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

Module - 1
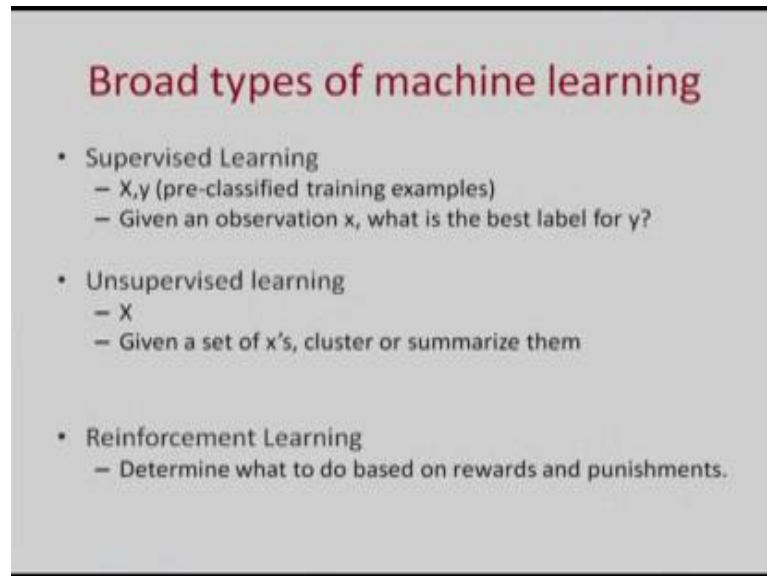**Lecture - 02**
Different Types of Learning

Good Morning. We start with second part of the first module that is introduction module. We will introduce the Different Types of Machine Learning.

(Refer Slide Time: 00:32)
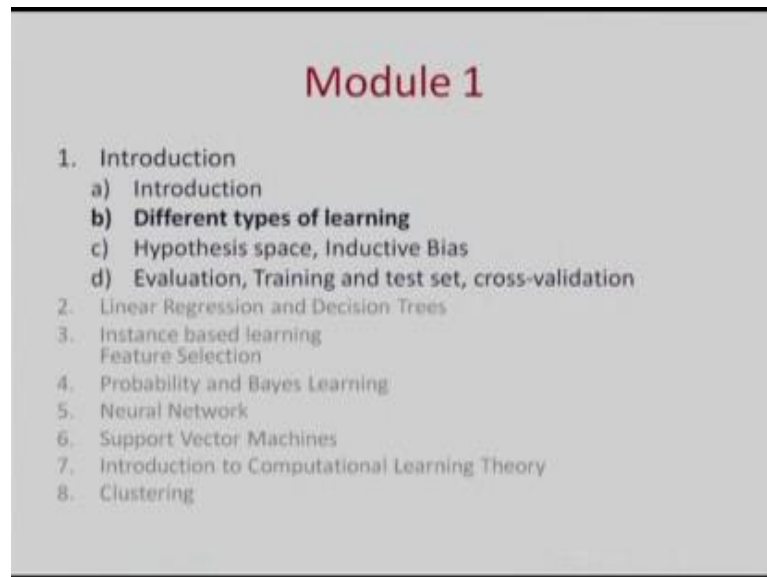
So, there are few broad types of machine learning.

(Refer Slide Time: 00:41)

There is Supervised learning.

(Refer Slide Time: 00:48)

In supervised learning what we have is. We said that in learning we use data and data we can say can comprise of input and the corresponding output. For every data instance we can have the input x and the corresponding output y. And, from this the machine learning system will build a model so that given a new observation x will try to find out what is the corresponding y. This called supervised learning because for every instance we tell what is the output. So, this is called labeled data.

Then we have Unsupervised learning. In unsupervised learning you are only given x, there is no label to the data. And given the different data points you may like to for example plaster them or summarize them or find some patterns in them. The third type of machine learning is Reinforcement learning. In reinforcement learning you have an agent who is acting in a environment, and you want to figure out what actions the agent must take at every step. The action that the agent takes is based on the rewards or penalty is that the agent gets in different states.

