BYU **IDAHO** Machine Learning & Data Mining | CS 450

08 Prove : Assignment
Support Vector Machine Experimenting

I. Objective

Be able to appropriately apply Support Vector Machines to classify in data.

II. Overview

For this assignment you will not be implementing a Support Vector Machine of your own, but rather, will be using an existing implementation to classify data and experimenting with its parameters.

In addition, as we move to start using algorithms, rather than implementing them, we will also switch to begin using the R language. For this assignment, you are welcome to use Python if you would like, but I would encourage you to start using R, and I will provide several code examples to help get you started.

III. Instructions

Please read sections 1-4 of the document: A Practical Guide to Support Vector Classification. This document is written by the creators of the popular LIBSVM implementation of a Support Vector Machine Library.

Your assignment is to apply their prescribed procedure (in essence, a grid search for hyper-parameters of the RBF Kernel) to see how

good of classification you can make on three datasets.

You may use any implementation of the SVM algorithm that you choose, but a popular choice to consider is LIBSVM.

USING R

R is a statistical programming language that is very popular in Data Science. It is excellent at facilitating exploratory data analysis and producing graphs, and it also has the power to integrate with Databases, APIs, Machine Learning algorithms, and all the toolset of a data scientist.

A popular development environment for R, that I would recommend is R Studio. It is free and used by industry professionals. You can download it here.

Coming up to speed on the basics of the language will come with time, but as a great place to start, please read R for Programmers which mentions some of the basics and subtleties of R for those with a background in more traditional programming languages.

USING SVMS IN R

The LIBSVM package for R is called, "e1071", and this tutorial walks you through the process of how to use it to train and evaluate an SVM.

Also, please read A few helpful hints for SVMs in R that I have written to guide you in the right direction.

IV. Experiment Guidelines

Use a Support Vector Machine to classify each of the following datasets. For each one, follow the approach outlined above (The

Practical Guide to SVM) to experiment with different values for C and gamma. Then, report how many different parameters combinations you tried, and then the highest accuracy you obtained along with the parameters you used to obtain this accuracy.

Although not required, you are welcome to write code to automate the process of trying different parameters, but do not simply use a predefined "tune" method.

You should perform a minimum of 10 parameter combinations on each dataset.

The datasets are:

- » [Vowel](#) - Make sure to use "vowel-context.data" rather than "vowel.data". A nicely prepared version is available for you here: [vowel.csv](#)

- » [OCR for Letter Recognition](#) - You are classifying the letter here, based on visual properties. A nicely prepared version is available here: [letters.csv](#)

V. Submission

When complete, fill out the [submission form](#) and upload it to I-Learn.