IST 707 Applied Machine Learning

HW1: What is Machine Learning

Task 1: What, Why, and Where?
Combine what you read from the following articles and the week 1 lecture to answer the following questions. **Write a short paragraph for each question.**

PwC 2021 AI Predictions
McKinsey: Notes from the AI Frontier

1. What is machine learning?

Artificial intelligence is an area of computer science where the goal is to enable computers and machines to perform human-like tasks and simulate human behavior.
Machine learning is a data analysis technique that automates the creation of analytical models. It is an artificial intelligence subfield founded on the notion that machines are capable of learning from data, spotting patterns, and making decisions with little assistance from humans.

2. Where can you use machine learning?

Machine learning has a wide usage, including:
- Image and speech recognition
- Natural language processing
- Recommender systems
- Predictive maintenance
- Fraud detection
- Healthcare
- Robotics
- Self-driving cars
- Marketing and customer service.

It is used in various industries such as finance, healthcare, retail, transportation and manufacturing to improve efficiency, reduce costs and make data-driven decisions.

3. Why should we use machine learning?

There are several reasons why we should use machine learning:
- Processes for making decisions can be automated and improved, which boosts effectiveness and lowers costs.
- It can examine massive volumes of data and spot patterns that humans might miss.
- In a variety of applications, including fraud detection, medical diagnosis, and recommendation systems, it may make predictions and suggestions based on data.
- Additionally, it can be utilized for jobs that are challenging or impossible for humans to complete, like speech and picture recognition, natural language processing, and others.

- It can be used to automate monotonous operations, freeing up human workers to concentrate on harder and more imaginative jobs.
- Smart cities and self-driving cars are only two examples of complex systems that may be managed and observed using it.
- It can be utilized to develop brand-new, cutting-edge goods and services that weren't before possible.

4. What benefits does machine learning have over other traditional methods of analysis, prediction, and decision-making?

Machine learning has a number of advantages over conventional methods of analysis, prediction and decision-making:

- Scale: Machine learning algorithms can assess and generate predictions on a scale that would be hard for people to manually perform. They can handle big and complex datasets.
- Automation: Machine learning algorithms can automate repetitive operations like feature extraction and data cleaning, which can reduce errors and save time.
- Self-learning: Unlike traditional approaches that need human input to update and improve, machine learning algorithms may enhance their performance over time as they are exposed to additional data.
- Flexibility: Machine learning algorithms can be tailored for certain tasks and datasets and used to a wide range of problem domains.
- Accuracy: When working with huge and complicated datasets, machine learning algorithms frequently outperform conventional methods in terms of accuracy.
- Missing data handling: Unlike traditional approaches, which frequently require complete data to function, machine learning algorithms can frequently tolerate missing data in the training set.

5. What business problems are machine-learning and artificial intelligence most suited to?

Machine learning and artificial intelligence are well-suited to a wide range of business problems including:

- Predictive modeling: Machine learning algorithms can be used to forecast demand for goods and services, find patterns in sales data, and predict future customer behavior.
- Fraud detection: By examining trends in financial data, machine learning algorithms can be used to find fraudulent transactions and other questionable activity.
- Customer segmentation: Based on their behavior and other characteristics, customers can be divided into several groups using machine learning algorithms, which can then be utilized to create focused marketing campaigns.
- Systems that provide individualized recommendations for clients based on their browsing and purchase histories are known as recommender systems.

- Image and speech recognition: Machine learning algorithms may be used to identify images and sounds, which is useful for automated transcription, voice-activated assistants, and facial identification, among other uses.
- In order to perform tasks like sentiment analysis, language translation, and language production, natural language processing (NLP) uses machine learning algorithms to analyze, comprehend, and synthesize human language.
- Fraud detection, network security, and manufacturing quality control can all benefit from the ability of machine learning algorithms to spot unexpected or aberrant patterns in data.
- Time series forecasting: Machine learning algorithms can be used for time series forecasting, which has a variety of uses, including the ability to anticipate the stock market and the weather.

Task 2: Practice Your Critical Thinking and Writing

Read the following two news articles. One criticizing Google Flu Trends, and the other defending it. **Write one paragraph to summarize the criticism, and another paragraph for the defense. Write the third paragraph to offer your own thoughts**, e.g., is the criticism valid? Does the defense make sense? What other problems or benefits do you see in Google Flu Trend or similar big data applications?

[NYT: Google Flu Trends: The Limits of Big Data](#)

Google Flu Trends was a project that aimed to use search data to predict flu outbreaks in real-time, however, it faced limitations and struggled with accuracy, and ultimately was discontinued by Google. The project highlighted the importance of considering the limitations and complexities of using big data for predictions, and the need for a better understanding of the underlying relationships between data and real-world phenomena.
Google Flu Trends did not account for other factors that might influence flu-related search activity, such as media coverage or changes in the way people use the internet.

[Atlantic: In Defense of Google Flu Trends](#)

According to the article, Google Flu Trends was able to accurately predict flu outbreaks in some regions, and its predictions were often more timely than those of traditional surveillance methods, such as lab reports or surveys.
The article also notes that Google Flu Trends faced limitations, such as the difficulty of accounting for changes in search behavior, but argues that these limitations are not unique to Google Flu Trends and are inherent to any big data-based surveillance system.
The article also emphasizes that the Google Flu Trends project should be viewed in the context of a broader effort to develop new and innovative methods for public health surveillance. It argues that while Google Flu Trends alone was not able to provide a complete picture of the flu situation, it was a valuable tool that could be used in combination with other methods to improve our understanding of flu outbreaks.

In my perspective, Google Flu Trends was a pioneering work in the sector and it opened the door for new projects in the future, even though there were not proper results and there were some restrictions and issues. Utilizing search data was a cutting-edge method of monitoring influenza, and it created new opportunities for the use of big data in public health surveillance.

Public health officials and researchers were able to better analyze and address flu outbreaks because to the project's useful information.

The cost-effectiveness of using Google search data to monitor influenza activity was due to the lack of extra data gathering or analysis needed.

There might have been some ways on how to improve the project including:

- Increasing the number of data sources used, such as social media and electronic health records, to improve the model's accuracy.
- Taking into account potential environmental elements that could affect flu activity by including real-time data, such as weather and air quality.
- Enhancing the model's capacity to account for variations in search activity over time, as people may refer to the flu with various phrases.
- Incorporating regional public health data and working with regional public health bodies to improve the amount of regional data available.
- Using historical data to periodically test and validate the model's performance to make sure it is functioning properly.

Submission requirements:

Word document or PDF, using 12-point Arial or Times New Roman and at least 1-inch margins on all sides.

Grading criteria:

All prompts are thoroughly answered and demonstrate a critical understanding of the material. Writing should be precise and concise. Points would be deducted for both incorrect and irrelevant content.