Apart from these three, we have also Semi-supervised learning. In semi-supervised learning it is a combination of supervised and unsupervised learning. That is you have some labeled training data and you also have a larger amount of unlabeled training data, and you can try to come up with some learning out of them that convert even when the training data is limited.
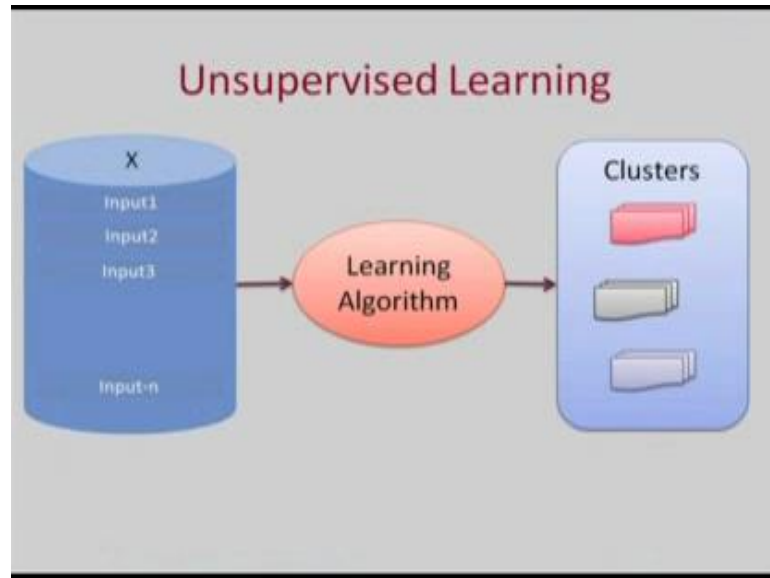
(Refer Slide Time: 03:13)
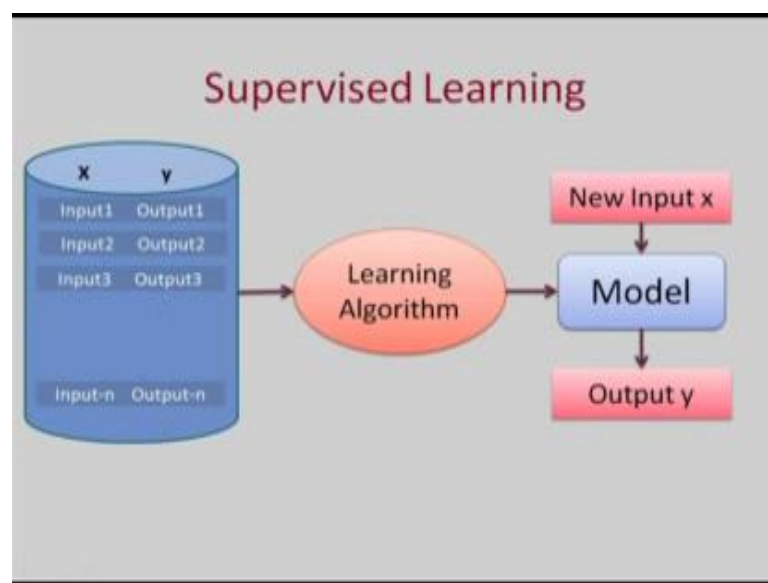
So, let us look at this picture which gives us a schematic diagram of supervised learning. In supervised learning you have a number of training instances. This contains the training instances, for the training instances each instance comprises of input and output. This is the first training instance, the second training instance, the third training instance, the nth training instance, given all these training instances the learning algorithm will come up with a model.

