Lecture 1

BU.330.775 Machine Learning

Minghong Xu, PhD.
Associate Professor

Today's Agenda

- Course Overview
  - Requirements and evaluation
- Introduction to Machine Learning
- Programing Basics
- Hands-on Learning: Python tutorial

Instructor Bio

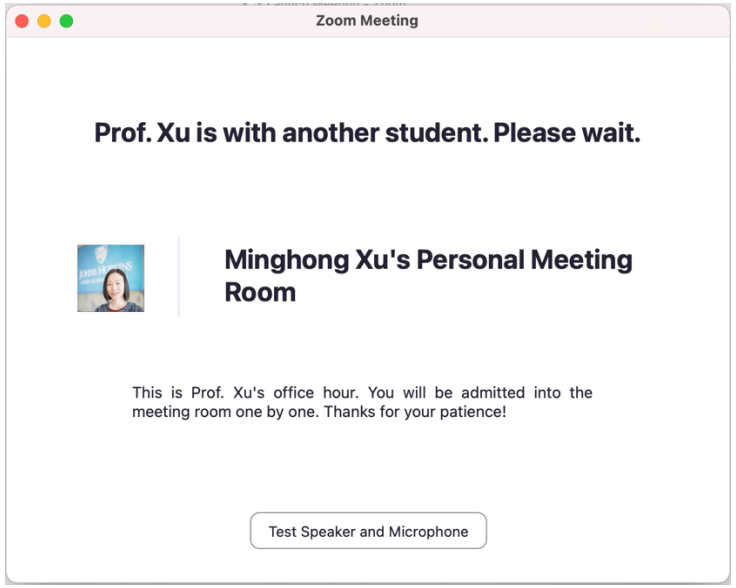
- Minghong Xu, Ph.D.
- Email: xu.minghong@jhu.edu
  - Best way to connect
- Affiliation: Center for Digital Health and Artificial Intelligence (CDHAI)
- Research and teaching: Artificial Intelligence and Big Data

Offices Hours

- Wednesdays 11:30am-1:30pm, and by appointment

https://jhucarey.zoom.us/j/4658557490?pwd=Y2NvL0M0RjdFb3RpUjlVOFBSSkFLZz09

- Link available on Canvas
- Waiting room enabled, admit one by one
- First come first served
  - unless book ahead of time



Teaching Assistant

- Boxi Jiao
- Email: bjiao1@jh.edu
- MSIS, Carey 2024

Course Expectation

- We focus on **both** design (theory) and deployment (practice)
  - How does it work, how can you apply

- *Why you need to know the technical details of ML models?*
- Programming and analytics made easier with AI
  - You will learn why in Generative AI course
- Prepare you for deep learning and generative AI

Class Logistics

- Two 10-min-breaks each class

- 50/50 split between theory and practice
- Hands-on learning using Python every week except the last week

Recommended Textbooks

- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems
- Introduction to Machine Learning with Python: A Guide for Data Scientists

- Most class examples are curated from these two books
- Check the original code on GitHub

About Python

- Top language for AI/ML

- *No experience before?*
- No worries!
- Build skills throughout the course
- Generative AI tools make programming and learning programming much easier before, “AI-assisted Programming”
  - We will demo it in class

Classroom Policy

- No cellphone
- No excessive talking
  - Questions during lecture should be addressed to me
  - Suggest to sit closer to me
- Academic honesty (important!)
- Rectifying your score
  - Discuss with TA or me your concerns on grading asap, within 1 week after scores are posted
  - Request after 1 week may not be entertained

Requirements

Assignment	Weight
Attendance and participation in class discussion	5%
Homework	60%
Final Exam	35%
Total	100%

Class Participation (5%)

- **Please use your name tent**

- Class participation is an important part of learning
- Discussion questions every week

- Expect cold calls
  - First row exempted
  - Last row having the highest chance

Assignments (60%)

- 6 weekly assignments
- Each of 10 pt
- Instructions in each week's hands-on learning session

- Due the next week before class
- Subject to TA interviews
- Responsible for your submission

Final Exam (35%)

