

# Assignment 2

---

In this assignment you are given an input file composed of points on a 2D plane. Each point is given as a triplet of (x coordinate, y coordinate, color) attributes. The value of the color attribute is 1 if the point is red and 0 if the point is black. The main idea behind the assignment is to train a system with the given points such that when a new point with an unknown color is given, your system will predict the color of the given point correctly based only on x and y coordinates of the given point. The classifiers you implement will be tested with a new input file similar to the input file given with this assignment.

You are required to submit your MATLAB codes and the necessary workspace for running the codes. You should also submit a report (maximum 5 pages) demonstrating your progress with results, plots etc. The assignment is due **April 27, 2014 23:59 PM**. For all your questions, you can send an e-mail to Çağla Çığ (ccig@ku.edu.tr).

## Instructions

Using libsvm (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>) and MATLAB, train 3 different SVMs for the input file. Train the first SVM using a linear kernel, the second SVM using a radial basis function (RBF) kernel and the third using a polynomial kernel. For all SVMs:

1- Use 5-fold cross-validation to separate the input file into training/validation and test data. Training data is used for calculating the support vectors of a support vector model. Validation data is used for finding the optimum cost and gamma parameter values before training a model. Test data is used for measuring the actual predictive power of a model.

2- Use 5-fold cross-validation to further separate the training/validation data into training and validation data. Plot the 2D heatmap matrix that shows the cross-validation accuracy value for each cost and gamma parameter pair you have tried during grid search. The ranges for cost and gamma parameters should be as follows:  $(-1 \leq \log_2 \text{cost} \leq 10)$  and  $(-10 \leq \log_2 \text{gamma} \leq 1)$ .

For each SVM, report the number of support vectors you have in your trained model. Submit each trained model separately as a MATLAB file.

In your report, give brief explanations to the following questions:

- What is the purpose of cross-validation in machine learning?
- What is the purpose of grid search in machine learning?

After you are happy with your classifier, test your classifier with the input file. Document the accuracy of your classifier (Number of correctly labeled points / Number of all points) along with a plot of the input file points. Plot the points according to your color labels instead of the original color labels.