And this model can be used to classify or to find the output or corresponding y value for a new observation x. So given a new input x you can use the model to find out the output y. This is the schematic of supervised learning.
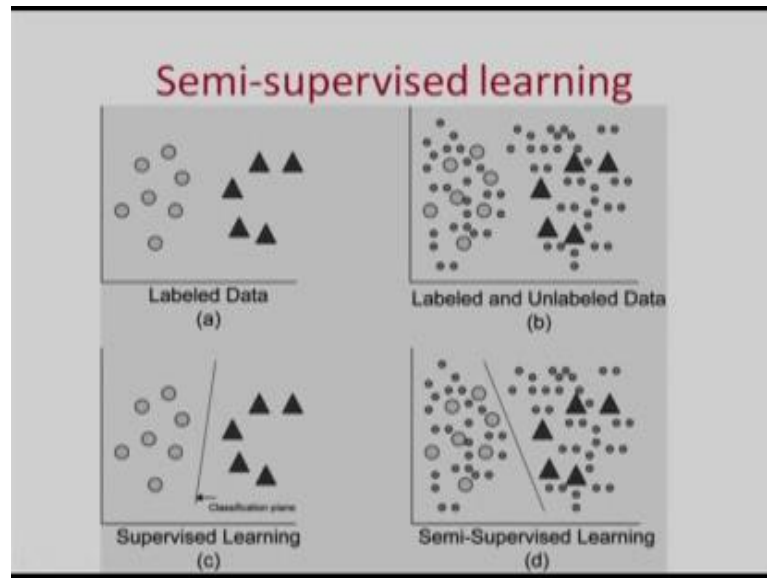
(Refer Slide Time: 04:12)

(Refer Slide Time: 04:17)
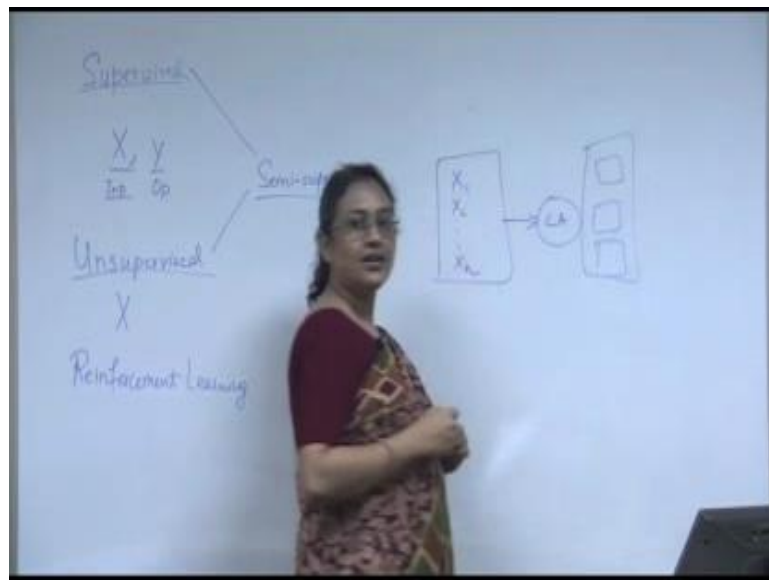
In unsupervised learning as we discussed, we only have x's. We have different x's; x 1, x 2, x n this is the data. And the learning algorithm will produce clusters will group this data. Based on the similarity of the data items to each other we can find out certain groups among the data. So that is called Unsupervised learning.

And as I said we also have semi-supervised learning. In semi-supervised learning, we have a combination of labeled data and unlabeled data. This is labeled data, the data which belong to two different classes so one class is circle the other class is triangle; in semi-supervised learning, apart from having data from the two classes you also have unlabeled data which is indicated by the small circles.

