linear and non-linear problems. The algorithm creates a line or a hyperplane which separates the data into classes.

The hyperparameters we have taken are kernel:{'rbf','linear','poly'} and C:[1, 10,100]and then performed GridSearchCv on this classifier to get the best parameters .

Using these best parameters we have built an SVM model by fitting training data and calculated its CV scores , mean accuracy and accuracy for testing data.

KNN Model Building :

Performed GridSearchCV on KNN classifier. The hyperparameters taken for gridsearch on SVM are `n_neighbours`.

K-Nearest Neighbors (KNN) is for regression and classification problems. The algorithm uses data and classifies new data points based on similarity measures (example: distance function). Classification is done by votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression)

The hyperparameters we have taken are `n_neighbours=list(range(1, 9))` and then performed GridSearchCv on this classifier to get the best parameters .

Using these best parameters we have built an KNN model by fitting training data and calculated its CV scores , mean accuracy and accuracy for testing data.

MLP Model Building :

Performed GridSearchCV on MLP classifier. The hyperparameters taken for gridsearch on SVM are `max_iterations`.

MLPClassifier stands for Multi-layer Perceptron classifier which in the name itself connects to a Neural Network. It is a class of feedforward artificial neural network (ANN). It utilizes a supervised learning technique called backpropagation for training. The major use cases of MLP are pattern classification, recognition, prediction and approximation. It consists of three types of layers-the input layer, output layer and hidden layer

The hyperparameters we have taken are `max_iter : (100, 500, 1000)` and then performed GridSearchCv on this classifier to get the best parameters .

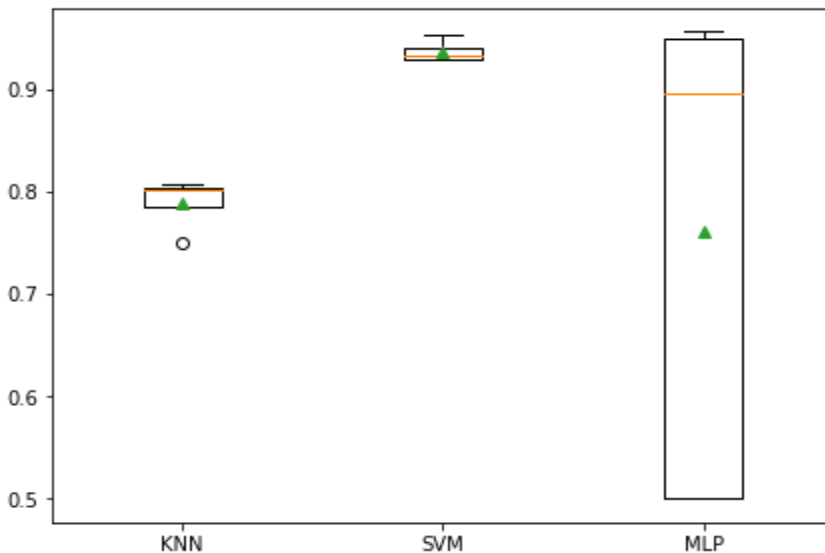
Using these best parameters we have built an MLP model by fitting training data and calculated its CV scores , mean accuracy and accuracy for testing data.

Comparison of Models :

Model	CV Scores	Training Accuracy	Testing Accuracy
KNN Model	[0.78476821, 0.81125828, 0.80794702, 0.81395349, 0.85049834]	0.8824	0.8145
SVM Model	[0.93377483, 0.94039735, 0.93046358, 0.93023256, 0.95348837]	0.9642	0.9536

MLP Model	[0.95033113, 0.50166113, 0.89735099, 0.95681063]	0.9440	0.9205
-----------	--	--------	--------

Comparing the three models by cross validation and plotting the box plot.



From the box plot we can tell that the SVM model has high mean accuracy , so it can be a better model in terms of accuracy while compared to the other models.

Challenges Faced :

Importing the data was one of the challenges,as the data samples are images.we have tried several times to convert the images into arrays .

Selecting the best classifier was one of the challenges for us ,as we have tried several classifiers to get better accuracy and performance .

To increase the accuracy was also a challenge for us ,for this we have trained the models by changing parameters of the classifiers by using GridSearchCv.

Conclusion :

By comparing every model the the SVM model has high mean accuracy , so it can be a better model in terms of accuracy while compared to the other models.And we have learnt how to convert images into useful numpy arrays using various methods for classifying the models which helped in gaining lots of information.

Contribution of Team Members :

Hari Bhutanadhu :

Data Importing , tried converting images into numpy arrays , built SVM model ,KNN model and the respective classification reports .Tried for improving accuracy .

Guvvala Sujitha :

Tried converting images into numpy arrays, built MLP model and the respective classification report ,done model performance comparison .

Report was written in the presence of both .

Reference Links :

<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>

<https://stackoverflow.com/questions/49685924/extract-google-drive-zip-from-google-colab-notebook>