- “Interview” type of questions
  - Descriptive
  - Understanding

- Python questions in format of multiple choices
  - No syntax (write code)
  - Understanding only

Final Exam Format

- In week 8, closed-book
- Administered via Respondus LockDown Browser

- Install LockDown Browser from https://download.respondus.com/lockdown/download.php?id=123533816
- Sample test will be posted on Canvas for you to test

Tentative Schedule

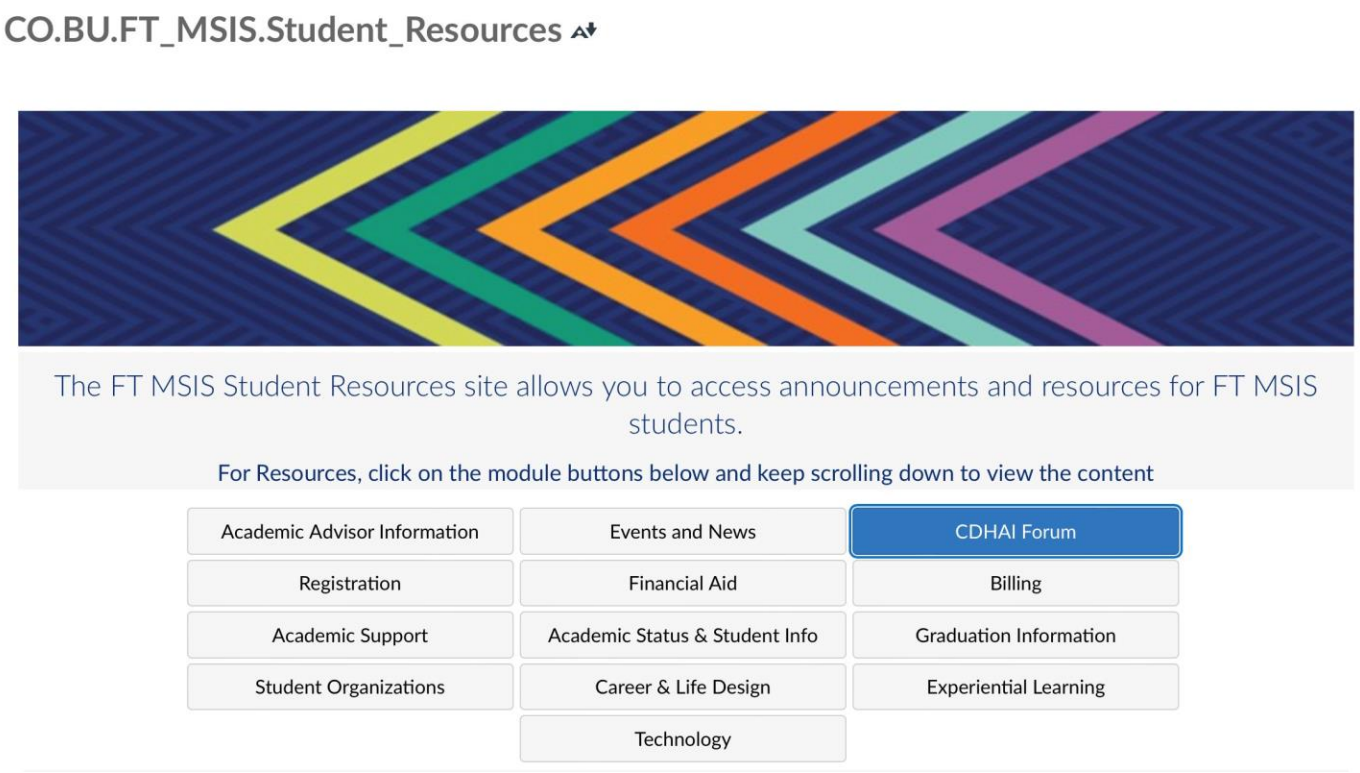
Week	Topic	Hands-on Learning	Due
1	Introduction to machine learning and programming basics	Python tutorial	HW 1 release
2	Data preprocessing and exploratory data analysis	Explore and clean a housing dataset for machine learning algorithms	HW 1 due HW 2 release
3	Design of supervised machine learning models and training (I), regularization and evaluation	Deploy supervised models on MNIST dataset	HW 2 due HW 3 release
4	Design of supervised machine learning models and training (II), ensemble and boosting	Deploy supervised models on breast cancer dataset	HW 3 due HW 4 release
5	Design of unsupervised machine learning models (I), dimensionality reduction and feature engineering	Deploy unsupervised models on breast cancer dataset	HW 4 due HW 5 release
6	Design of unsupervised machine learning models (II), clustering and applications	Deploy unsupervised models on MNIST dataset	HW 5 due HW 6 release
7	Reinforcement learning and final review	Reinforcement learning using gymnasium	HW 6 due
8	Final exam		

