Introduction to Course

Machine Learning for Data Science

Muhammad Atif Saeed

Ph.D Scholar (Computer Science), MS (Data Science)

Department of AI & DS

Slides Credit: Dr. Akhtar Jamil

Instructor Information

Instructor	Muhammad Atif Saeed	E-mail	Atif.saeed@isb.nu.edu.pk
Office	D – 111A		

Course Information

Course Code	DS-5004
Course Title	Machine Learning for Data Science
Credit Hours	3

Prerequisite(s)

- There is no prerequisite for this course.

- However, having good programming skills will be an added benefit especially with one of the following tools:
  - Programming Skills
  - Python or C++
  - Scikit-Learn
  - Keras or Tensorflow or PyTorch
  - Or any other framework related Machine Learning

Goals of this Course

- This course is designed for students to gain theoretical and practical knowledge of Machine Learning.

- Course will cover from basics of Machine Learning to advanced level algorithms.

- It will covers algorithms, theory and practical applications

- Provide guideline for doing research

- How to compose a technical paper and present it.

Course Learning Outcomes

- Understanding various algorithms of Machine Learning.

- Given a real work problem, analyze the problem and design a machine learning pipeline to address the problem.

- Be able to implement the algorithms in any programming language.

Grading Policy

- Absolute grading
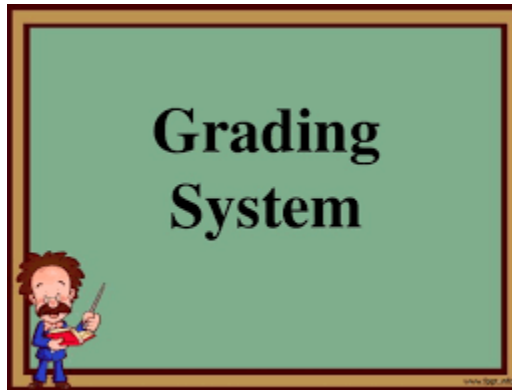
Grading Policy

Assessment Item	Number	Weight (%)
Assignments	4	10
Quizzes	4	10
Presentation	1	5
Paper Writing	1	10
Sessional-I	1	12.5
Sessional-II	1	12.5
Final Exam	1	40

Explanation of Assessment

- All assignments and Quizzes carry equal weightage

- Enough Time will be given for Assignments for their implementation

- Quizzes can be announced or surprise

Missed Assessment

- Retake of missed assessment items (other than sessional/ final exam) is NOT allowed.
- Missed assessment item (other than sessional / final exam) will earn zero marks
- Late submission will NOT be accepted .
- For missed sessional/ final exam due procedure will be followed.
- No change is any deadline

Course Plagiarism Policy

- Plagiarism in any kind of assessment including project or sessional/ final exam, assignments quizzes, will result in F grade in the course.

- Plagiarism is copying the work of others and present it as yours.

- Properly cite the work of others

Dishonesty, Plagiarism

You can fool some of the people all of the time,
and all of the people some of the time,

but you can not fool all of the people all of the time.

Abraham Lincoln,
16th president of US (1809 - 1865)

Attendance policy

- Students are supposed to have 100% attendance.

- The minimum attendance requirement at all levels and in all courses is 80%.

- The relaxation of 20% attendance has been given only to cover any planned events or unforeseen situations.

- I will take attendance at the start of lecture

- Anyone reaching after 10 minutes will be marked as Absent.

Text Book

- *No Text Book*

Reference Book(s)

1. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 2nd Edition by Aurélien Géron

2. Neural Networks and Learning Machines 3rd Edition by Simon Haykin

3. The Elements of Statistical Learning by Jerome H. Friedman, Robert Tibshirani, and Trevor Hastie

4. Deep Learning (Adaptive Computation and Machine Learning series) Illustrated Edition by Ian Goodfellow (Author), Yoshua Bengio (Author), Aaron Courville

5. Pattern Recognition and Machine Learning (Information Science and Statistics) by Christopher M. Bishop

https://libgen.is/

Course Materials

- Will be shared via Google Class Room

Class Code

lsmdbirp

Paper Writing

- All students must select a research topic and write a complete paper in IEEE format.

