Indian Institute of Space Science and Technology – Thiruvananthapuram

MA613 Data Mining

Assignment-IV

Date: 17-10-2014

1. (a) Develop 2nd degree, 3 degree, and 7th degree polynomial models using ridge regression for the attached data (Data.txt).

   (b) Use leave one out (1 fold) cross validation.

   (c) Plot λ values against training error and validation error, in a single figure.

   (d) Plot J_reg(w). Plot the graph of the resulting models.

   (e) Compare the performance of the models and select the best model.

   (f) Compare the performance of the best model with its least square counterpart.

2. Apply regularized linear regression on Boston housing data and compare its performance with linear regression.**Note:** Normalize only the input data.

3. Apply k nearest neighbour algorithm on Breast Cancer Wisconsin data (Diagnostic). Use 5 fold cross validation. Report the performance of the model. Plot ROC curve.

4. Using Weka software, apply decision tree on Adult data. Report the results.
   Note: Set apart 20% of the data as testing data for all the analysis above and report the results on the testing data.

5. A database has five transactions. Let $min_{sup} = 60\%$ and $min_{conf} = 80\%$. List all the

TID	Items bought
100	(M, O, N, K, E, Y)
200	(D, O, N, K, E, Y)
300	(M, A, K, E)
400	(M, U, C, K, Y)
500	(C, O, O, K, I, E)

strong association rules using Apriori algorithm by (a) manually (b) Weka software.

Notes

- All the files related with the assignment should be saved in a single folder and send to sumitra@iist.ac.in.

- Last date of submission: 24-10-2014.

- **As far as assignments are concerned, students are expected to observe academic honesty and integrity. Though the students can collaborate and discuss, copying directly other students' assignment or allowing your own assignment to be copied constitute academic dishonesty and is highly discouraged.**