Final Grades

- I may curve up or curve down at the end

- A/A-: 20%
- B+: 60%
- B and below: 20%

Extra Credit

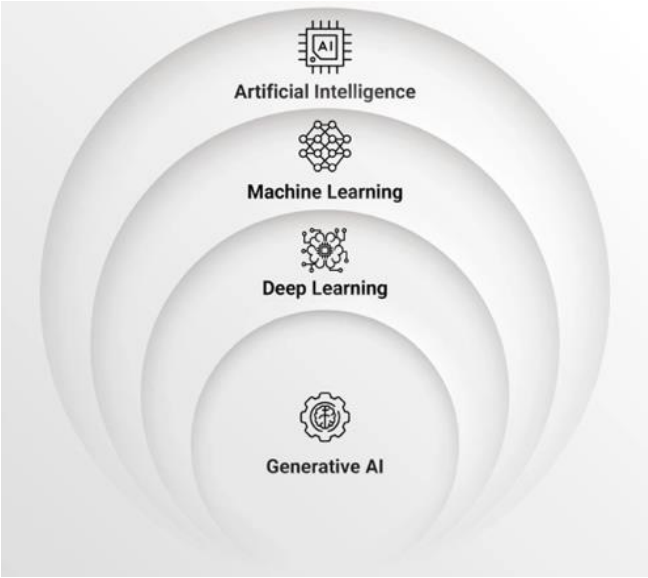


- 1pt for each qualified article
- Need guidance?
- You chance to list a research experience on your resume
- Networking opportunity available

Introduction to Machine Learning

AI Paradigm

- **Artificial Intelligence**: broad field of developing machines that can replicate human behaviors
  - Perceiving, reasoning, learning, problem solving,…
- **Machine Learning**: methodology of teaching machines to learn patterns in data
- **Deep Learning**: branch of ML that uses neural network models
- **Generative AI**: branch of DL that produces new data that is similar to given dataset
  - Text, image, audio, video,…



AI News!

- *Let's share some recent news about AI*

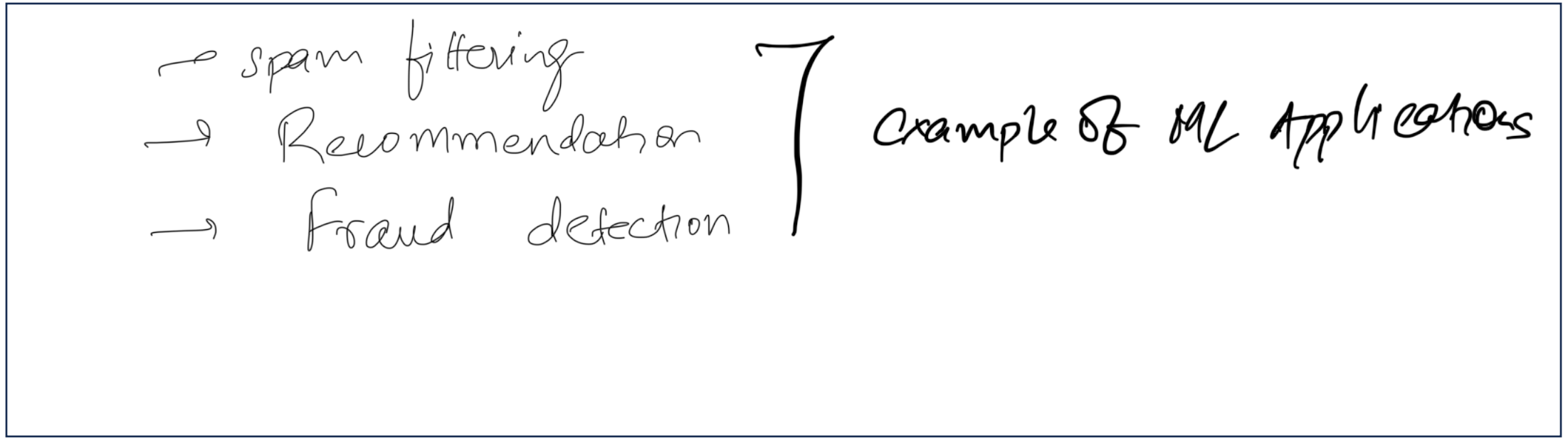
Data → Information → Knowledge.

