Introduction to Machine Learning

Course Logistics

Varun Chandola

Computer Science & Engineering
State University of New York at Buffalo
Buffalo, NY, USA
chandola@buffalo.edu
Class Details

- Lecture Information
  - Monday, Wednesday, Friday (9.00 - 9.50 AM)
  - 109 Knox

- Recitations
  1. 10.00 - 10.50 AM Monday, *Norton 210*
  2. 01.00 - 01.50 PM Tuesday, *Bell 337*
  3. 08.00 - 08.50 AM Friday, *Cooke 127a*

- Recitation topics are listed in the syllabus
  - No recitation this week.

- Class web page
  - http://www.cse.buffalo.edu/~chandola/machinelearning.html
  - https://piazza.com/buffalo/spring2018/cse474/home
Varun Chandola
- http://www.cse.buffalo.edu/~chandola
- Email: chandola@buffalo.edu
- Office: 304 Davis Hall
- Phone: (716) 645-4747

Office Hours: 1.00 PM - 3.00 PM (Mondays)
Teaching Assistants

▶ Xin Ma
  ▶ Email: xma24@buffalo.edu
  ▶ Office Hours: 10.00 AM - 11.00 AM (Fridays)

▶ Rudra Prasad Bakshi
  ▶ Email: rudrapra@buffalo.edu
  ▶ Office Hours: 11.30 AM - 12.30 PM (Wednesdays)

▶ Hongfei Xue
  ▶ Email: hongfeix@buffalo.edu
  ▶ Office Hours: 4.00 PM - 5.00 PM (Mondays)
Piazza

- Primary medium of communication
- All announcements, teaching notes, slides, polls, etc. will be made available through Piazza.
- Questions?
  1. General post to all (*Name will be visible*).
     - Choose appropriate folder.
  2. Private post to instructor, TA.
- Interact.

**Piazza Incentive**

- Top 3 contributors (questions or answers) will get recognized
- Award - *To be decided*
Topics Covered

Theoretical Machine Learning

- Concept Learning
- Mistake Bound Online Learning
- Vapnik-Chervonenkis Dimension
- PAC Learning
- Statistical Learning Theory

Machine Learning Tools

- Bayesian Inference
- Expectation Maximization
- Optimization

Machine Learning Algorithms

- Linear Regression
- Linear Classification
- Neural Networks
- Support Vector Machines
- Kernel Methods
- Latent Space Models (PCA)
- Mixture of Models
- Bayesian Networks
Topics Covered

Theoretical Machine Learning

- Concept Learning
- Mistake Bound Online Learning
- Vapnik-Chervonenkis Dimension
- PAC Learning
- Statistical Learning Theory

Machine Learning Tools

- Bayesian Inference
- Expectation Maximization
- Optimization
### Topics Covered

#### Theoretical Machine Learning
- Concept Learning
- Mistake Bound Online Learning
- Vapnik-Chervonenkis Dimension
- PAC Learning
- Statistical Learning Theory

#### Machine Learning Algorithms
- Linear Regression
- Linear Classification
- Neural Networks
- Support Vector Machines
- Kernel Methods
- Latent Space Models (PCA)
- Mixture of Models
- Bayesian Networks

#### Machine Learning Tools
- Bayesian Inference
- Expectation Maximization
- Optimization
Topics Not Covered

- Deep learning
- Reinforcement Learning
- Probabilistic Graphical Models and Bayesian Networks
- Decision Trees
- Association Analysis
- Applications of Machine Learning (See Piazza post)
Textbooks

- No prescribed text
- Primary references
- Optional reading list

![Textbooks](image-url)
Grading

- **Grading Scheme**
  - Short weekly quizzes using Gradiance (12) – 20%
  - Programming Assignments (3) – 30%
  - Mid-term Exam (in-class, open book/notes) on 03/16/2018 – 20%
  - Final Exam (in-class, open book/notes) on 05/16/2018 – 30%

- All components will be individually curved

- **Final grade:**
  - A: [92.5, 100]
  - A-: [87.5, 92.5]
  - B+: [82.5, 87.5]
  - B: [77.5, 82.5]
  - B-: [72.5, 77.5]
  - C+: [67.5, 72.5]
  - C: [62.5, 67.5]
  - C-: [57.5, 62.5]

