**Introduction to Machine Learning**
Prof. Sudeshna Sarkar
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Module - 1
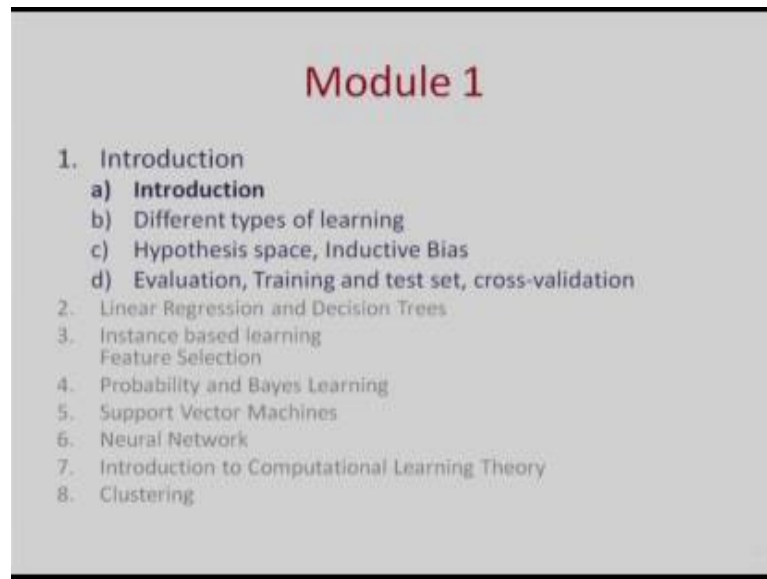**Lecture - 01**
Introduction

Good morning, I am Sudeshna Sarkar. Today, we start the first lecture on machine learning. This is module one, part A. Today, we will introduce machine learning, basic, go through the basics of the course, discuss the brief history of machine learning and discuss what learning is about and some simple applications of machine learning.

(Refer Slide Time: 00:46)
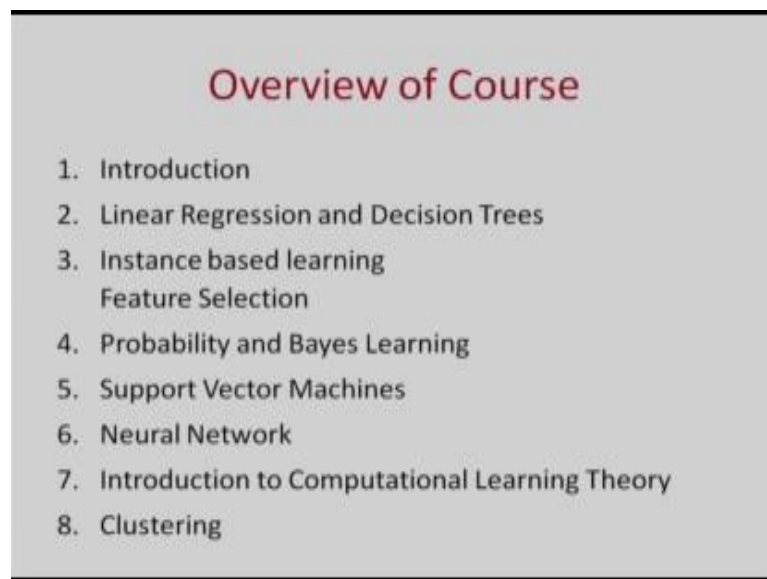
First, this is the overview of the course. The course is over 8 weeks and will have 8 modules. The 1st module is Introduction; in the 2nd module we will discuss about Linear Regression and Decision trees; 3rd module, Instance Based Learning and Feature selection; 4th module, Probability and Bayes learning; 5th module, Support Vector Machines; 6th module, Neural Networks; 7th module, we will do an Introduction to Computational learning theory and possibly a little bit on sample learning and then, the last module we will talk about Clustering.