Machine Learning

- AI systems able to acquire their own “knowledge”
  - By extracting patterns from raw data
- Machine learning is the scientific study of algorithms and statistical models to perform a task using inference instead of instructions
- Also known as **predictive analytics** or **statistical learning**
- *What is non-ML AI?*

Use Cases

- *Name a few machine learning applications that you know*



Three Main Types

- Supervised learning
  - Identify patterns in data that is already **labeled**
- Unsupervised learning
  - Uncover and create the labels itself
- Reinforcement learning
  - Learn to make decisions by receiving rewards or penalties for actions

- Sometimes you may have labeled data in unsupervised learning
- Key difference is whether you use the label in the learning process

Machine Learning Tools

- Python tools
  - Jupyter Notebook
  - JupyterLab
- Python libraries
  - Scikit-learn
  - Pandas
  - Matplotlib
  - Seaborn
  - NumPy and SciPy
- We will use cloud solution

Machine Learning Job Roles

- Data Scientist
  - Applying knowledge of statistics and analytical skills to interpret data
  - Data scientists often have degrees in statistics, computer science, or economics
  - Some programming skills are required
- Machine Learning Engineer
  - Emphasis on programming and system design skills
  - Often have background as a developer or software architect
  - Some knowledge of statistics required
- Applied Science Researcher
  - Applies machine learning technology to a specific domain
  - Requires knowledge of both the domain and machine learning

Programming Basics

What is Programming

- A computer program is a set of instructions
- Computers are only smart because we program them to be
- *If the Dad is the machine, is it a machine learning?*
- *If no, how to convert to ML?*

Programming Language

- Similar to a real language
- Only a written language
  - Which gives us leverage
- Syntax: grammar of a programming language
  - A set of rules to follow
  - If you do not follow the syntax: forget a semicolon, misplace a character, etc.
  - You will receive a syntax error

Variables

- Something that can store information
  - Type, name, and a piece of information
- Type
  - Numerical: integer, float, single/double
  - String: in quotation marks
  - Boolean: True/False
- Naming recommendation
  - Do not capitalize the first word
  - But capitalize the first letter of all words after

Manipulate Variables

- Define/declare

```
int i = 10;
string studentName;
```

- Assign

```
studentName = “Daniel”
num1 = num2 + num3
str1 = str2 + str3
num2 += num3
```

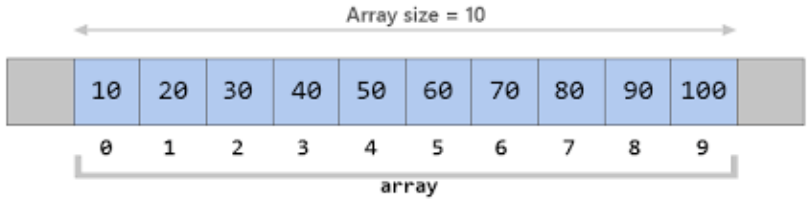
Arithmetic Operators

Name	Symbol	Example	Result
Exponentiation	^	4^2	16
Multiplication	*	16*2	32
Division	/	16/2	8
Addition	+	16+2	18
Subtraction	-	16-2	14