  - Both word and Latex IEEE formats can be obtained from following link

  - https://www.ieee.org/conferences/publishing/templates.html

Paper Writing

- A typical paper must consists of following main section
  - Abstract
    - Should be one paragraph explaining what is being presented in the paper
  - Introduction
    - A short explanation of the problem and reviews of the existing research.
    - You must download and review at least 15 papers (Journal Papers are highly encouraged)
  - Data Set
    - Which data has been used for this study
  - Methodology
    - What method(s) has been implemented
  - Experimental Results
    - Explanation of the results obtained. Should include figures, tables etc.
  - Conclusion
    - Summarize what you did overall
  - References
    - Bibliography of which papers were reviewed or used in the paper

Reading Papers

- Every week students must read a research paper related to their topic and write a short report.
  - Your weekly work will directly contribute to your assessment items.
- Particularly following items are required for discussion
  - Methodology of the paper
  - Main Contributions
  - Experimental Results
  - Strong points and weak points

Presentation

- Each student will be given 10-15 minutes for their presentation

- Students will explain every step they have performed for preparing their paper and the results obtained.

- Implementation must be shown to the instructor

Programming resources

- https://scikit-learn.org/stable/

- https://docs.microsoft.com/en-us/cognitive-toolkit/

- https://www.tensorflow.org/

- https://pytorch.org/

- https://www.nltk.org/

High Impact Journals/Conferences

- High Impact Conferences
  - https://www.scimagojr.com/journalrank.php?area=1700&type=p&category=1702

- High Impact Journals
  - https://www.scimagojr.com/journalrank.php?category=1702&area=1700&type=j

Tentative Schedule

Week	List of Topics
1	Introduction to course and Fundamentals of Machine Learning
2	Introduction to Linear classifier: Linear Regression
3	Logistic regression
4	Introduction to Neural Networks
5	Multi-Layer Perceptron
6	Sessional-I

Tentative Schedule

7	Support Vector Machines	3
8	Decision Trees	3
9	Ensemble Learning and Random Forests	3
10	Unsupervised Learning: Clustering	3
11	Unsupervised Learning: Dimensionality Reduction	3
12	Sessional-II	

Tentative Schedule

13	Introduction to Deep neural networks	3
14		3
15		3
16	Presentations and Discussions	3
17	Presentations and Discussions	3

It's Your Turn

- To Introduce Yourself
  - Your Name
  - Your Experience So Far At Fast?
  - Your Suggestions/Expectations?

netiquette – Brown Car Guy

Let's Start …..☺

Lecture #1-1

Goals

This lecture will cover:

- Why Machine Learning?
- Development Environment Setup/ Mendeley/Overleaf
- Research Stages

Where are we going?

Job Title

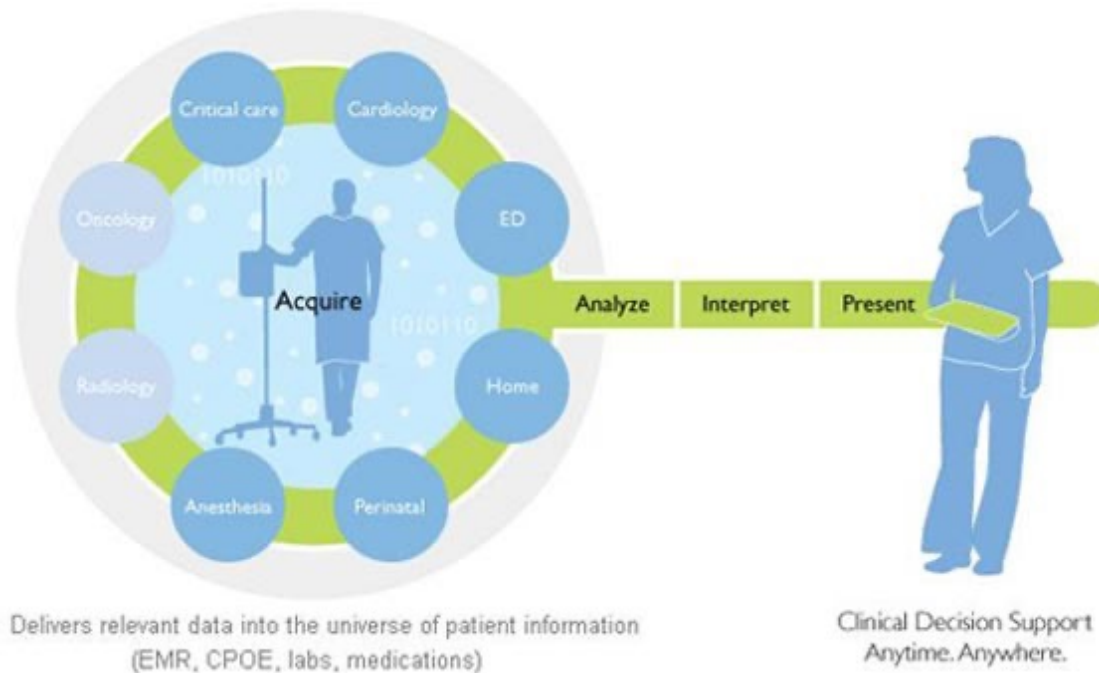|	
#1	Enterprise Architect
#2	Full Stack Engineer
#3	Data Scientist
#4	Devops Engineer
#5	Strategy Manager
#6	Machine Learning Engineer
#7	Data Engineer
#8	Software Engineer
#9	Java Developer
#10	Product Manager
#11	Back End Engineer
#12	Cloud Engineer
#13	HR Manager
#14	Business Development Manager
#15	Information Security Engineer
#16	Physician

