

Assignment: K-means Clustering and Radial Basis Functions (RBF)

Problem 1 (20 points)

In this problem you will use the data file “vizData.csv” (x and y denote the input and target) to visualize the effect of gamma in the parametric RBF model on a simple data set. Use all training points in the data set as cluster centers.

1. Create a regression RBF model with γ equal 5. Plot the final hypothesis curve along with the training points on the same plot.
2. Create a regression RBF model with γ equal $1/5$. Plot the final hypothesis curve along with the training points on the same plot.
3. Explain your observations.

Problem 2 (20 points)

In this problem you will use the data file “kMeansData.csv” (x_1 and x_2 denote the input features) to create 3 clusters using unsupervised Lloyd’s k-means algorithm.

The training should only stop if the difference between the cluster center locations in two consecutive iterations is less than 0.001 or if the number of iterations has reached 1000. For the initial selection of cluster locations choose 3 points from the data set randomly.

After convergence, report the final cluster centers. Plot the 3 clusters in different colors with cluster centers clearly marked on the plot.

Problem 3 (30 points)

In this problem you will use the data file “rbfClassification.csv” to create an RBF classification model. x_1 and x_2 denote the input features and cls denotes the target class of the corresponding data points.

1. Use k-means clustering to determine the location of 2 cluster centers that you will use in your RBF model. Report the coordinate of the cluster centers.
2. Train an RBF model using $\gamma = 0.5$. Report the correct classification rate of your model.