(Refer Slide Time: 01:23)

In the 1st module, that is introduction, we will have four parts. Today, we will give a brief introduction, in the next more lectures we will have discuss about different types of learning supervise and supervised, etcetera. Then, we will have the 3rd module where we will talk about hypothesis space and inductive bias. Following this, we will talk about evaluation training and test set and cross-validation.

(Refer Slide Time: 01:51)
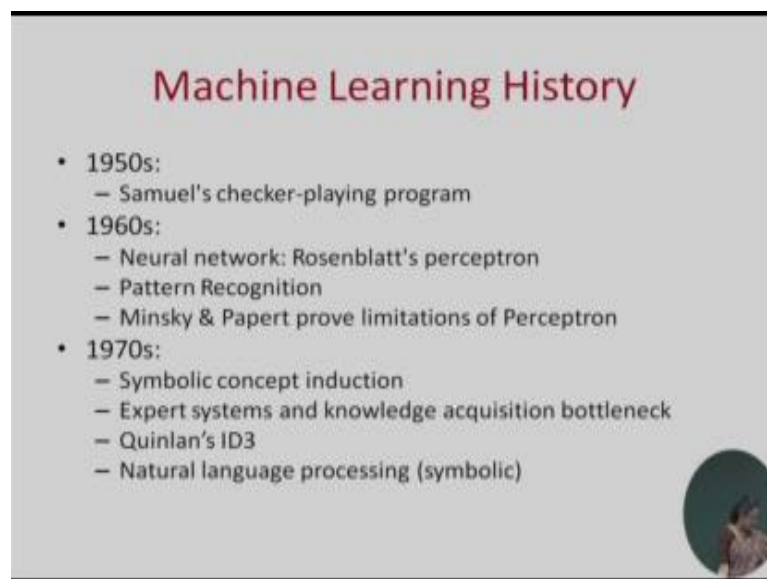
First, I will like to start with a brief history of machine learning. A machine that is intellectually capable as much as humans has always fired the imagination of writers and also the early computer scientist who were excited about artificial intelligence and machine learning, but the first machine learning system was developed in the 1950s.

In 1952, Arthur Samuel was at IBM. He developed a program for playing Checkers. The program was able to observe positions at the game and learn a model that gives better moves for the machine player. The system played many games with the program and observed that the program was able to play better in the course of time with getting more experience of board games. Samuel coined the term machine learning and he defined learning as a field of study that gives computers the ability without being explicitly programmed.

In 1957, Rosenblatt proposed the perceptron. Perceptron is the simple neural network unit; it was a very exciting discovery at that time. Rosenblatt made the following statement; the perceptron is designed to illustrate some of the fundamental properties of intelligent systems in general without becoming too deeply immersed in the special and frequently unknown conditions, which hold force particular biological organisms.
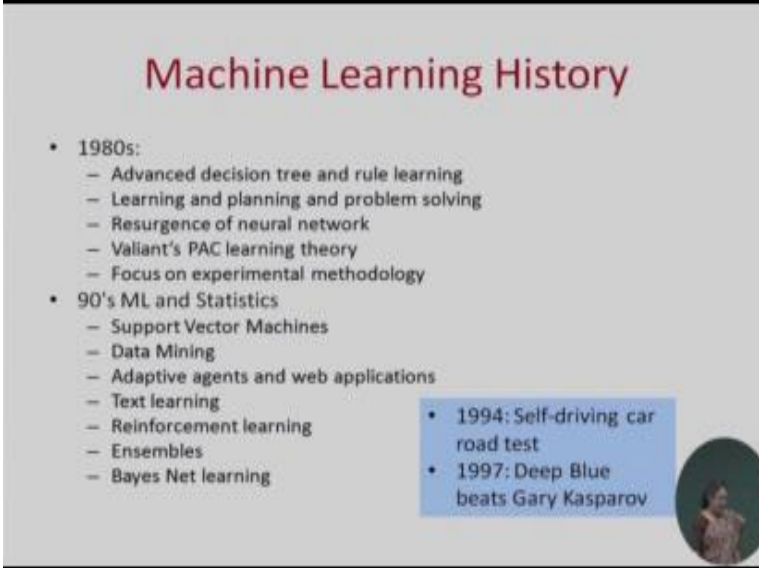
But after 3 years, (Refer Time: 03:52) came up with the delta learning rule that is used for learning perceptron. It was used as a procedure for training perceptron. It is also known as the least square problem. The combination of these ideas created a good linear classifier. However, the work along these lines suffered a setback when Minsky in 1969 came up with the limitations of perceptron. He showed, that the (Refer Time: 04:27) problem could not be represented by perceptron and such inseparable data distributions cannot be handled and following this Minsky's work neural network research went dormant up until the 1980s.