Machine Learning in Action
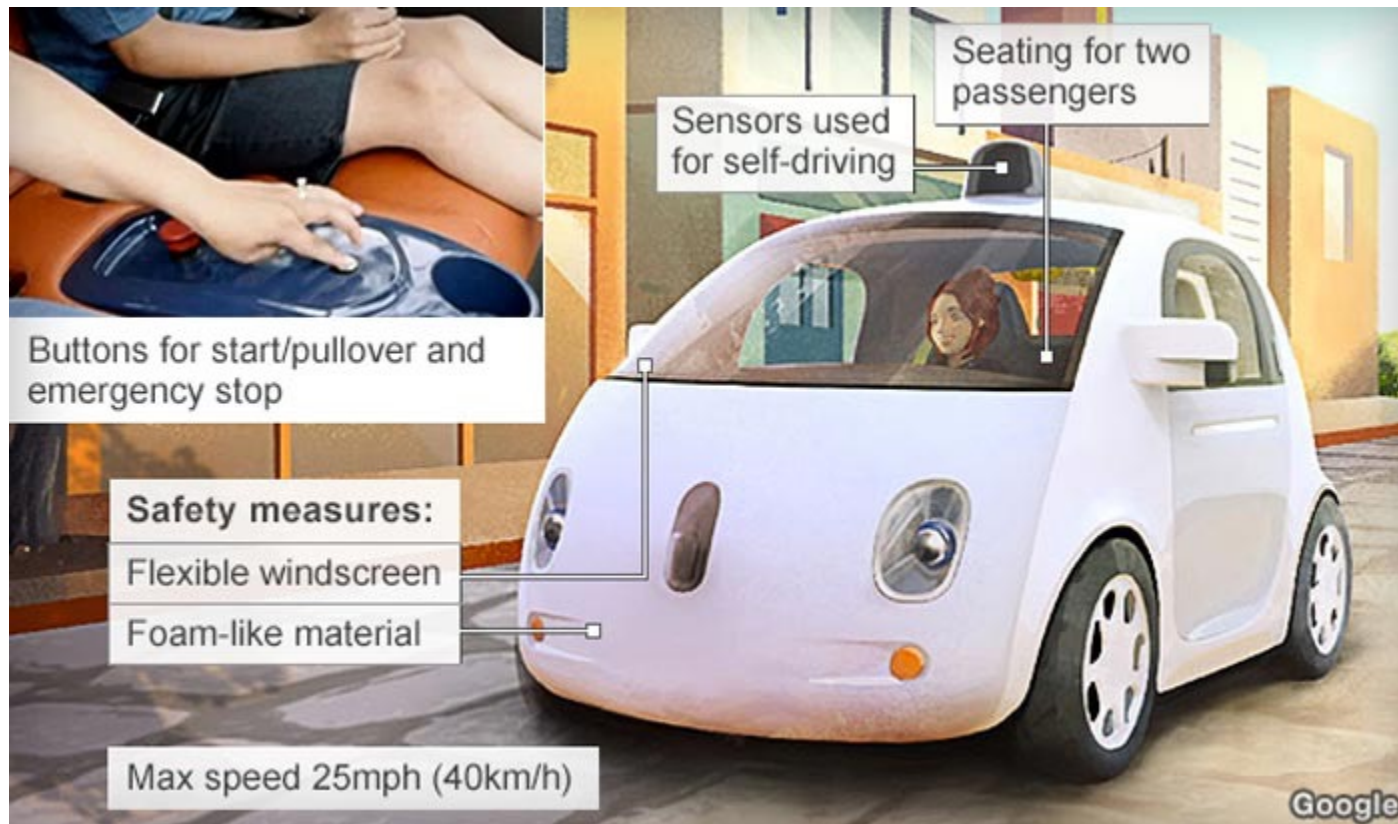
Machine Learning in Action

