Tecnológico Nacional de México

INSTITUTO TECNOLÓGICO CAMPUS TIJUANA

ING. EN SISTEMAS COMPUTACIONALES

Subdirección Académica

Departamento de Sistemas y Computación

BDD-1704 TI9A - 6:00pm-7:00pm

ASIGNATURA:

Datos Masivos

SEMESTRE:

Septiembre- Enero 2020

Tarea:

Explicación de lo que yo entendí del código:

OnevsRestExample.scala

MAESTRO:

JOSE CHRISTIAN ROMERO HERNANDEZ

04/12/2020

//importamos las librerias correspondientes

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

import org.apache.spark.ml.evaluation.MulticlassClassificationEvaluator

// cargar archivo de datos.

var inputData = spark.read.format("libsvm").load("C:\Users\DELL\Desktop\ejercicioonevsrest/sample\_multiclass\_classification\_data.txt")

// generar la división de tren / prueba.

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

// instanciar el clasificador base

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

// Crea una instancia del clasificador One Vs Rest.

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

// entrena el modelo multiclase.

val ovrModel = ovr.fit(train)

// puntuar el modelo en los datos de prueba.

val predictions = ovrModel.transform(test)

// Obtener evaluador.

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

// calcula el error de clasificación en los datos de prueba.

val accuracy = evaluator.evaluate(predictions)

println(s"Test Error = ${1 - accuracy}")

//Resultados

