

CIS 635 Data Mining

Homework 5

Description

This homework has you practicing some of the practical concepts related to SVM.

Instructions

Part 1 – predicting

In this first part, you are given the vector w that determines the linear decision boundary for a particular data training set. You are to assume that SVM has been used on the data set to find w . Your task is to make predictions on new data, given w . The new data instances are in the table to the right and $w = \{-4, 1, -2\}$.

x_0	x_1	x_2
1	2.1	3.8
1	9.7	1.3
1	7.2	1.3

Predict the class of the new instances (1, -1 or 0). The class is 1 if $x \cdot w > 1$, and -1 if $x \cdot w < -1$. If it is between -1 and 1, the class is indeterminate (i.e. 0).

Part 2 – using SVM for a problem with 3 class values

Using the training data hw06dataTrain.txt and test data hw06dataTest.txt for this exercise. The training and test sets both have 2 numeric attributes and a nominal class. The class can be one of three values 0, 1 and 2. It may be helpful for you to plot the data in a scatter plot.

You are to use SVM to classify the test records. Since it is a binary classifier you cannot just create a single model. The approach you will take is pairwise. You will create three models. First you will create a new training set with just the instances that are class 0 and class 1 (in the new sets modify the class values so they are 1 and -1). Use SVM to train a model (use the default kernel, which is radial) on this set (I will refer to this as mod01). Then create another training set with just the class 0 and 2 (again, change the classes to 1 and -1). Train another model on this set (mod02). Finally train another model (mod12) on a set with classes 1 and 2. It is very important that you are organized; think clearly about which model is predicting which class.

After you have the three models trained you can use them to predict the test set. Use the majority rule to assign class. That is, if mod01 predicts 0 and mod02 predicts 2 and mod12 predicts 2, then assign it class 2. In the unlikely case where the three models predict three different classes, assign it a class of 0. When you are finished print the confusion matrix and a script of the R commands that you used.