CS6464: Concepts In Statistical Learning Theory

Jan-May Semester 2023
Slots: Mon (4:30- - 6:00 PM) and Fri (4:30-6:00 PM)

Prof. Sukhendu Das, SSB 302; Phone: 4367 Email:
sdas@cse.iitm.ac.in, sdas@cse.iitm.ac.in
TA(s): Sonam Gupta
(cs18d005@smail.iitm.ac.in),Riya Verma
(cs21D012@smail.iitm.ac.in)

Updated on Jan 31, 2023

Note: The course webpage is http://www.cse.iitm.ac.in/~vplab/statistical_learning_theory.html. A google group will be formed for course related communications; please check the emails regularly.

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 scholars who intend to specialize in this field, to 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 Online Learning, Transfer Learning etc.

* To have hands-on experience in implementing various ML and PR techniques on different datasets.

* To learn how statistical distribution in datasets affect performance of ML and PR techniques.

* To learn to compare the performance of two learning systems.

* To study a few optimization methods used to estimate the parameters of a model during training.

3   Course prerequisite(s)

CS5011 or equivalent preferable.

4   Classroom Mode

Traditional Lectures (2, 90 mins slots]). Tutorials may be taken outside class. Tutorial problems have to be solved in class. Attendance will be taken in random mode by TAs during class.

5   Textbooks

* T. Hastie, R.Tibshirani, J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference and Prediction", Springer Series in Statistics, 2009.

* V. N. Vapnik; "Statistical Learning Theory", Wiley, 1998.

6   Reference Books

* Christopher M. Bishop, "Pattern recognition and machine learning", Springer, 2006.

* Kevin Murphy, "Machine Learning, a Probabilistic Perspective", MIT Press, 2012.

* Journal of the Royal Statistical Society: Series B (Statistical Methodology).

* Foundations and Trends in Machine Learning; Now Publishers Inc.

* Journal of Machine Learning Research; JMLR, Inc. and Microtome Publishing (United States).

* Proceedings of ICML, NIPS, ICLR

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 matrial being discussed in the class. Regular feedback from the class regarding the lectures will be very much appreciated.

8   Planned Syllabus

- Learning Problem, Risk functions, Statistical Decision Theory.

- Ill posed and well posed problems.

- Least Square Regression, Bias Variance tradeoff.

  ERM (+Tikhonov Regularization), Iterative regularization by early stopping, SRM.

  Linear Models of Regression, Subset Selection methods, Shrinkage methods, Ridge regression.

- LASSO, LAR

  Bag-of-Words, Online Learning and Transfer Learning.

- ADMM, Proximal gradient
- SVM, Kernel, VC dimension, RKHS

9   Tentative Grading Policy

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

End Sem	45- 50
Mid Sem	15-20
Software Assignments $\times$ [2]	25-30
Total	100

10   Tentative Dates

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

Tutorial Dates:	TBD
Software Assignment 1 Announcement:	TBD
Software Assignment 1 Deadline:	TBD
Software Assignment 2 Announcement:	TBD
Software Assignment 2 Deadline:	TBD
Mid Semester Exam	TBD
End Semester Exam	TBD

Keep a watch on web-link for updates on dates

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 the IITM Senate Discipline and Welfare (DISCO) Committee. In case of assignments, for the first instance of violation, an overlap of greater than or equal to 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 will be intimated of the matter and will result in a drop of one-penalty level in final course grade. Please protect your Moodle account password. Do not share it with ANYONE. Do not share your academic disk drive space on the Campus LAN.

Attendance will strictly be maintained in all classes and monitored; Rules of academic section will be strictly followed for your eligibility in appearing for the End Sem exams and final grades will be released subject to such conditions as laid out by academic section rules of attendance.

For any 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.

For Makeup and Supplementary – pl, follow official protocol (check academic ordinances and email/instructions from Dean AC office). Do not disturb the teacher or TAs on this.

Plagiarism: Code submitted by you will be cross verified by an automated software tool with similarity measures among all students. Anything with greater than 50% similarity will be rejected, upto 15-20 % similarity will be allowed.