

Name- Mahizebin shams-F- Mofiz

ID - 1712639692

Section- 05

" Assignment - 03 "

Date: 25 May, 2021

(A)

## Ans to the Qus NO-01

Q1.

PCA is considered only as a variance to the features, but it is not considered as a relationship between features and labels while doing feature reduction.

But in case of regularization acts directly on the relationship between features and labels. Thus, regularization helps to develop models that are better at showing predictions for a given set of features.

As we apply PCA to reduce features in a data set so regularization makes the process better for this reason we still need regularization.

(2)

Ans to the Ques NO-02

Q2.

F<sub>1</sub> score:

F<sub>1</sub> score is the balance between the precision and the recall.

$$\therefore F_1 = \frac{2 * (\text{precision} * \text{recall})}{(\text{precision} + \text{recall})}$$

I would use F<sub>1</sub> score as a performance metric for classification algorithms and gives a better measure of the incorrectly classified cases than the accuracy metric.

The methods I use to handle missing or corrupted data in a dataset are —

(3)

## Method-01

### Deleting rows or columns:

We usually use this method when it comes to empty cells. If the majority of our data is missing for a column or for a row, we can simply delete them.

## Method-02

### Replacing the missing data with

### aggregated values:

We can calculate the aggregated value based on the rest of the values we have in the column and put the received numbers to the empty spot.

## Method-03

(1)

### Creating an unknown category:

Categorical features have a number of possible values, so we can create one more category for the missing values. This way the lower variance by adding new information to the data, will create. This also could be used when the original information is missing or cannot be understood.

## Method-04

### Predicting missing values:

Where we have no missing values, we can train a machine learning algorithm in order to predict the missing values since the samples for which this training

is performed. There are missing values, it is necessary to replace initially using one of the simplest methods for recovering gaps.

This way it will create and give us better performance unless a missing value should have a high variance.

Ans to the Ques No-03

03

Principal Component,

Principal component are the eigenvectors of a covariance matrix. The idea to reduce the dimensionality of the dataset by transforming the variables to a new set of variables, it is known that the principal components ordered in such a way that retention of variation present in the original variables decreases as we move down.

The difference between the first and the last principal component,

The difference is that the first principal component

(A)

has the direction in space along with projections and the largest variance.

On the other hand, the last component has the variance-maximizing direction orthogonal to the previous component.

KNN,

KNN algorithm is said to work better on small data sets because when the dataset is small, the classifier completes execution in shorter time duration. This algorithm doesn't have a training step as it's a case-based learning algorithm to make the predictions.

## Ans to the Ques No-04

04.

In my opinion, No. The figure demonstrates a decision boundary and there is a under fit to the training set, we so if would say we would like to lower the bias or increase the variance of the SVM. Hence the value of  $c$  should be increased and the value of gamma should be decreased.

Q

Ans to the Ques NO-05

05.

When the max-depth value is none, it is set in default mode and the nodes are derived from a single class. The decision tree takes its natural course of expanding and is not expanded forcefully.

When the max-depth value is higher than the default value, the nodes are derived from more than one class and it overfits the decision tree model.

When the max-depth value is lower than the default value, the nodes are derived from a single class and it underfits the decision tree model.

In my opinion, a node is said to be impure if it is split from different classes. For instance, if a node is 100% derived from a single class then it is pure but if a node is 50% derived from one class and the another 50% derived from other then the node is impure.

Ans to the Ques NO-06

(11)

Q6,

In the random forest construction,

The number of decision trees that we need to use to get a good result is hard to express.

There is a threshold to the number of trees that need to be used. Moreover, there is no significant performance gain that will only increase the computational cost.

A random forest is a supervised learning algorithm that performs both regression and classification task with using multiple decision trees. It can be used for

improving both classifications and regression problems. To use for regression problem, in random forest algorithm first the "forest" has to be built. It is an ensemble of decision trees and this process is known as "Bagging." The ~~bi~~bagging method is a combination of learning models and it will increase the overall result.

