TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE TIJUANA

SUBDIRECCIÓN ACADÉMICA

DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN

SEMESTRE:

Enero - Junio 2021

CARRERA:

Ingeniería en Sistemas Computacionales

MATERIA:

Datos Masivos

TÍTULO:

Practica-1

UNIDAD A EVALUAR:

Unidad -2

ALUMNO: JUAN ANTONIO ACEVES ZAMORA

NO. CONTROL:16210502

NOMBRE DEL DOCENTE :

JOSE CHRISTIAN ROMERO HERNANDEZ

One-vs-Rest Classifier

Import the libraries that will be essential for the regression and classification of the data.

```
import org.apache.spark.ml.classification.{LogisticRegression,
OneVsRest}
import
org.apache.spark.ml.evaluation.MulticlassClassificationEvaluator
```

Upload the file that will generate the results.

```
val inputData =
spark.read.format("libsvm").load("sample_multiclass_classification
_data.txt")
```

Generate the train division and test set.

```
val Array(train, test) = inputData.randomSplit(Array(0.8, 0.2))
```

Create an instance of the base classifier.

```
val classifier = new
LogisticRegression().setMaxIter(10).setTol(1E-
6).setFitIntercept(true)
```

An instance of the One Vs Rest classifier is created.

```
val ovr = new OneVsRest().setClassifier(classifier)
```

The multiclass model is trained.

```
val ovrModel = ovr.fit(train)
```

The model is scored based on the test data.

```
val predictions = ovrModel.transform(test)
```

The evaluator is obtained.

```
val evaluator = new
MulticlassClassificationEvaluator().setMetricName("accuracy")
```

The classification error in the test data is calculated.

```
val accuracy = evaluator.evaluate(predictions)
println(s"Test Error = ${1 - accuracy}")
```