

SMAI-M20-01: CSE-471: Statistical Methods in AI

C. V. Jawahar

IIIT Hyderabad

August 10, 2020

About the course: Scope

- A **fundamental** course on machine learning.
- **Not an** advanced course on ML/AI
- Emphasis on:
 - Basics Concepts, Principles, Basic Maths
 - Connect to Practical and Effective ML Algorithms
- Note: Class could be highly heterogeneous:
 - UG, PG, Working Professionals, New Students, People who have taken an ML course else where, students who use these ideas in their own research/work etc.

About the course: Getting Ready

- Exposure:
 - Maths in the form of UG Courses and Schools.
 - Programming/Algorithms: Comfortable with adaptation/hand-on.
- Maths:
 - Typical Engg Maths. Not super advanced.
 - Topics: Linear Algebra, Probability, Differential Calculus
- Programming:
 - Most tools/libraries are in python. No plan to teach python or programming. Familiarity with plotting (graphs) etc. will be useful.
 - We will use: **jupyter notebook**. If you have not used, try it out before the next class.
 - No super-heavy compute expt. planned. Some TA support will be provided.
- Infrastructure:
 - Some compute (say a laptop) advisable. Though VPN/Remote/Cloud may also work in many cases.
 - Some internet connection (say 4G) expected. Course accounts buffers to take care of unfortunate network failures.

About the course: Course Coverage

① Part 1: Basics

- Mathematical Foundations; Role of Linear Algebra and Probability; Supervised Learning Formulation and Challenges, Sample Algorithms

② Part II: Fundamental Algorithms

- Linear Methods in Machine Learning; Regression, PCA, Logistic Regression, Perceptrons, Gradient Descent, Multiclass

③ Part III: Powerful Ideas

- SVMs, Kernels, Nonlinear Methods, Ensemble, Semi-Supervised, Unsupervised and Self-Supervised Learning

④ Part IV: Neural Network Learning

- Artificial Neural Networks, MLP and Back Propagation; Intro to Deep Learning, Intro to CNN, RNN

Approximately 25% emphasis/time on each part.

Course Structure and Evaluation

- Course will be fully online
- What does it may imply?
 - You need to pay more attention to all content/input. You can't just pick up from the lectures.
 - You need to be more regular, and more disciplined.
 - Communicate and discuss technical stuff in online mode.
- How is this course (SMAI-M2020)?
 - Will try to align to the traditional models: Regular meetings, some lectures/explanations, some in person QA, etc.
 - Some specific material and background: (i) Lecture Notes (ii) Some Videos (iii) Some Links
 - Learn by Regular Work: Home Works, Assignments, In Class Reviews, Quizzes etc.

Course Structure and Evaluation

- Lectures, Tutorials (as per time table)
 - Lec: M,W,F: 9.30-10.30AM IST
 - Tut: ?? (TBD)
- Office Hours (OH) (Additional Support; Weekly (?)):
 - Logistics, Exceptions, Semi-Technical, Administrative
 - TA (One Common; One on Reserving)
- Emails:
 - Use Office Hours, if your request can wait. Get personal attention.
 - Any course specific emails: please send to: **smai.m2020@gmail.com**
- Expected course load:
 - Regular activities (Homeworks, Assignments)
 - Some preparation (20-30 mins before the lecture)
 - Reading, Thinking, Discussing
 - Meetings: Lec (*), Tut, OH (Optional)
- **If attending, attend seriously and disciplined.** This is a large class. May be not everyone was not given opportunity.

Course Structure and Evaluation

- Regular Homeworks: 40%
 - Handwritten and Programming
 - Approximately 200 Points. Your best 80% (Approximately out of 160) will be used for grading
 - Stretch beyond the class/lecture. Learning+Evaluation is the goal.
- Quiz (3): 25%
 - Closest to the traditional exams.
 - Evaluation is the primary purpose
- Assignment (3): 25%
 - Programming and Exploration
 - Exposure to the depth is the focus.
- In Class Review/Recap: 10%
 - Objective questions and regular review. Light weight. Learning is the primary objective.

Let us start with an Introductory Video

- How Can Data be Useful in Solving Problems?
- `https://www.youtube.com/watch?v=8xniRSjRyCQ&feature=youtu.be`
- Around 10 mins; Let us wait for 15 mins.

$$y = f(\mathbf{W}, \mathbf{x})$$

- \mathbf{W} is the learnable parameters.
- ML algorithms aims to learn/find \mathbf{W} from the Data $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}$
 $i = 1, \dots, N$.
- Usually only
 - a portion of the data is used for “Training” (developing the solution or computing \mathbf{W}) and
 - the rest of the data is used for “Testing” (or measuring the performance).

What Next:?

- We will use this week (until 15 Aug) for streamlining the Course:
 - New students
 - Teething problems in online mode.
- Topics:
 - K Nearest neighbour Algorithm
 - Performance Metrics of ML solutions.
- Recap of (do it yourself!):
 - Vectors and Matrices
 - Probability Distributions and Bayes Theorem
- Logistics:
 - Formal Details on Moodle
 - TAs on Board
 - Office Hours to Start
 - Mechanism for submitting regular homeworks (shiksha.iiit.ac.in)
- A good text book:
 - Mathematics of Machine Learning (pdf available)
<https://mml-book.github.io/>
 - More on Moodle