In the meantime, in the 1970s, machine learning symbolic following the symbolic type of artificial intelligence, good old fashioned artificial intelligence, those types of learning algorithms were developed, concept induction was worked on. And then, J.R. Quinlan, in 1986 came up with decision tree learning, specifically the ID3 algorithm. It was also released as software and it had simplistic rules contrary to the black box of neural

networks and it became quite popular. After ID3 many alternatives or improvement ID3 were developed such as cart, regression, trees and it is still one of the very popular topics in machine learning.

During this time symbolic natural language processing also became very popular.

(Refer Slide Time: 05:46)

Machine Learning History

- 1980s:
  - Advanced decision tree and rule learning
  - Learning and planning and problem solving
  - Resurgence of neural network
  - Valiant's PAC learning theory
  - Focus on experimental methodology
- 90's ML and Statistics
  - Support Vector Machines
  - Data Mining
  - Adaptive agents and web applications
  - Text learning
  - Reinforcement learning
  - Ensembles
  - Bayes Net learning

- 1994: Self-driving car road test
- 1997: Deep Blue beats Gary Kasparov

In 1980s, advanced decision trees and rule learning were developed. Learning, planning, problem solving was there. At the same time, there was a resurgence of neural network. The intrusion of multilayer perceptron was suggested by (Refer Time: 06:08) in 1981 and neural network specific back propagation algorithm was developed. Back propagation is the key ingredient of today's neural network architectures. With those ideas neural network research became popular again and there was acceleration in 1985, 86 when neural network researchers presented the idea of MLP, that is, multilayer perceptron with practical BP training. (Refer Time: 06:39), Williams, (Refer Time: 06:41), Nielsen were some of the scientists who worked in this area.
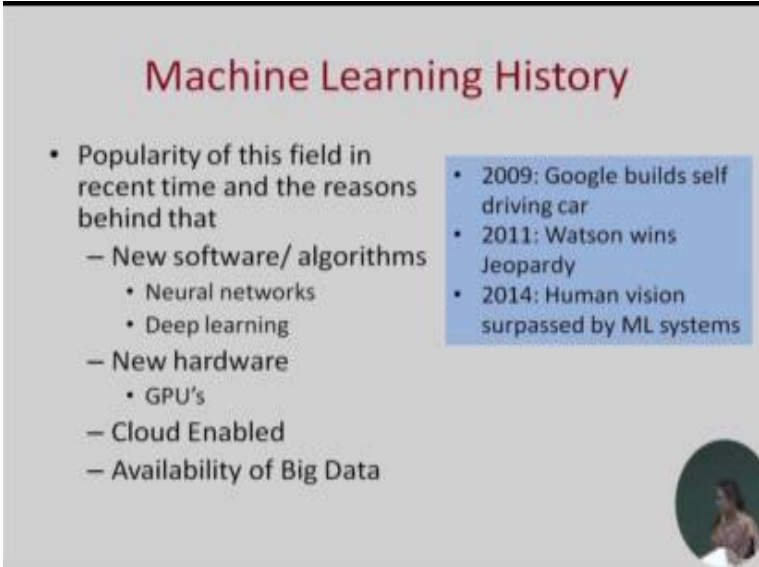
During this time, theoretical framework of machine learning was also presented. Valiant's PAC learning theory, which stands for probably approximately correct learning, it was developed and the focus shifted on experimental methodologies. In the 90s,

machine learning embraced statistics to a large extent. It was during this time, that support vector machines were proposed. It was a machine learning breakthrough and the support vector machines was proposed by Vapnik and Cortes in 1995 and S.V. Hem had very strong theoretical standing and empirical results.