For example, for supervised learning based on the data. Supervised data you will come up with some function and if you also have unlabeled data in addition to the labeled data you might try to come up with the better function.
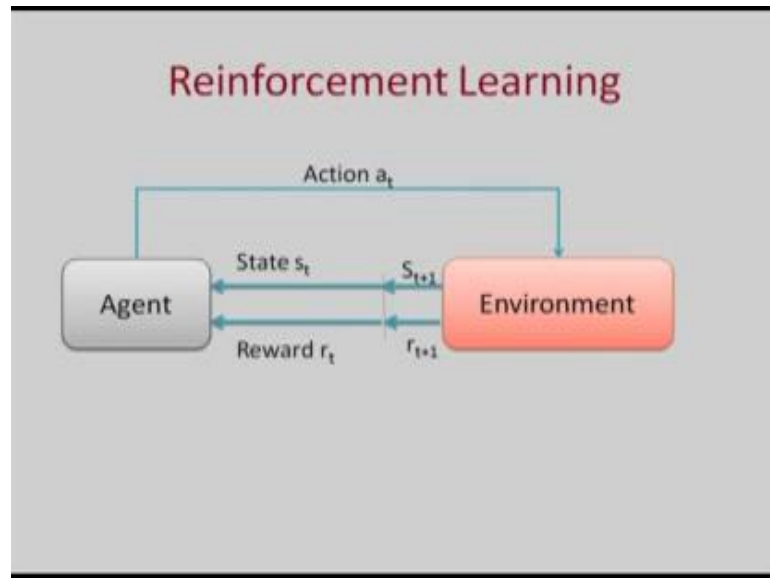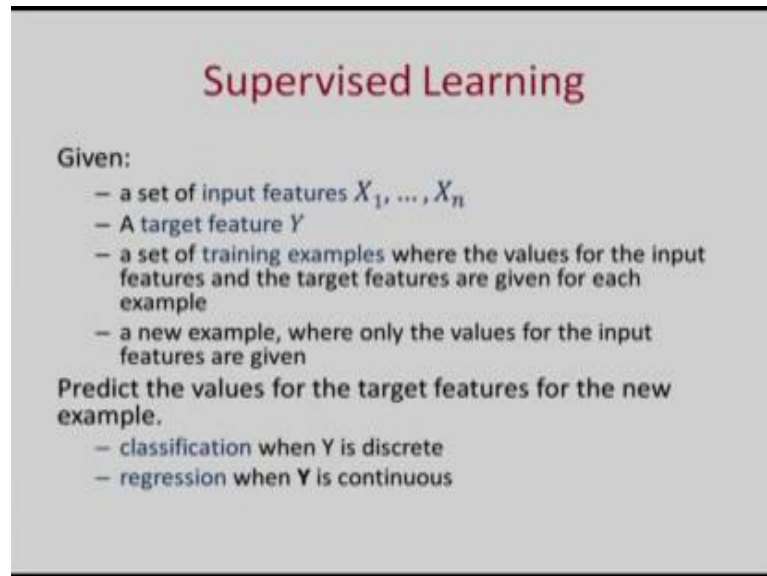
And we will very briefly talk about reinforcement learning; in fact we will not cover reinforcement learning in this introductory course on machine learning. But in reinforcement learning we have an agent which acts in the environment. The agent can take action and this action can impact the environment. In a particular stage, the agent takes an action and the environment goes to a new state and gives some reward to the agent, that reward may be a positive reward can be a negative reward or penalty or can be nothing at that particular time step. But the agent is continually acting in this world.
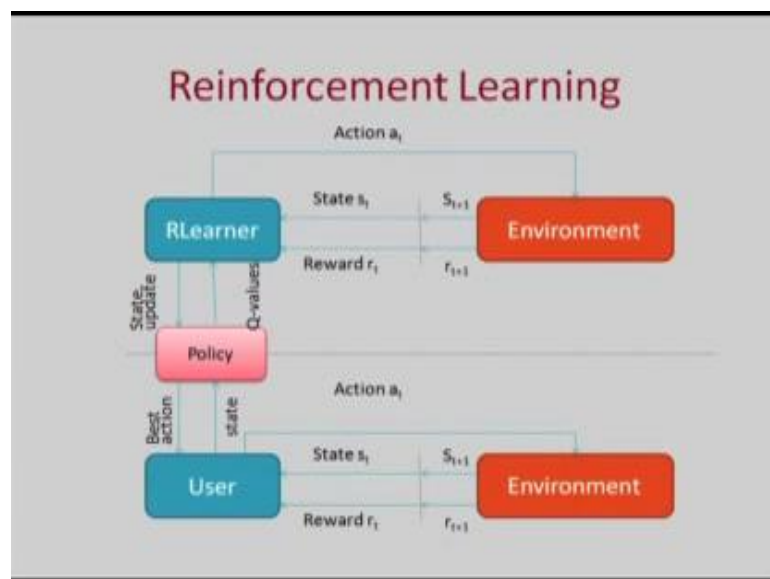
And the reinforcement learner what it will do is, it will learn a policy. That is, given a state what action to take so that not only the short term reward is optimized, but the overall utility of the agent over its entire time horizon is optimized. This is what reinforcement learning is about we will not talk about details of reinforcement learning in today's class.

