

CS5011: Machine Learning

Feb-May Trimester 2021

Slots: Tue (1630-1800) & Thu (1700-1830 HRS), Prof. Sukhendu Das, BSB 312;

Phone: 4367 Email: *sdas@iitm.ac.in*,
sdas@cse.iitm.ac.in

Updated on February 09, 2021

Note: The course webpage is http://www.cse.iitm.ac.in/~vplab/E_machine_learning.html.

1 Course Objectives

In the recent past, algorithms of solving many ill-posed problems in the field of multi-dimensional signal processing and big data analytics have gained importance. New methods of signal representation, modeling, optimization and learning have been formulated, which spans over various areas of Machine Learning, Pattern Recognition, Vision and Natural Language Processing. This course will provide an overview of the theories and current practices, required by students and industrial practitioners who intend to specialize in this field, to understand and solve complex problems in Machine Learning Applications for image, video, text and bioinformatics.

2 Learning Outcomes

- To learn existing statistical algorithms of Machine Learning (ML) and Pattern Recognition (PR).
- To understand the difference between Classification and Regression.
- To be aware of recent advances in the field of ML such as Graphical models, Ensemble methods etc.
- To have hands-on experience in implementing various ML techniques on different datasets.
- To learn to compare the performance of two learning systems.
- To study few optimization methods used to estimate the parameters of a model during training.

3 Course prerequisite(s)

Basics of Linear Algebra, Probability theory and statistics.

4 Classroom Mode

Online Lectures (2 × [1 hr. 30 mins. slots]).
Quizzes will be taken outside class hours.

5 Textbooks

- Christopher M. Bishop, “Pattern recognition and machine learning”, Springer, 2006.
- T. Hastie, R. Tibshirani, J. Friedman, “The Elements of Statistical Learning: Data Mining, Inference and Prediction”, Springer Series in Statistics, 2009.
- Jiawei Han and Micheline Kamber, “Data Mining: Tools and Techniques”, Elsevier, 2012.

6 Reference Books

- V. N. Vapnik; “Statistical Learning Theory”, Wiley, 1998.
- Foundations and Trends in Machine Learning; Now Publishers Inc.
- Journal of Machine Learning Research; JMLR, Inc. and Microtome Publishing (United States).
- Conference Proceedings of ICML, NIPS, ICLR.
- Kevin P. Murphy, “Machine Learning, a Probabilistic Perspective”, MIT Press, 2012.

7 Course Requirements

You are required to attend all the lectures. If you miss any of them it is your responsibility to find out what went on during the classes and to collect any materials that may be handed out.

Class participation is strongly encouraged to demonstrate an appropriate level of understanding of the material being discussed in the class. Regular feedback from the class regarding the lectures will be very much appreciated.

8 Planned Syllabus

- Introduction to machine learning - different forms of learning; Basics of probability theory, linear algebra and optimization.
- Linear regression, ridge regression, Lasso, Bayesian regression, regression with basis functions.
- Linear Discriminant Analysis, Logistic regression, Perceptrons, Large margin classification, Kernel methods, Support Vector Machines, Classification and Regression Trees, Multilayer Perceptrons and Back propagation.
- Bayesian Belief Networks, Markov Random Fields, Exact inference methods, approximate inference methods.
- Boosting - Adaboost, Gradient Boosting; Bagging - Simple methods, Random Forest.
- PAC Learning, VC Dimension, Bias/Variance Tradeoff.
- Partitional Clustering - k-means, k-medoids; Hierarchical Clustering - Agglomerative, Divisive, Distance measures; Density based clustering - DBScan; Spectral clustering
- Apriori Algorithm; FP-Growth

9 Tentative Grading Policy

The following allocation of points is tentative. These may change during the semester.

Quizzes × 2	20-30
End Sem	40-50
Software Assignments × [2]	20-25
Total	100

10 Tentative Dates

The following allocation of dates is tentative. These may also change during the semester.

Quiz Dates	29/02/2021, 28/03/2021
Software Assignment 1 Announcement	20/02/2021
Software Assignment 1 Deadline	25/03/2021
Software Assignment 2 Announcement	21/03/2021
Software Assignment 2 Deadline	20/04/2021
End Semester Exam	TBD

11 Academic Honesty

Academic honesty is expected from each student participating in the course. NO sharing (willing, unwilling, knowing, unknowing) of assignment code between students, submission of downloaded code (from the Internet, or anywhere else) is allowed.

Academic violations will be handled by IITM Senate Discipline and Welfare (DISCO) Committee. In case of assignments, for the first instance of violation, an overlap of >75% will result in ZERO marks for all students involved. Overlap of 50% - 75% will result in 50% deduction and overlap of 25% - 50% will result in deduction of 20% of the marks originally awarded. In case of the second instance of code copying, the DISCO Committee of IIT Madras will be intimated of the matter.

Please protect your Moodle account password. Do not share it with ANYONE. Do not share your academic disk drive space on the Campus LAN.

Each proxy in the attendance will be penalized by 5% of (absolute) marks. It becomes 2% each for the donor & beneficiary, if both accept the fault.

For online exams – cameras must be ON. System being used for exam should not have any other window open – other than moodle/google-meet/google-form etc. (as specified by TAs). Any disruption of power/network link must be provided immediately after restoration of power/network by email with screen shots of mobile/system, ensuring that date/time stamps are visible.