

CS 620

Formal Methods in Machine Learning

Instructor: Supratik Chakraborty
Dept of CSE, IIT Bombay

Spring 2021

Course Logistics

- Slot 12: Mon & Thurs 5.30 - 6.55 pm
- Week 1: Thurs, Jan 7, 5.05 - 5.55 pm
- Pre-recorded lectures + weekly live interactions (Slot 12)
 - Recordings for week to be made available by Mon 5.30 pm
 - No lectures during live interactions -- only doubt clearing
- Primary references:
 - Research papers
- Hands-on project
- Evaluation: Mid-sem, end-sem, paper presentation, project -- 90%
 - 10% based on quizzes (more attention-recall than deep thinking)

What this course is about

- Algorithmic techniques
 - Prove formal “properties” of ML components
 - Or show that they violate the “properties”
- ML components
 - Feed-forward neural networks
 - Recurrent neural networks
- Formal “Properties”
 - Must be mathematically precise and admit unique interpretation
 - Some properties
 - All images of cats are labeled 1
 - Two images that differ in less than 5% of pixels can't be labeled differently
 - Robot never gets into the region within coordinates (x_0, y_0) and (x_1, y_1)

What this course is **NOT** about

- Fundamentals of logic, automata theory
- Program verification
- Fundamentals of machine learning
- Push-button techniques to prove any property about your favourite neural network
- All paper-and-pencil theory

Expectations from students

- Must have taken pre-requisite courses
 - BTechs: CS228, CS213, CS337
 - MTechs: CS433/CS771, **CS725 (if you haven't taken this, talk to me separately)**
- Weekly readings of papers
 - Without this, lectures may be difficult to follow
- Attempting quizzes, even if the weightage is low
 - Those found not attempting may be penalized
- Willingness to get hands dirty with tools
- Willingness to get minds challenged with logic-based formal reasoning