Then, another strong machine learning model was proposed by Freund and Schapire in 1997, which was part of what we called ensembles or boosting and they came up with an algorithm called Adaboost by which they could create a strong classifier from an ensemble of weak classifiers.

The kernalized version of SVM was proposed near 2000s, which was able to exploit the knowledge of convex optimization, generalization and kernels. Another ensemble model was explored by Bremen in 2001 that ensembles multiple decision trees where each of them is curated by a random subset of instances. This is called random forest. During this time, Bayes net learning was also proposed. Then, neural network took another damage by the work of (Refer Time: 08:31) showed that gradient loss after the saturation of neural network unit happens when we apply back propagation so that after a certain number of epochs neural networks are inclined to over fit. But as we come closer today we see, that neural networks are again very much popular.

(Refer Slide Time: 08:51)

We have a new era in neural network called deep learning and this phrase refers to neural network with many deep layers. This rise of neural network began roughly in 2005 with the conjunction of many different discoveries for people by Hinton, LeCun, Bengio, Andrew (Refer Time: 09:20) and other researchers.

At the same time, if you look at certain applications where machine learning has come to the public forefront. In 1994, the first self driving car made a road test; in 1997, Deep Blue beat the world champion Gary Kasparov in the game of chess; in 2009 we have Google building self driving cars; in 2011, Watson, again from IBM, won the popular game of Jeopardy; 2014, we see human vision surpassed by ML systems. In 2014-15, 2015 we find, that machine translation systems driven by neural networks are very good and they are better than the other statistical machine translation systems where certain concepts and certain technology, which are making headlines.

Now, in machine learning we have GPU's, which are enabling the use of machine learning and deep neural networks. There is the cloud, there is availability of big data and the field of machine learning is very exciting now. So, with this brief introduction of machine learning history we will now discuss what is learning? What is machine learning? What is a learning algorithm?

(Refer Slide Time: 11:06)
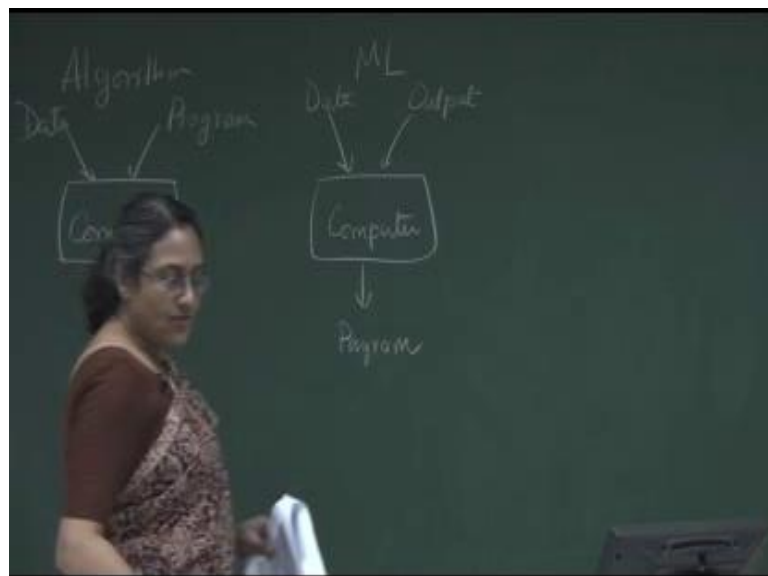
First, let us look at how a machine learning solution differs from a programmatic solution. When you have a program or algorithm to solve a problem, this is how you use the computer. This is your computer and you write a program. The program takes data as the input and the program produces output. On the other hand, when we are using machine learning and you have the computer, you are feeding the data, input as well as examples of output. So, you are putting examples of input, output data and you are getting a program or a model with which you can solve subsequent tasks. So, this is what learning is about.

