Sanjana Jadhav
CS 4375.004

Overview of ML

What is **Machine Learning**?
Machine Learning is a field of analysis incorporating statistics to accurately predict patterns, learn from data, and act when necessary.

Data, Pattern Recognition, and Accuracy
Models are trained using data, which is the core of machine learning. **Data** is usually presented in the form of tables and includes many observations and features about specific topics. **Pattern recognition** is key to predicting results from new data that has not been used in training the model. An ML model's **accuracy** is used to measure its performance. As an example, 80% of the collected data is used to train the model, and 20% is used to check the model's accuracy.

Relationship between AI and ML
AI is the concept of computers trying to replicate "human intelligence." Therefore, AI software can perform a large variety of tasks, just like a human. **ML** is a subfield of AI where the computer learns from data to make decisions. A computer can only predict relationships based on existing data and perform tasks for which it has been trained.

Examples:

1) **Targeted Advertising**
   - ML is used in Google's search engine to target specific advertisements to consumers based on search history. For instance, if one looks up "Ipads," this data is fed into the ML model and advertisements for Ipads begin to show up on other websites visited by the user. Traditional programming cannot be used in this instance as the search history is dynamic.
2) **Medical Imaging**
   - ML is especially beneficial in the healthcare field. To accurately predict illnesses, the software uses patterns derived from datasets including information about various types of diseases. Diseases vary from individual to individual, so traditional programming is not effective in achieving this goal.

Definitions:
When looking at a table of data, it is important to know the terminology. For example, a **row** in a table is called an **observation**. It is a single, specific **instance** or **example**. A **column** in a table is called a **feature**. These are also known as **attributes** or **predictors**, and they provide more information about each observation. There are also other classifications of features: quantitative and qualitative. **Quantitative data** represents finite numeric values such as grades or hours of studying. **Qualitative data** represent categories such as the difficulty of a class. A **target** (or response) is a variable that is being predicted (for instance, grade).

Sanjana Jadhav
CS 4375.004

Figure 1: Sample Data about the Difficulty, Hours of Studying and Grade

qualitative feature, attribute, predictor quantitative (also the target)

Difficulty	Hours of Studying	Grade
easy	1	92
hard	10	97
medium	4	85
hard	2	70 (observation, instance, example)

Data can be read using the wrong data type or it can be incomplete. It is important to know how to read the data accurately. It helps us understand the data and use appropriate algorithms to find relationships.

Personal Interest:
My interest in ML started through my exposure to sci-fi movies. In recent years, my interest has only been magnified due to my passion for math/statistical analysis. This past summer, I worked in a team to create an ASL translator that uses a CNN model to accurately predict letters and numbers from hand gestures. However, due to time constraints and my lack of knowledge in the field of ML, I worked on the interface of the application and the communication between the front end and the ML model. As I worked on this software, I realized the impact of increasing accessibility for those who communicate in ASL.

My goal in the future is to create meaningful technology that will solve real-world problems and expand inclusivity. The first step is gaining a solid foundation in the field of ML :)

Sources:
- Class Notes/Textbook
- AI vs. ML
- Targeted Ads
- Medical Imaging