- **Use UBLearns for all electronic submissions**
Mid-term and Final Exams

- All material covered in class
  - Final exam will not be comprehensive
- All multi-choice objective problems
- No partial credit
Gradiance

- An online quiz system
- One quiz per week released on Monday by 8.59 AM and due next Sunday by 11.59 PM
- 3 - 4 multiple choice problems about topics covered that week
- A warm up quiz (ungraded) is posted
- 5-minute delay between successive submissions
- Only 3 tries allowed, maximum score will be used
- Every wrong answer will result in 1 negative point per try

Gradiance Enrollment

- Go to http://www.newgradiance.com/services
- Register and use the class token FC4761F5
- Make sure you register using your UBIT name as the username
- No other username will be accepted
Python

- All programming assignments and class demonstrations using Python
- Resources:
  - Installing python, ipython
  - Python IDE - Canopy
  - More about ipython notebooks
  - Python for Developers, a complete book on Python programming by Ricardo Duarte
  - CodeAmerica - Python
  - An introduction to machine learning with Python and scikit-learn (repo and overview) by Hannes Schulz and Andreas Mueller

Github Repo

- https://github.com/ubdsgroup/ubmlcourse
- http://nbviewer.ipython.org/github/ubdsgroup/ubmlcourse/tree/master/notebooks/
Online student response system
  ▶ Random number generator!
▶ http://m.socrative.com/student/
▶ Enter class ID - 259432
▶ Optional
Academic Integrity and Honor Code

▶ http://www.cse.buffalo.edu/shared/policies/academic.php
Machine Learning Honor Code

- Against the ML honor code to:
  1. Collaborate on Gradiance quizzes
  2. Collaborate or cheat during exams
  3. Submit someone else’s work, including from the internet, as one’s own for any submission
  4. Misuse Piazza forum

- You are allowed to:
  1. Have discussions about homeworks. Every student should submit own homework with names of students in the discussion group explicitly mentioned.
  2. Collaborate in groups of 2 or 3 for programming assignments. One submission is required for each group.

- Violation of ML honor code and departmental policy will result in an automatic F for the concerned submission

- Two violations ⇒ fail grade in the course
Checklist and Resources

1. Sign-up for Piazza
2. Sign-up for Gradiance, try warm-up quiz
3. Read the department’s academic integrity policy

Resources

- Piazza - piazza.com/buffalo/spring2018/cse474/home
- Video Channel - TBA
- Course slides and handouts - www.cse.buffalo.edu/~chandola/machinelearning.html
- Github Repo - github.com/ubdsgroup/ubmlcourse
- Notebooks - nbviewer.ipython.org/github/ubdsgroup/ubmlcourse/tree/master/
Warmup Exercise

A fair coin

- Probability of heads?
- 5 heads in a row?
- 5\textsuperscript{th} head after seen 4 heads in a row?
Warmup Exercise

A fair coin

- Probability of heads?
- 5 heads in a row?
- 5th head after seen 4 heads in a row?
- Gambler’s Fallacy
A fair coin

▶ Probability of heads?
▶ 5 heads in a row?
▶ 5\textsuperscript{th} head after seen 4 heads in a row?
▶ \textit{Gambler’s Fallacy}
▶ If I know that probability of two people bringing a bomb on a plane is very low, should I bring a bomb along to make myself safer?
## Warmup Exercise

### A fair coin
- Probability of heads?
- 5 heads in a row?
- 5\(^{th}\) head after seen 4 heads in a row?
- *Gambler’s Fallacy*
- If I know that probability of two people bringing a bomb on a plane is very low, should I bring a bomb along to make myself safer?

## Matrix Vector Products
- Let \([3, 4]\) denote a vector in a 2D space
- Multiply with a number?
  \[
  2 \begin{bmatrix} 3 \\ 4 \end{bmatrix}
  \]
- Multiply with a matrix?
  \[
  \begin{bmatrix} 2 & 1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix}
  \]
- For a matrix, find a vector such that matrix-vector product \(\equiv\) scalar-vector product.