(Refer Slide Time: 12:23)

So, with this let us come to the formal definition of machine learning. So, learning is the ability to improve one's behavior with experience. So, it is about building computer systems, that automatically improve with experience and we have to discuss what are the fundamental laws that governed the learning processes. Machine learning explores algorithms that learn from data, build models from data and this model can be used for different tasks. For example, model can be used for prediction, decision making or solving tasks.

Now, we will discuss formal definition of machine learning as given by Tom Mitchell and this is our definition that is followed very popularly.

So, Mitchell's definition of machine learning; it says, that computer program is said to learn from, what does it learn from? It learns from experience. So, the computer program learns from experience E. So, E is the experience of the, that the computer program uses. It is passed data with respect to some class of tasks T and performance measure P; if its performance on tasks in T, on tasks in T as measured by P improves with experience E.

So, learning is about using experience data that is from passed problems, solving data and there is a task and the task belongs to a class of tasks T and the tasks are evaluated by performance measure P, and a machine is said to learn tasks in T if, if performance at tasks as measured by P improves with experience E. So, what we see is that the components of a learning algorithm are follows; task is the behavior of the task, behavior of the task that the learning program is seeking to improve. For example, there are different types of task like prediction, classification, acting in an environment, etcetera.

The second component is the data or the experience. So, the experience is also called the data. This is what is used for improving the experience at the task. And then, there is a measure of improvement P. For example, you might want to increase accuracy in prediction or you might want to have new skills to the agent which it did not earlier

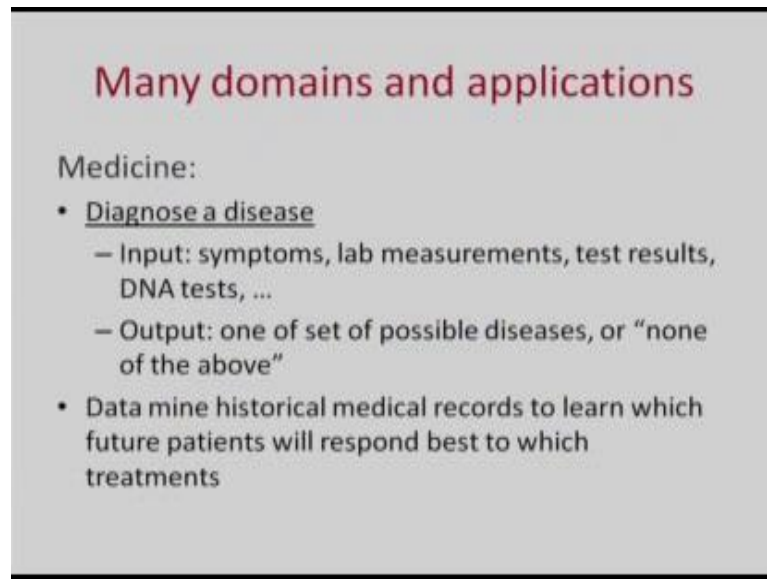process or improve efficiency of problem solving, corresponding to this you can define the performance measure.
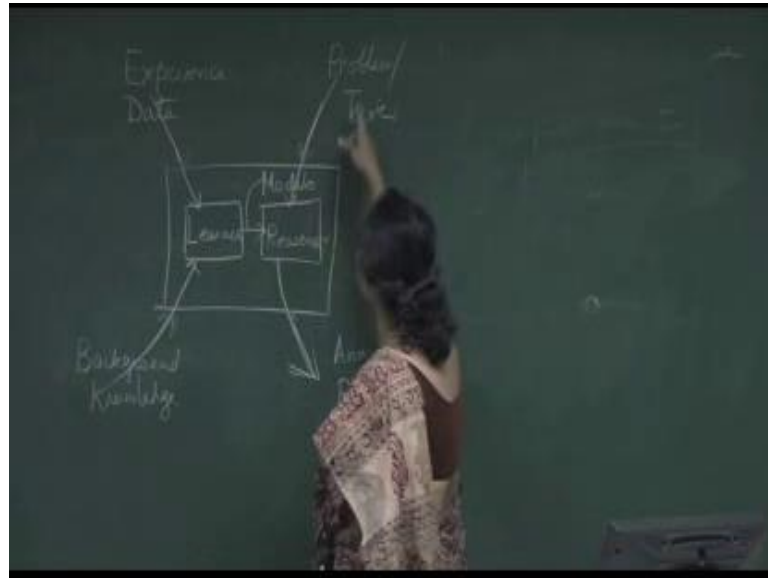
So, based on this definition we can look at learning system as a box. So, this is our learning system. It is a box to which we feed the experience or the data and there is a problem or a task, that requires solution and you can also give background knowledge, which will help the system. And for this problem or this task the learning program comes up with an answer or a solution and its corresponding performance can be measured.

