Indian Institute of Space Science and Technology – Thiruvananthapuram

MA613 Data Mining

Assignment-III

Date: 22-09-2014

1. (a) Apply logistic regression, Gaussian discriminant analysis and Naive Bayes on the following data:
   i. Data1 (please find the attached data)
   ii. Haberman's Survival Data Set (download the data from UCI web repository)

   (b) Report the preprocessing techniques you have done on the data, if any.

   (c) Compare the performance of the algorithms.

   (d) Plot the decision boundary for the three algorithms. Plot the logistic regression curve.

   (e) Plot multivariate Gaussian density function corresponding to the positive & negative classes and their contours for Data1.

   (f) Report the values of the parameters of the final model.

2. Write short notes on analysis of variance (ANOVA) & students t distribution and apply their concepts on the above data.

3. Check whether the emails having the content (a) "review your account" (b)"review us now" are spam, on the basis of the below information:
   D1: "send us your internet banking password" spam
   D2: "send us your review" ham
   D3: "review your internet banking password" ham
   D4: "review us" spam
   D5: "send your internet banking password" spam
   D6: "send us your account details" spam

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: 01-10-2014.
- **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.**