Arrays

- A list of something
  - Use `students` instead of `student1, student2, student3`, etc…

- Indexing: most programming languages start from 0, including Python



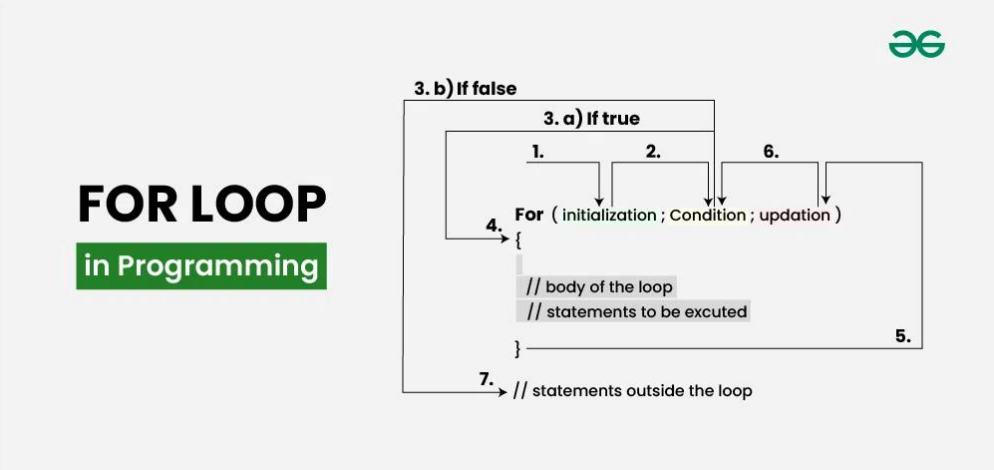
Conditional Statements

- If statement, to function differently under different conditions
- Express the logic: if something is true, do instruction A; otherwise, do instruction B
- Can stack multiple if

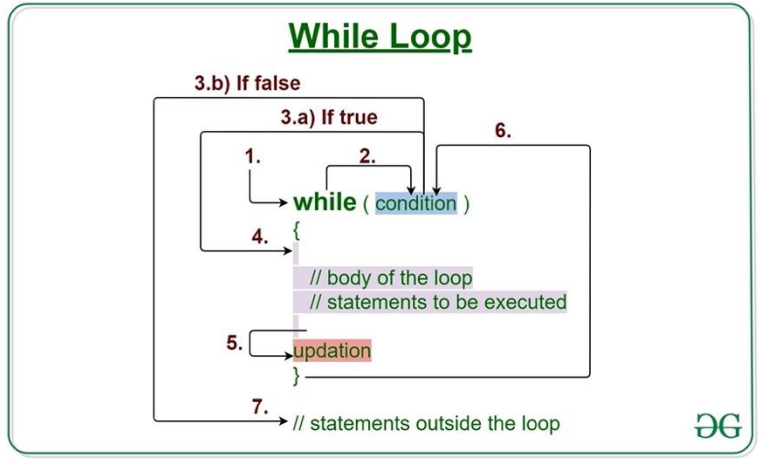
```
if…
elseif…
elseif…
else…
```

Loops

- Run certain instructions **repeatedly**
- For loop



while loop



- Infinite loop caution!
- Set up the condition so that it will be met at some point

Functions

- A segment of code that can be easily run by calling the function name
- Can be called numerous times, in numerous places
- Arguments
  - Variables we pass into a function in order to be manipulated
  - For example, the addition function takes two arguments