So, this is the schematic diagram of a machine learning system or a learner system. Inside there are two components, two main components, the learner L and the reasoner. See, the learner takes the experience and from that it can also take the background knowledge and from this the learner builds models and this models can be used by the reasoner, which given a task finds the solution to the task. So, the learner takes experience and background knowledge and learns a model and the reasoner works with the model and given a new problem or task, it can come up with the solution to the task and the performance measure corresponding to this.
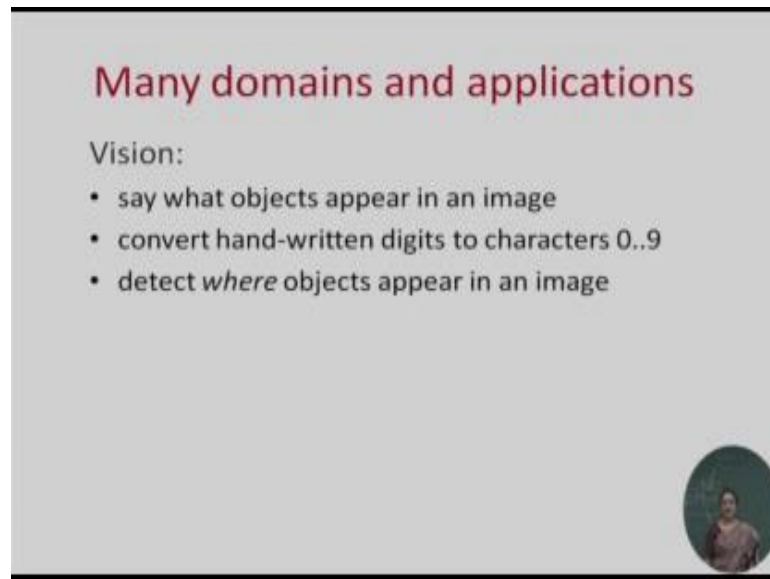
Now, we will like to look at some examples of machine learning system. So, machine learning, there are many domains and applications of machine learning. For example, in medicine you can use machine learning to diagnose a disease where the input at the symptoms, lab measurements, test results, DNA tests, etcetera. Output could be one of a set of possible diseases or none of them. For doing this, one can determine historical medical records and learn which future patients will respond best to which treatments.
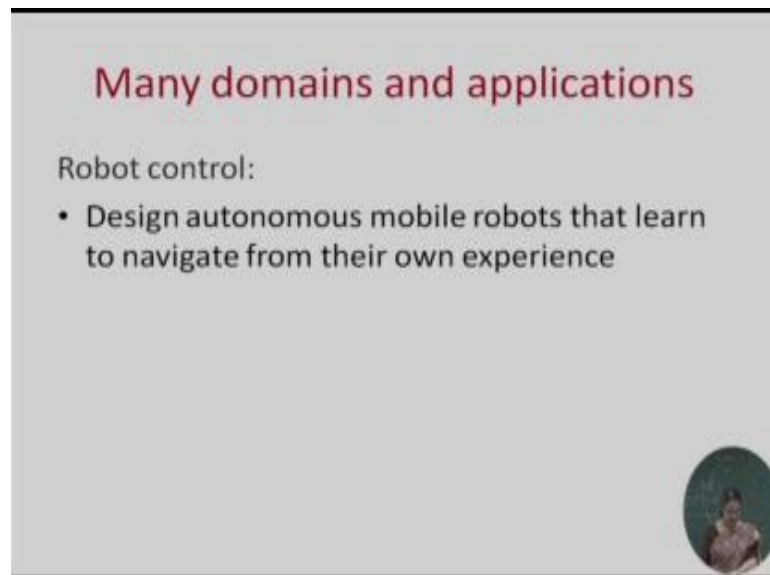
(Refer Slide Time: 21:31)

