Indian Institute of Space Science and Technology – Thiruvananthapuram

MA613 Data Mining & MA633 Data Mining Lab

Assignment-II

Date: 02-10-2020

1. Analyze data1 and data2 using linear regression
   (a) Plot $J(w)$ against iteration.
   (b) Plot $J(w)$ against w for data1.
   (c) Plot the hyperplane obtained. In the same figure, plot the output points also.
   (d) Describe the cross validation techniques used.
   (e) Report the average mean square error and standard deviation.
   (f) Report the parameters of the model.

2. Analyze data3 using linear regression
   (a) Find the parameters using direct method and iterative methods.
   (b) Compare the time taken and space requirements of two approaches.
   (c) Describe the search space and cross validation techniques used to find the values of the hyperparameters of the model.
   (d) Report the values of the hyperparameters.
   (e) Report the parameters of the model.
   (f) Compare the performance of the two models using statistical techniques.

3. Analyze data4 using linear regression
   (a) Describe the cross validation techniques used.
   (b) Describe the method used to find the parameters.
   (c) Report the values of the parameters of the model.
   (d) Report the average mean square error and standard deviation.

4. Analyze Wine quality data set using linear regression (download data from UCI web repository)
   (a) Analyze the data with normalization and without normalization.

(b) Describe how you applied normalization techniques on training and testing data.

(c) Apply k fold cross validation and hold out method.

(d) Assess the performance of the model.

(e) Report search space & the values of the hyperparameters and the parameters of the model.

(f) Apply batch as well as online optimization algorithms and compare their performance using statistical measures. Compare the time taken by the two algorithms.

Notes

- All the files related with the assignment should be saved in a single folder.
- Last date of submission: 11-10-2020.
- **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.**