Indian Institute of Technology Madras

ID5055 Foundations of Machine learning

Tutorial VII

Due date: 11:59 pm, October 15, 2023

Instruction

1. Assignment shall be submitted on the due date. Late submissions will not be entertained. If you cannot submit the assignment due to some reasons, please contact the instructor by email.
2. All the assignments must be the student's own work. The students are encouraged to collaborate or consult friends. In the case of collaborative work, please write every student's name on the submitted solution.
3. If you find the solution in the book or article or on the website, please indicate the reference in the solutions.

Problems

1. Refer to the tutorial notebook for details. Create datasets with number of features - 1000, 2000, 3000, 4000 and 5000. For this question use the same dataset generation parameters as Dataset 1 in notebook. Change only the number of features (total and informative) per datapoint. Use time module to calculate runtime of Naive Bayes for each dataset created above and plot results
2. For the datasets with class imbalance used in the tutorial today (Dataset 3 and Dataset 4), try improving the performance in the specificity in the Gaussian naive Bayes model. [Hint: Set the priors beforehand]
3. Use ComplementNB class in *sklearn* to fit a naive Bayes classifier and see whether the performance (specificity) increases for the imbalanced dataset.