

Lecture-1

September 2, 2020

1 Welcome to Data 602: Introduction to Data Analysis and Machine Learning!

1.1 Today's Agenda:

- Accessibility
- Let's get to know each other more: Ice Breaker
- Set some ground rules especially for this new environment
- Assignments and homeworks

1.2 Accessibility

- Remote setting might create challenges and barriers for some of you.
If this is the case reach out to me at mguner@umbc.edu
or SDC (Student Disability Center) from disAbility@umbc.edu or (410) 455-2459.

Some of the concerns might be but not limited to:

- Internet speed
- Not having available space
- If using video/audio is not available to you
- We will have in-class surveys, quizzes etc. if you need some accommodation
- In this class, I am planning to have some group work

1.3 Ice Breaker:

- Introduce yourself and tell us a funny/interesting thing about yourself!
– I go first :)

1.4 Let me know you more!

- Take the following survey:

[Data Science Toolkit](#)

1.5 Ground Rules for Remote Setting

- Be Kind
- Be Inclusive
- Be Constructive
- Be Active
- Be Professional

1.6 Course Logistics

- Lectures will be in two parts:
 - 4:30 - 5:30
 - Break 15 mins
 - 5:45 - 7:00

1.7 During the Lectures

- Please mute yourself if someone else is speaking!
- If you have a question please ask them when I get the questions.
- Please use the chat for only class-relevant discussions (i.e. Please don't write distracting side stories in the chat.)
- Lecture materials (slides, notebooks, recordings will be shared with you after the lectures)
- Let me know immediately if:
 - You cannot hear me,
 - You cannot see the screen I am sharing
 - Font size is too small to read.

1.8 Grading

Assignment	Percentage
Attendance/Engagement	15%
Homework	30%
Midterm Project	30%
Final Project	30%

1.9 Attendance/Engagement

What counts as being present:

- You show up at most 15 mins after class started.

What counts as being engaged:

- Readings before class.

- Take quizzes and answer questions successfully.
- Answer questions, make comments, participate.

1.10 Homework

- 3 Mini Projects. Due Dates - Tentative topics:
 - Week-4: EDA with Python and Linear Regression.
 - Week-10: A project with supervised/unsupervised problems.
 - Week-13: A project with deep learning models.

1.11 Homework - Deliverables

- Github repo:
 - Code (clean, modular, reproducible, etc.)
 - ReadMe (with clear details on project goals, tools and directions)
- Blogpost

1.12 Projects

- Think of them as bigger homework.
- You can build your projects on previous homework.
- 2 projects: 1 for midterm and 1 for final.
 - Project-I: Due Week-7
 - Project-II: Due the last week of the classes.

Projects - Deliverables

- Github: Code, ReadMe, Presentation, Technical Notebook
- Blogpost.

1.13 Books

- [ISLR](#)
- [Python Machine Learning](#)
- [Hands-on Machine Learning](#)

1.14 Other Books Might be Helpful

- [Deep Learning Book](#)
- [Pattern Recognition and Machine Learning](#)
- [Elements of Statistical Learning](#)

1.15 Tools

- [umbcdatasci - slack](#)
- [umbcdatasci #data602_fall_2020](#)

- [Jupyter Notebooks](#)
- [Github](#)
- [Anaconda Individual Edition](#)
- [Visual Studio](#)

1.16 Any Questions?

1.17 PART - II

1.18 What is this course about?

- Machine Learning - sklearn
- Deep Learning - Tensorflow

1.19 Machine Learning and Scikit-Learn

- What is Machine Learning and its place in AI, data science framework
- Types of problems in machine learning - supervised, unsupervised, semi-supervised, reinforcement etc.
- Terminology: Algorithm, optimization, cost function, evaluation, training, parameters, hyper-parameters, over-fitting, under-fitting, bias, variance, random variables, prior, posterior etc.
- Commonly used algorithms in machine learning: Lasso, trees, support vector machines, logistic regression, k-means, PCA, T-SNE, kernel-techniques, regularization techniques etc.

1.20 Deep Learning and TensorFlow

- What is deep learning and why it is a big deal?
- Some problems that deep learning is very powerful: Computer vision, machine translation, speech-recognition, Representation Learning, text and speech generation.
- Challenges of deep learning: Vanishing gradient, Exploding gradient, computational challenges, implementation challenges etc.
- Commonly used architectures in deep learning: FNN, RNN, GNN, CNN, LSTM etc.
- Terminology: Neuron, networks, activation, loss, validation, learning-rate, initialization, relu, sigmoid, stochastic process, gradient descend, backpropagation, memory, gate, encoder, decoder, kernel, filter etc.

[Quiz: What can AI do?](#)

1.21 Readings - Discussion - Breakout Rooms

[Andrew Ng - What is Machine Learning](#)

[IBM - What is ML](#)

[Andrew Ng - AI is new electricity](#)

1.22 Why now? - Discussion

1.23 Some examples

- [Quick Draw](#)
- [Another Drawing](#)
- [Deep Fake Video](#)
- [image captioning](#)
- [tinkering with neural networks](#)

1.24 Extra Resources:

[IBM - AI-ML-Deep_Learning-NN](#)

[Drawbacks of Deep Learning](#)

[AI-Karaoke](#)

[Google Assistant - Hair Cut - Starts 1:11](#)