CS361: Machine Learning

Project Guideline

Ashish Anand

Professor, Dept. of CSE, IIT Guwahati

Associated Faculty, Mehta Family School of Data Sc and AI, IIT Guwahati

Objective

- Hands on experience on
  - classical ML algorithms implementation
  - using exploratory data analysis
  - entire ML pipeline

Project Administration

Project Group

- Number of Registered Students: 125

- Number of students in a group: 4

- Assign Roles to each member
  - Marks will be based on Individual Contribution

Important Timelines

- Project Group Information: 11th Feb, Sunday [11:45 PM]
- Project Proposal: 23rd Feb, Friday [11:45 PM]
- Mid-Term Report: 22nd Mar, Friday [11:45 PM]
- Final Report and Code: 19th Apr, Friday [11:45 PM]
- Presentation & Viva: 20th-28th Apr
- No extension feasible

Marks Distribution

- Project: 20-30
  - Group Work + Individual Contribution
  - Divided across several evaluation stages
  - Report across evaluation stages
  - Plagiarism Check

Project Proposal

- Introduction: Should discuss motivation, the target problem, major challenges and outline of proposed direction

- Methods: Intended methods you are planning to apply. Try to choose methods from different paradigms

- Intended Experiments

- Relevant References

- Max 2-page content, references extra

- Evaluation: Primarily based on how concise and good a report is. Report must be in your own words with proper citation.

Mid-Term Report

- Objective: More formalized version of initial problem description submission

- Presented in form of extended abstract
  - We will provide latex template

- Organize into following sections
  - Abstract, Introduction, Method, Progress, Conclusion, References

- 3-4 Pages, Max 4 Pages

- Evaluation: Based on Progress made, clarity in the work direction, report

Final Report and Presentation

- Follow a full paper format
  - Latex template will be shared again
  - 6-7 Pages

- Evaluation: Based on clarity on the problem formulation, model selection, challenges handled, analysis of reported results, and report

What I'll be expecting

- Report
  - Explain problem: definition or formulation, motivation, challenges, existing methods: adv. and disadv.
  - Explain your data: basic statistics, pre-processing
  - Explain your method: comparative study, discussion on comparative adv. of the chosen methods or project, novelty aspect
  - Explain Implementation (You may include if any specific challenges faced, or if you have used certain verifications)
  - Explain Result: Provide insights from obtained results.
  - Explain Future Scope: what next and what could be done differently
- Code
  - Well commented and readme file
  - Anybody should be able to execute and replicate results as reported in report.

How to start working on project?

- Choose theme: Environment, Economics, Education, Finance, Agriculture, Health, Infra Development etc.
- Define main question(s) you want your model to answer
  - Prediction
  - Important Features and their relative significance
- Accordingly, choose appropriate models
- Define Appropriate Experiment settings
- Analyze Results

What kind of work will be more appreciated

- Comprehensive studies
- Well reported

Relevant Resources

- Challenges [Adv: Data availability]

- Kaggle

- Papers with Code

Relevant Conferences

- ICML, WWW, KDD, NeuRIPS/NIPS

- IJCNLP, ICLR, AAAI, IJCAI

- ArXiv: Machine Learning

- ML/AI for social good