Introduction

CS771: Introduction to Machine Learning

Purushottam Kar

Course Details

- Name: CS771(A) – Introduction to Machine Learning
- Nickname: ML
- Lectures: Wednesdays and Fridays, 1800-1930 hrs, L16
- Instructor: Purushottam “Puru” Kar
- Teaching assistants: wait for next slide
- Course websites
  - Internal website: http://tinyurl.com/ml17-18ai
  - Piazza website: http://tinyurl.com/ml17-18adf (Enroll !!)
- Office hours:
  - Puru: Thursdays 1800-1930 hrs, RM509 (by appointment)
  - TA hours to be announced

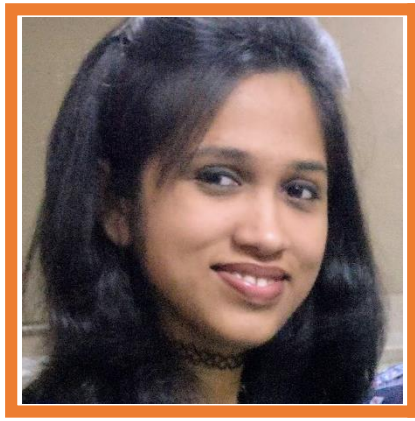
Course Mentors

Jayant Agrawal (agjayant)

Computer vision, ML

Nishit Asnani (nishit)

NLP, Reinforcement learning



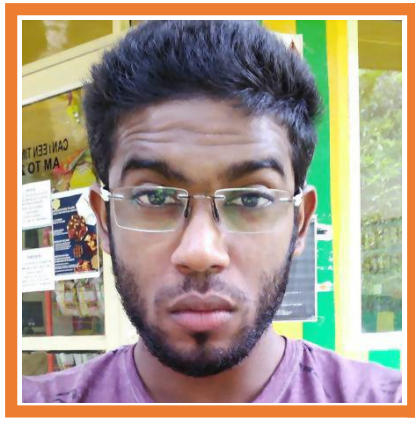
Ankita Bishnu (ankitab)

Computer vision and NLP, VQA

Shibhansh Dohare (sdohare)

NLP, Reinforcement learning

Course Mentors



Atul Gangwar (atulgang)

Information retrieval

Abhinav Garg (abhigarg)

NLP, Computer vision,VQA, Graph algorithms

Govind Gopakumar (govindg)

Optimization, ML

Purushottam Kar (purushot)

Online learning, Optimization

Course Mentors

Rohit Singh Kharanghar (rsk)

Computer vision, Deep learning

Gopichand Kotana (gopick)

Computer vision, ML

Pawan Kumar (kpawan)

Deep learning, NLP, Machine translation

Bhaskar Pratim Mukhoty (bhaskarm)

Adversarial ML, Non-convex optimization

Course Mentors

Vishak Prasad (vishak)

Bayesian ML, Deep learning, Visual recognition

Munot Rushab Preetam (rushab)

NLP, Learning with kernels

Arindam Sarkar (arindam)

Probabilistic ML, Computer vision

Susmit Wagle (waglesmi)

ML, Computer vision, NLP, Computer security

Auditors

- Please send a mail to Puru to be included in mailing list
- Auditors will have access to several aspects of the course
  - Lectures and lecture material
  - Discussion forum activities
  - Assignment and examination questions
  - Project presentations
- We regret our inability to extend the following services
  - Submit assignments and receive graded submissions
  - Appear for examinations and receive graded answer scripts
  - Take part in project groups with registered students
  - If few auditors wish to form a separate project group, talk to Puru

Grading Scheme

- Assignments – 30%
- Term project – 20%
- End semester examination - 30%
- Mid semester examination - 20%

Excellent work in term project will be very well rewarded*

Assignments

- "Pen-paper" type
  - Reasoning about ML algorithms
  - Developing extensions to ML algorithms
  - Derivations and calculations
- Must be submitted electronically (details later)
- Must be typeset in LaTeX using style file provided
  - Several resources available online – start learning today!
  - Scanned/photocopied submissions will be rejected without review
  - Use Lyx with caution (must confirm to style file)
  - Use other typesetting software at your own risk