Another domain is computer vision where given an image; you want to find out what objects appear in an image and where the objects appear in an image.
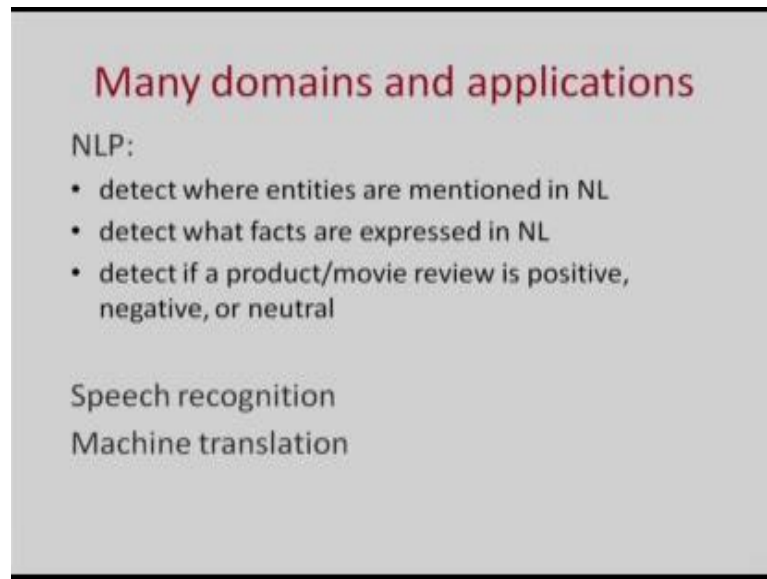
(Refer Slide Time: 21:47)

A third domain, robot control; one can use machine learning to design autonomous mobile robots that learn to navigate from their own experience.

Then, in the domain of natural language processing, one can detect where entities are mentioned in natural language and detect what facts are expressed in natural language. One can look at a product or movie review and find out if it is positive, negative or neutral, that is, the sentiment on the review.

Other applications in NLP include speech recognition, machine translation etcetera. In the financial domain, one can try to predict if a stock will rise or fall; one can predict if a user will click on an advertisement or not.

There are many applications in business intelligence. You want to robustly forecast product sales quantities taking seasonality and trend into account. Identify cross selling promotional opportunities, identify price sensitivity of a consumer product, and optimize product location at a super market shelf and so on.

Then, there are other applications such as fraud detection, credit card fraud detection. Understand consumer sentiment, forecast women's conviction rates based on external macroeconomic factors, etcetera. So, these are some of the many, many applications of machine learning. Machine learning is a part of many products and systems that we routinely use.