- Computers learning from medical records which treatments are most effective for new diseases

Delivers relevant data into the universe of patient information (EMR, CPOE, labs, medications)

Clinical Decision Support
Anytime. Anywhere.

http://www.healthcare.philips.com/main/products/hi_pm/products/clinical_support.wpd

Machine Learning in Action

- Houses learning from experience to optimize energy costs based on the particular usage patterns of their occupants

Machine Learning in Action

- Helicopters can learn aerial tricks by watching other helicopters perform the stunts first

Machine Learning in Action

- Document Classification

Sports
Science
News

Machine Learning in Action

- Stock Market Prediction

Machine Learning in Action

- Weather Prediction

Why ML?

- Data is everywhere and is generated at high speed.
  - Example: Every second, on average, **around 6,000** tweets are tweeted 350,000 tweets sent per minute, 500 million tweets per day and around 200 billion tweets per year.
- Writing a computer program with if-else conditions to solve a problem including lot of data is practically not feasible
- Machine Learning algorithms can provide a better solution
  - Understand the data and make prediction on unseen data

Why ML?

- Traffic Alerts (Maps)
- Spam Detection
- Social Media
- Products Recommendations
- Speech Recognition
- Speech to Text Conversion
- Text to Speech Conversion

Why ML?

- Fraud Detection
- Sentiment Analysis
- Chatbots
- ML for health care
  - Detection and diagnosis
  - Discovery of new drugs
- Security Purposes
  - Cyber threat detection
  - Weapon Detection

Why ML?

- Faults Detection
- Character recognition
- Automatic License Plate Recognition
- Fingerprint Recognition
- Face Recognition
- Autonomous Car driving
- Object Detection and Tracking

Development Environment Setup

- Download and Install Anaconda
  - https://www.anaconda.com/
  - Individual Edition is Free

- Install Scikit-Learn
  - https://anaconda.org/anaconda/scikit-learn

Mendeley

- Mendeley is a company, which provides products and services for academic researchers.
- It is most known for its reference manager
- To manage and share research papers
- Generate bibliographies for scholarly articles
- Download : https://www.mendeley.com/

Overleaf

- The easy to use, online, collaborative LaTeX editor.

- Overleaf is used by over nine million students and academics at 6,800 institutions worldwide

- Work from anywhere

- https://www.overleaf.com

Getting Started with Research

- Step 1. Select a Topic
  - Your interests and Skills
  - Supervisor / Project
  - Availability of resources
- Step 2. Search for Relevant Information
  - IEEE Explore
  - Springer
  - Scopus
  - Web of Science
  - Google Scholar

Getting Started with Research..

- Step 3. Organize Resources
  - Zotero
  - EndNote
  - Mendeley
  - JabRef
  - License
  - MarginNote
- Step 4. Write Paper
  - Outline the Paper
  - Take Notes
  - Avoid Plagiarism

Reference

- https://www.glassdoor.com/

- https://www.dsayce.com/

- Neural Networks and Learning Machines 3rd Edition by Simon Haykin

Thank You ☺

Presented by Muhammad Atif Saeed