SMAI-S25-01: CSE-471-A: Statistical Methods in AI

C. V. Jawahar

IIIT Hyderabad

January 3, 2025

Announcement

There are two batches of SMAI running in parallel. Class room No. 105 and 205. Please make sure that you are in the right class.

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.
  - No super-heavy compute expt. planned. Some TA support will be provided.
- Infrastructure:
  - Some compute (say a laptop) advisable. Though Cluster/VPN/Remote/Cloud may also work in many cases.
  - Some internet connection (say 4G/5G) expected. Course accounts buffers to take care of unfortunate network failures.

About the course: Course Coverage

1. **Part 1: Fundamentals**
   - Mathematical Foundations; Role of Linear Algebra and Probability; Supervised Leaning Formulation and Challenges, Sample Algorithms; ML problem formulation; Role of Data
2. **Part II: ML Algorithms**
   - Linear Methods in Machine Learning; Regression, PCA, Logistic Regression, Perceptrons, Gradient Descent, Multiclass
   - SVMs, Kernels, Nonlinear Methods, Ensembe, Semi-Supervised, Unsupervised and Self-Supervised Learning
3. **Part III: Neural Network Learning**
   - Artificial Neural Networks, MLP and Back Propagation; Intro to Deep Learning, DL Architectures

Approximately equal emphasis/time on each part (say 7-8 lectures per part)

Course Structure and Evaluation

- Course will be fully offline and students are expected to be in the lectures (twice in a week; 20-25 in a Semester)
  - Some preparatory work expected before the lecture.
- What does it may imply?
  - Lecture/class is for learning, interacting
  - Inclass activities and grading
  - Communicate and discuss technical stuff in online mode.
- How do we grade?
  - Traditional Exams (Quiz, Mid, Final) - 40%
  - In class activities - 30%
  - Homeqoek/Quiz - 30%
  - We may have a $\pm 5 - 7\%$ change in this depending on how things progress

Problem Space

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`
- `https://tinyurl.com/ydsec8pa`
- Around 10 mins; Let us wait for 15 mins.

Problem Space

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

- $\mathbf{W}$ is the learnable parameters.
- ML algorithms aims to learn/find $\mathbf{W}$ from the Data $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}$ $i = 1, \ldots, N$.
- 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 10 Jan) 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
  - Possiblility of Office Hours
- A good text book:
  - Mathematics of Machine Learning (pdf available) https://mml-book.github.io/
  - More on Moodle