Machine Learning Concepts – Lec 1

# 1. Machine Learning

Formal Definition: Machine Learning (ML) is a field of computer science that uses statistical techniques to give computer systems the ability to learn from data, without being explicitly programmed.  
Layman Explanation: In simple terms, machine learning is about teaching computers to learn from examples and data, so they can make decisions without human intervention.  
Example: If you want to teach a computer to identify pictures of cats, you show it thousands of pictures of cats and let it 'learn' what a cat looks like.  
Real World Example: Netflix's recommendation engine uses machine learning to suggest shows and movies based on your viewing history.

# 2. Types of Machine Learning

There are three primary types of machine learning:

1. Supervised Learning: In supervised learning, the model is trained on labeled data. This means the model is given input-output pairs, where the output is already known. The goal is for the model to learn a mapping from inputs to the correct outputs.  
2. Unsupervised Learning: In unsupervised learning, the model is trained on data without labels. The goal is for the model to find patterns or structure in the data.  
3. Reinforcement Learning: In reinforcement learning, the model learns through trial and error by interacting with an environment. It receives feedback in the form of rewards or penalties based on the actions it takes.

# 3. Supervised Learning

Formal Definition: Supervised learning is a type of machine learning where the model is trained using labeled data, meaning the correct output is provided for each training example.  
Layman Explanation: It’s like a teacher giving students the correct answers during a test, so they can learn to find patterns and apply them to new questions.  
Example: A spam filter uses supervised learning by learning from emails marked as ‘spam’ or ‘not spam’ to classify new emails correctly.  
Real World Example: Credit card fraud detection systems are trained on data labeled as ‘fraudulent’ or ‘non-fraudulent’ to detect potential frauds.

Problem Example: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months. Should this be treated as a classification or regression problem? Answer : Regression

# 4. Unsupervised Learning

Formal Definition: Unsupervised learning is a type of machine learning where the model is trained on data without any labels. The goal is for the model to find structure or patterns in the data.  
Layman Explanation: It’s like exploring an unknown place without a guide. The machine learns to group and find similarities in the data without knowing what it is beforehand.  
Example: Clustering algorithms can group similar customers together based on purchase patterns.  
Real World Example: Market segmentation is done using unsupervised learning by clustering customers based on their buying behaviors.

Problem Example: Given a set of news articles found on the web, group them into a set of articles about the same story. This is an example of an unsupervised learning problem.

Problem Example 2 : Email detection, spam or not spam. This is an example of unsupervised learning.

# 5. Introduction to Machine Learning Algorithms

Machine learning algorithms can be broadly categorized into three types:  
1. Regression Algorithms: These algorithms predict continuous output values.  
2. Classification Algorithms: These algorithms predict categorical outcomes.  
3. Clustering Algorithms: These algorithms find structure in data by grouping similar data points together.  
Refer to the image for common examples of algorithms used in each type.

# 6. Libraries and Frameworks for Machine Learning

There are several libraries and frameworks that make it easier to develop and implement machine learning models. Some of the most popular ones are:  
1. Scikit-learn: A simple and efficient tool for data mining and data analysis, built on top of NumPy, SciPy, and Matplotlib.  
2. TensorFlow: An open-source platform for machine learning developed by Google. It is used for both research and production.  
3. PyTorch: A deep learning framework developed by Facebook's AI Research lab. It is known for its flexibility and ease of use.  
4. Keras: A high-level neural networks API that runs on top of TensorFlow. It is user-friendly and modular.  
5. XGBoost: An optimized gradient boosting library designed for speed and performance. It is widely used for structured data problems.