Lecture 5 - Machine Learning

Machine Learning
Just as we touched on before, machine learning is a charming little offshoot of artificial intelligence. It's all about exploring and learning from a treasure trove of data, identifying patterns, and making predictions - kind of like a data detective, if you will!
What we're really aiming for here is to dive into the world of algorithms that can learn from the data they interact with, and even come up with new algorithms of their own. It's like teaching a machine to think and grow, which is pretty cool, right?

Train and Test
Think of machine learning as a two-step dance. First, we have the training phase where we use our dataset to teach the prediction model to groove. Then, we move on to the testing phase where we introduce our trained model to new, unseen data and see how well it performs. So, it's learn, test, and repeat!

Train data and Test data
So, you might have picked up on the fact that we always split our data into two parts: the training data and the test data. It's like a secret pact between the two - they are strictly separate and never mix. That means our test data is like a secret set of questions that we don't use while we are training.

Think of it like studying for a math exam. You wouldn't want to know the exact questions that will appear on the test ahead of time, right? That's because it wouldn't really be a fair evaluation. If you study using the actual test questions, you'd just be solving those questions and might miss out on learning other important concepts.

It's the same with machine learning. If we train our model on the test data, it becomes a one-trick pony, only good at predicting what it's already seen. It won't perform well when faced with truly new data in the future. And that's not what we want, right? We're aiming for a model that's a jack of all trades and can tackle all sorts of unseen data with ease.

Data
Let's take a moment to chat about data, the lifeblood of machine learning. Data can wear many different hats. It could be a straightforward table or matrix filled with numbers, or it could get a bit fancy, appearing as a photo, video, or sound clip.

To make it easy to visualize, let's imagine our data as a big table of numerical values. In this table, we call each column a **'feature'** and each row a **'sample'**. Different folks might use different names, but let's stick with these terms for simplicity.

For instance, if we're looking at data from patients with diabetes, each column (feature) could represent different characteristics of a patient like age, weight, blood pressure, blood sugar

levels, and so forth. The rows (sample) then would represent each patient - Patient 1, Patient 2, Patient 3, and so on. The total number of rows would give us the total number of patients in our data. It's a neat, organized way to understand and work with all that information!

Supervised learning, Unsupervised learning, and Reinforcement learning

There are three major types of machine learning methods that we can categorize: supervised learning, unsupervised learning, and reinforcement learning.

Imagine supervised learning like a friendly guide. You're given data with the correct answers attached. Like if we had a bunch of photos, pictures of dogs would be labeled as dogs, and pictures of cats as cats. So, as it learns, it knows what each picture is, making it a very well-informed guide.

On the other hand, unsupervised learning is like a daring explorer. There are no correct answers attached to the data. Imagine being handed a big pile of animal photos. Unsupervised learning is all about finding patterns in such data. It's like that explorer finding a new path in the wild. An example would be clustering, which groups similar animals together.

Lastly, reinforcement learning is a bit like an adventurer in a game world, learning as it interacts with its environment, although sometimes it's given a bit of data. It's AI learning through real-world interactions within a specified environment.

Supervised learning

We'll delve into the intriguing details of unsupervised learning and reinforcement learning later. But for now, let's kick things off with our friendly guide, supervised learning.

We're about to dive into several supervised learning algorithms, which is pretty exciting! But before we do that, there's something you should know. Supervised learning splits into two main types: Classification and Regression.

Classification and Regression

Think of Classification like putting things in different baskets based on their type. The answer, or the label, is a discrete value. For instance, predicting whether a given picture is of a dog or a cat is a classification task - it falls into one basket or the other.

On the other hand, Regression is all about guessing a number that can be anywhere on a sliding scale. Here, the correct label is a continuous, or real, value. For example, predicting the price of a house based on various information about it is a regression task. It's like guessing the weight of a fruit - it could be any number within a certain range.