Assignments

- Programming type
  - Develop basic ML primitives
  - Use toolboxes to solve problems
  - Develop extensions
- All submissions will be done electronically (details later)
- Language of choice: Python
  - Several resources available online – start learning today!
  - Plagiarism checks will be carried out – renaming identifiers is futile!
  - Code as well as learnt models will have to be submitted

Term Project

- Form groups of 4 or 5 – start planning today!
  - Larger or smaller groups need to justify
  - Groups must be formed by August 15, 2017 – intimate Puru
- Several project ideas will be floated – you can propose your own
  - Talk with your friends, discuss with course mentors
  - Look around you for inspiration – ML can solve a lot of problems
  - Focus on getting motivated with the problem – solutions will come later
- A formal project proposal will be due on September 15, 2017
  - Pinpoint the problem you wish to investigate
  - Give a description of existing work on that problem
  - Give a rough outline on what plan of action you wish to take

Reference Material

- No textbook for the course
- Reference material will be put up on internal website
- Locally cached copies for some references

Use of Unfair Means

- The following are prohibited – severe penalties
  - Copying answers in pen-paper assignments
  - Copying code in programming assignments
  - Passing off known results as one's own
  - Manipulating experimental results

- The following are prohibited – credit deductions
  - Using material in project report (figures, text) without acknowledging
  - Using help from auditors or mentors in projects without acknowledging

What is Machine Learning?

and where is it used?

Machine Learning

> The art and science of designing adaptive algorithms

Subject: [all] New Pizza Counter at New SAC
From: "DOSA" <dosa@iitk.ac.in>
Date: Wed, October 28, 2015 10:07 am
To: all@lists.iitk.ac.in
Cc: dosa@iitk.ac.in (more)
Priority: Normal
Options: View Full Header | View Printable Version | Download this as a file

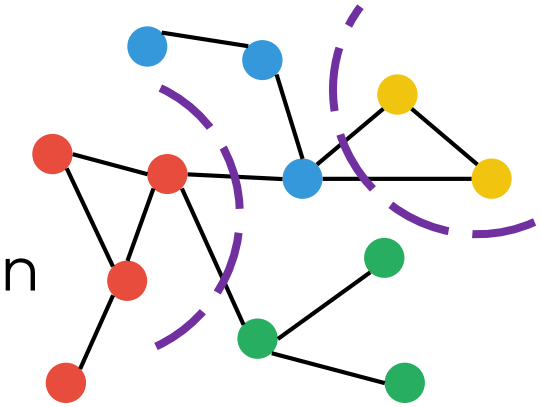
Subject: [all] Lost and Found
From: "DOSA" <dosa@iitk.ac.in>
Date: Wed, October 28, 2015 10:07 am
To: all@lists.iitk.ac.in
Cc: dosa@iitk.ac.in (more)
Priority: Normal
Options: View Full Header | View Printable Version | Download this as a file

Machine Learning

> The art and science of designing adaptive algorithms

How to identify ML applications

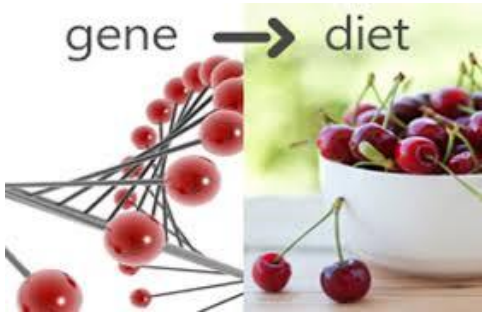
- Complexity: no “closed form” solutions
  - Human “experts” cannot successfully prescribe a solution
  - Graph cut problem not a good candidate for ML

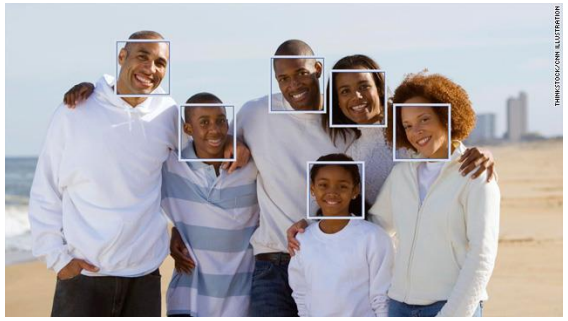


- Presence of immense variety
  - Too many variants to be solved independently
  - Recommendation systems excellent ML problems

- Need for automation
  - Scalability and speed are main criterion
  - Do we need to automate medicine, driving?