(Refer Slide Time: 23:45)
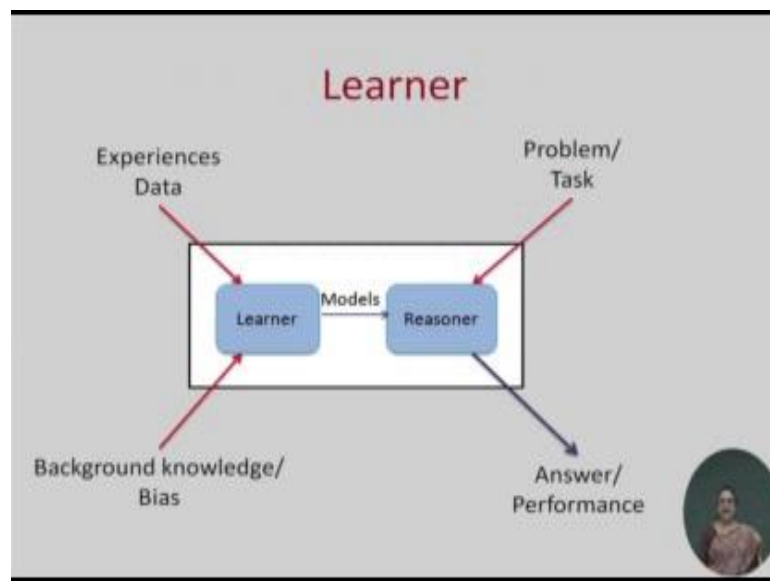
If you look at the box that we drew for machine learning systems and discuss how we can go about creating a learner, these are the following steps.
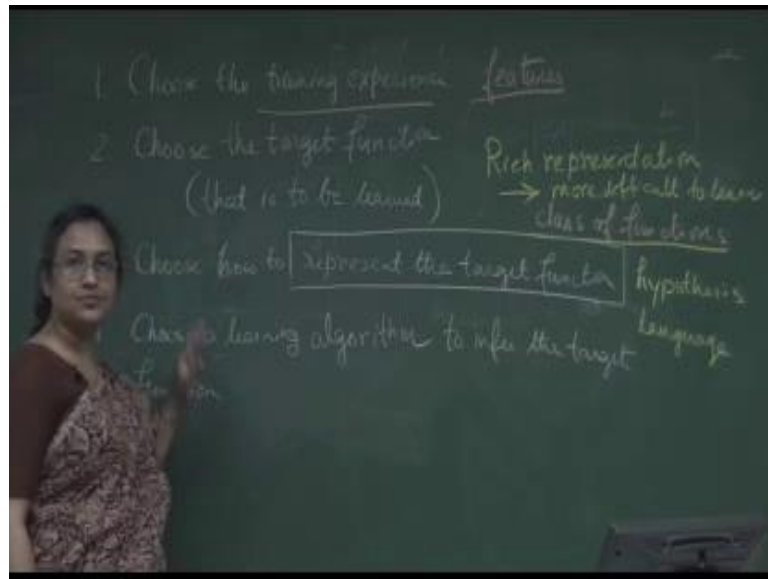
First of all, we choose the training experience or the training data. Then, we choose the target function or how we want to represent the model. So, this is what we want to learn, the target function that is to be learned. For example, if you are trying to write a machine learning system to play the game of checkers, the target function would be, given a board position, what move to take. And then, we want to have the class of function that we will use, the task is given a board position what move to take and we will design the function as a function of the input and we have to decide what type of function we will use, whether we will use a linear function or some other representation.

So, we choose how to represent the target function. And finally, we choose a learning algorithm to infer the target function. So, the learning algorithm will explore the possible function parameters so that based on the training experience it can come up with the best function given its computational limitations. So, what is very important in the designing of a learning algorithm is how to represent the target function. Before that what is important is how to represent the training experience.

So, the training experience, as we will see, can be expressed in the terms of features of the domain and then we have to decide how to represent the target function. So, we want to come up with the appropriate class of functions on the features. So, you have to decide

the class of the functions and when we are trying to find this class of the functions, we have to make very important decision. We can go for a very powerful function class, which is very complex, can represent complex concepts. But if you choose a powerful or rich representation, if you choose a rich representation of the class of functions, then we can represent complex function, but and it will be more useful for subsequent problem solving, but it may be more difficult to learn.

So, richer representations are able to represent many types of classes including complex classes, solve complex problems, but are more difficult to learn. And the components of representation, as we have said, are the features or the attributes of the domain or the vocabulary on which we define the functions. So, we have the features and then we have the class of functions, which we also call the hypothesis language.

We will talk more about this in the next lecture and when we in the course of this course as we study different learning algorithms, you can keep these steps in mind. With this we come to the end of introduction.

Thank you very much.