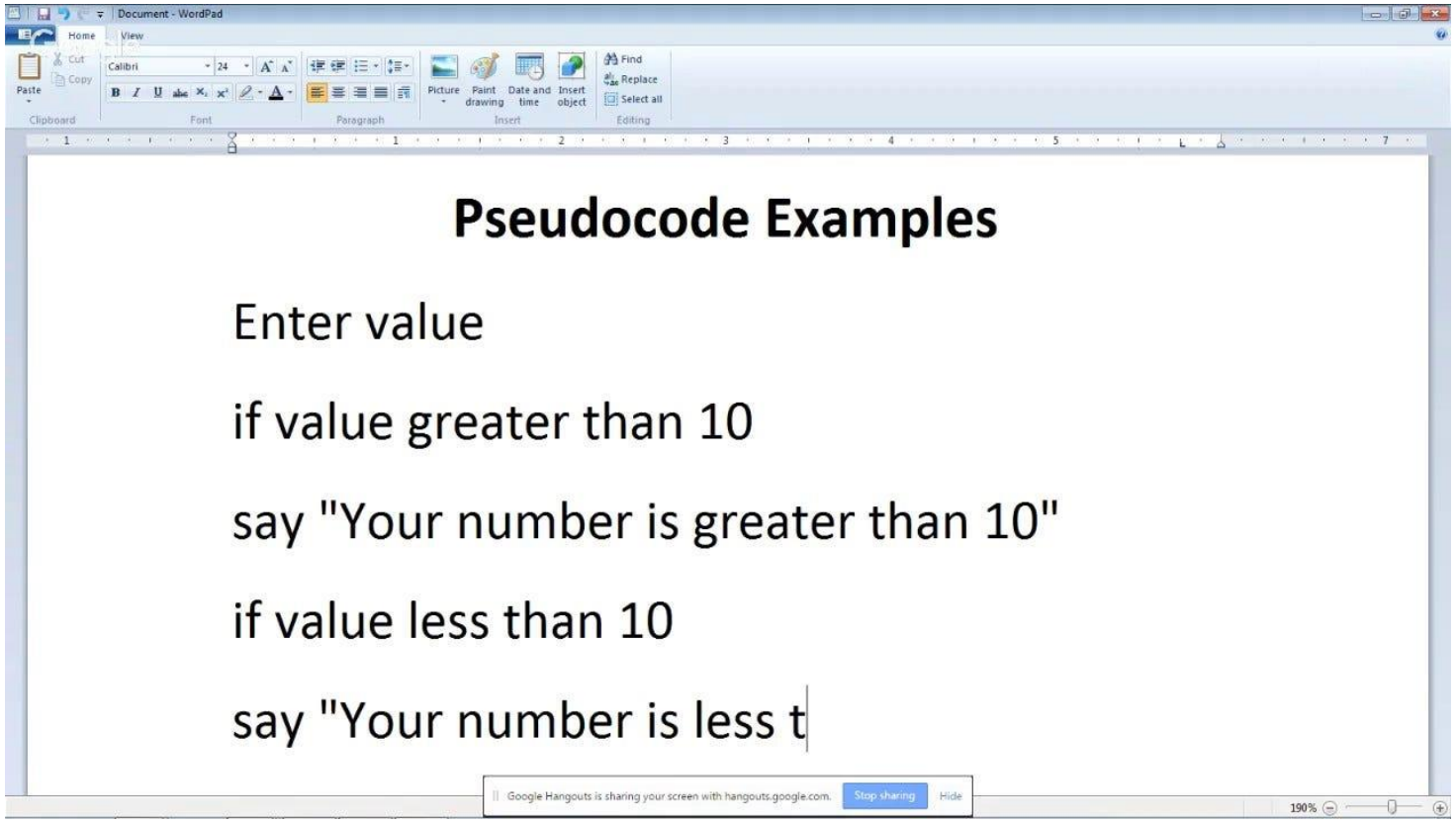
```
addition(int1, int2)
```

- Return values
  - A function can either return variable, or not, to the place where it is called

```
int i = addition(j,k)
```

Pseudocode

- *Pseudo*: not real
- Plan out the code
- Import thing I will ask you to practice
- Become more important after generative AI
- May be tested in final exam



Lab 1: Python Tutorial

- Practice Python basics and NumPy package

- We will use Google CoLab for Python programming
  - Cloud platform will be preferred for teaching purpose
  - You may need gmail account for the AI function
- You may install Anaconda if you want to set up the environment on your local machine
  - You will be responsible for maintaining your own tool

Next Week

- Data Preprocessing
- Exploratory Data Analysis

- Homework 1 due before class

References

- AWS Academy Machine Learning Foundations