Prominent ML applications

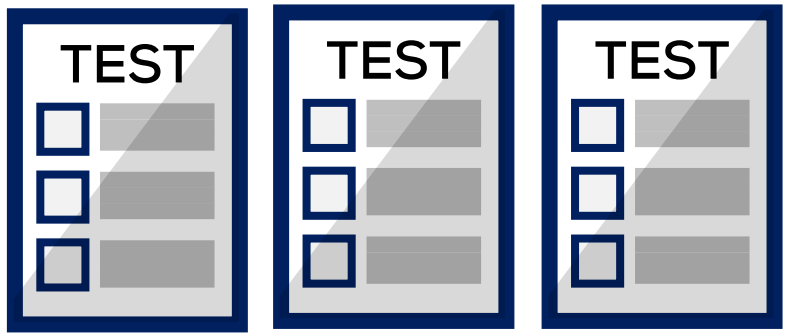




How is ML done?

the ultra short version

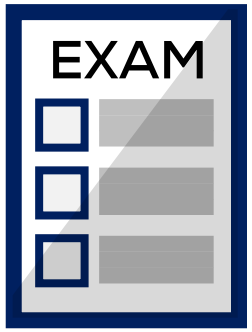
A typical exam prep cycle



Practice Tests

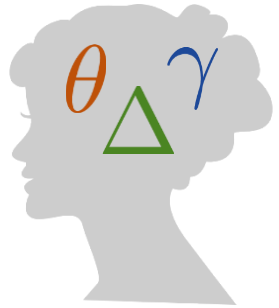
Preparation

Actual Exam



Pass / Fail

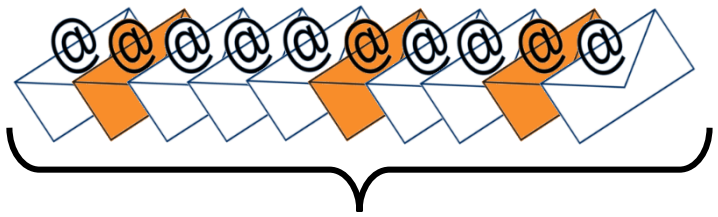
Subject Knowledge



Spam Filtering

Spam Filtering

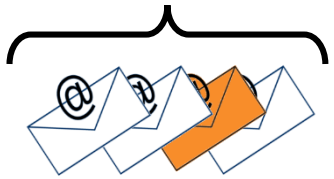
A typical ML workflow



Existing Emails

ML Algorithm

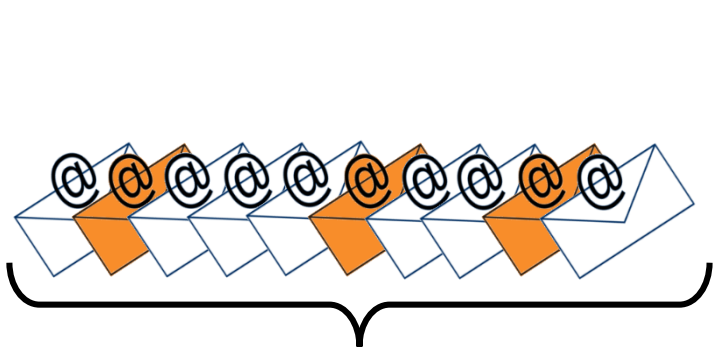
New Emails



Spam Filter

Spam/Non-spam

A typical ML workflow



Training Data

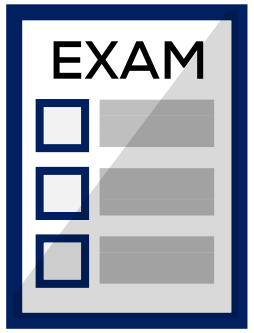
ML Algorithm

Test Data

Model

Output

ML as an "examination"

