School of Mathematics and Statistics
MAST90083: Computational Statistics and Data Science
Assignment 3
Due date: No later than 6:00 pm on Tuesday 18th October 2022
Weight: 15%

Question 1: Support Vector Machines

Use of any function or library other than what is mentioned in this assignment is not recommended. Use library e1071 that contains the SVM function for this assignment. Unless specified otherwise, set the seed to 50 for all instances i.e. whenever the random number generator is invoked by any function you should use a seed. You should also note that due to the way in which the plotting function is implemented in the library e1071 the decision boundary for linear kernel case might look jagged. You may use "rep" and "sample" function in addition to the functions that have already been mentioned in the assignment.

1. We are going to produce a random data of size 100×2 for each of the three classes (C=3). This can be generated as an aggregated random data $\mathbf{x}$ of size $N \times 2$ as "matrix(rnorm(N*2), ncol=2)", where $N = 300$. Each 100 entries in this matrix belong to a separate class, first 100 to class 1, next 100 to class 2 and last 100 to class 3, however since all observations were generated from the same distribution it is not possible to differentiate among them. To make these 300 entries distinctive and divide the data into 3 different classes, lets define class specific means in variable $\mathbf{z}$ as "matrix(c(0,0,3,0,3,0),C,2)". Also, generate a response vector $\mathbf{y}$ of size N that contains labels (1 to 3) for the data in $\mathbf{x}$. Using $\mathbf{z}$ and $\mathbf{y}$, assign class specific means to data points of each class and this operation will change the entries of the matrix $\mathbf{x}$ and divide it into three classes. Use ggplot from the library "ggplot2" to plot $\mathbf{x}$ as a data frame while using $\mathbf{y}$ as a factor for colour assignment.

2. Construct the data frame for the training data as "tdata=data.frame($\mathbf{x} = \mathbf{x}$, $\mathbf{y}$=as.factor($\mathbf{y}$))" and fit the support vector classifier using svm function by setting the kernel as linear, and cost as 10 and store the result in svmfit. Now, plot the results as "plot(svmfit, tdata)". Also generate summary using the object svmfit and answer how many support vectors were there in each class?

3. Using the training data from the previous question, perform a ten-fold cross-validation by utilizing the function "tune" and providing it with a list of cost values as $0.001, 0.01,$ $0.1, 1, 5, 10, 100$. Use summary on the object returned by the tune function to find out at what value of cost, the minimum cross validation error rate was found. For this best cost value, did the number of support vectors increase? How many support vectors were there in each class? Also, save the best model returned by the tune function as "bestmod".

4. Set the seed to 100 and generate a test data following the exact approach of question 1 and the syntax "testdata=data.frame(x=xtest , y=as.factor (ytest))", the only difference however is that ytest is now labeled randomly with replacement and not in a sequence of first 100 to class 1 (label 1) and so on. Now, use predict function with input arguments as "bestmod" (from previous question) and "testdata" to predict the class label of these test observations and store the results in $\mathbf{y}_p$. Use the function "table" to print the results in form of a table for the vector of predicted labels ($\mathbf{y}_p$) against the test labels ytest. How many observations are misclassified? Why in one case the number of correctly classified observations are greater than 100?

5. Initially, for training, cost and gamma are both set to 1 and then for the tuning purpose their values are set to $0.1, 1, 10, 100, 1000$ and $0.5, 1, 2, 3, 4$, respectively. Find how many observations are misclassified using the best model when the kernel is radial (i.e. repeat question 1 to 4 with radial kernel). Does the result imply that data is linearly separable and we do not need the radial kernel? What were the optimal (best) cost and gamma (parameter of the radial basis function) in this case?

Question 2: Artificial Neural Networks

Suppose we have a two-layer network with r input nodes x_m, $m = 1, ..., r$, a single layer ($L = 1$) of t hidden nodes Z_j, $j = 1, ..., t$ and s output nodes Y_k, $k = 1, ..., s$. Let β_{mj} be the weight of the connection $X_m \rightarrow Z_j$ with bias β_{0j} and let α_{jk} be the weight of the connection $Z_j \rightarrow Y_k$ with bias α_{0k}. The functions $f_j(.)$, $j = 1, ..., t$ and $g_k(.)$, $k = 1, ..., s$ are the activation functions for the hidden and output layers nodes respectively.

1. Derive the expression for the value of the kth output node of the network as function of α_{0k}, g_k, α_{jk}, f_j, β_{0j}, β_{mj} and X_m

2. Derive the matrix form for the vector of output of the network

3. Under which conditions this network becomes equivalent to a single-layer perceptron

4. What is the special case model obtained when the activation functions for the hidden and output nodes are taken to be identity functions.

Grading:

- Total: 15 points

- Question 1: 8 points

- Question 2: 7 points

The assignment is to be submitted via LMS