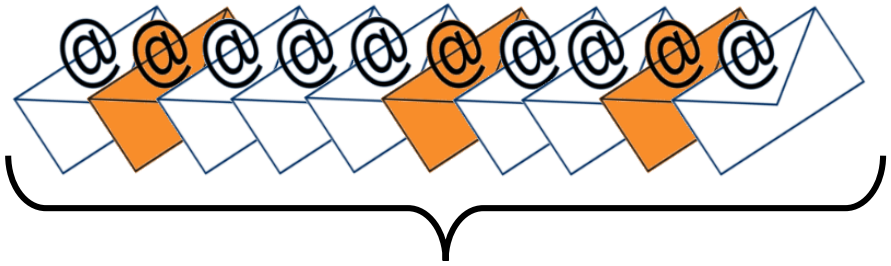


- Critical to do well on D-day
- Prep test results indicative
- No out-of syllabus questions
- Should not leak exam paper

- Critical to do well on test data
- Training accuracies indicative
- Training/test data are similar
- Should not look at test data

Many other similarities – future discussions

Input driven ML



Training Data

Batch Learning

Active Learning

Robust Learning

Semi-supervised Learning

Supervised Learning

Online Learning

Unsupervised Learning

Reinforcement Learning

Output driven ML

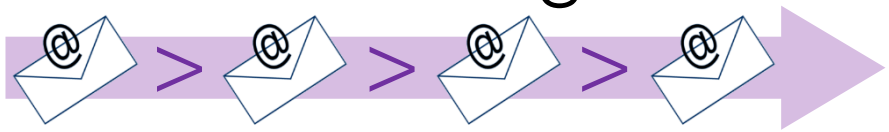
Test Data

Regression

```
Subject: [**SPAM**] Free movie tickets every month
X-Barracuda-Spam-Score: 4.89
```

Topic Modelling

Ranking



Multi-classification

Primary Social Promotions

Updates Forums

Tagging

URGENT, OFFICIAL, TAX

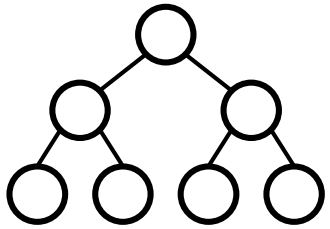
Binary Classification

```
Subject: [**SPAM**] Free movie tickets every month
X-Barracuda-Spam-Status: Yes
```

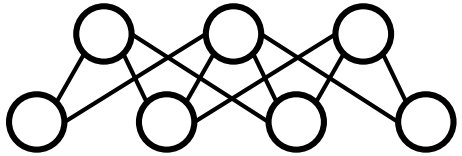
Process driven ML

ML Algorithm

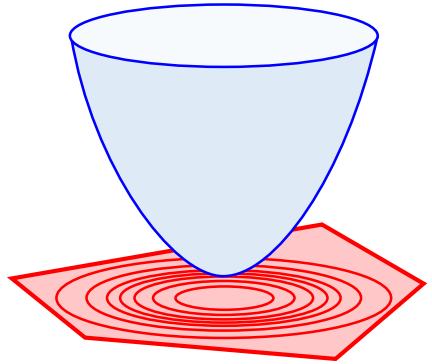
Local Methods



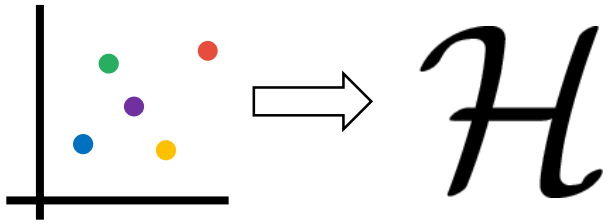
Neural Methods



FA Methods



Kernel Methods



Probabilistic Methods