(Refer Slide Time: 06:53)

Supervised Learning

Given:
- a set of input features $X_1, \dots, X_n$
- A target feature Y
- a set of training examples where the values for the input features and the target features are given for each example
- a new example, where only the values for the input features are given

Predict the values for the target features for the new example.
- classification when Y is discrete
- regression when Y is continuous

Most of this course we will talk about supervised learning and we will spend some time on unsupervised learning. So let us look at more details about what supervised learning is about. In supervised learning you have a set of input features.

(Refer Slide Time: 07:30)
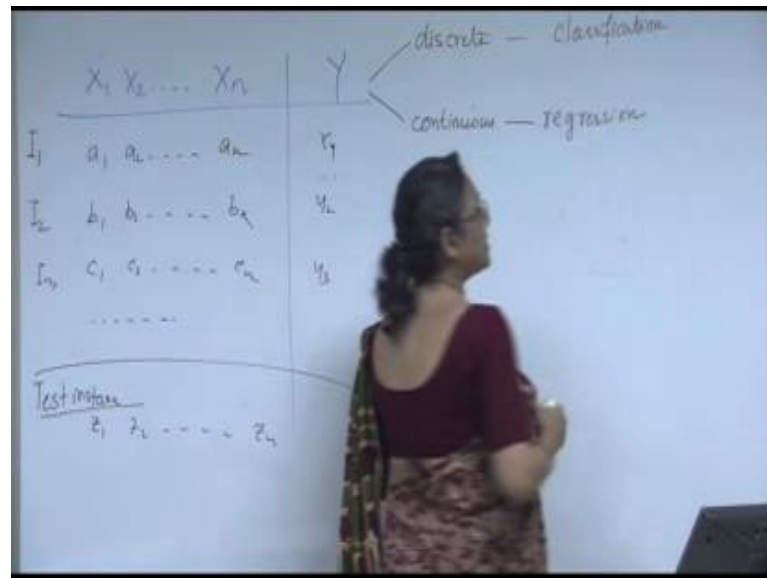
So what you have is you have some features x 1, x 2, x n, these are the features with respect of which you describe the instances. And you have a target fixture y, and you have several instances each instance comprises of values. Suppose, this is the instance 1 it comprises of values of this feature x so you may have a 1, a 2, a n, and you have instance 2 which is b 1, b 2, b n. Instance 3 which is c 1, c 2, c n. And the corresponding you have the y values; y 1, y 2, y 3, etcetera.

So, your instances are described in terms of features. There are input features and there is the output attribute. And you have a set of training examples this comprises the training examples, and in each training example the values for the input features and the target feature are given for each example. Each of these rows is one example and for each example we have the value for instance 3 the value for x 1 is c 1, the value for x 2 is c 2, the value for x n is c n and the value for y is y 3. This is used for training. During the testing time you are only given, so use like this you have many training examples for test you are given a test instance. In test instance you are only given the values of the attributes; z 1, z 2, z n and you have to output the corresponding y.

Now the y or the target feature can be a discrete valued feature or a continuous valued feature. If y is a discrete valued feature we call such problem classification problem. So y can be discrete valued or continuous valued. For example, we can have a discrete valued feature like whether it will rain or not tomorrow, such problems are called classification

problems. Or we can have a discrete valued feature which given the symptoms of the patient, predicts whether a patient has a particular disease. Or you have continuous valued feature then you call such problems regression.

Suppose given the values of certain features you want to predict the price of a house, so given the location of the house, the floor area of the house, the number of bedrooms and some such features, you want to predict the price of the house which is a continuous valued or real valued number and such problems are called Regression problems.

(Refer Slide Time: 11:16)
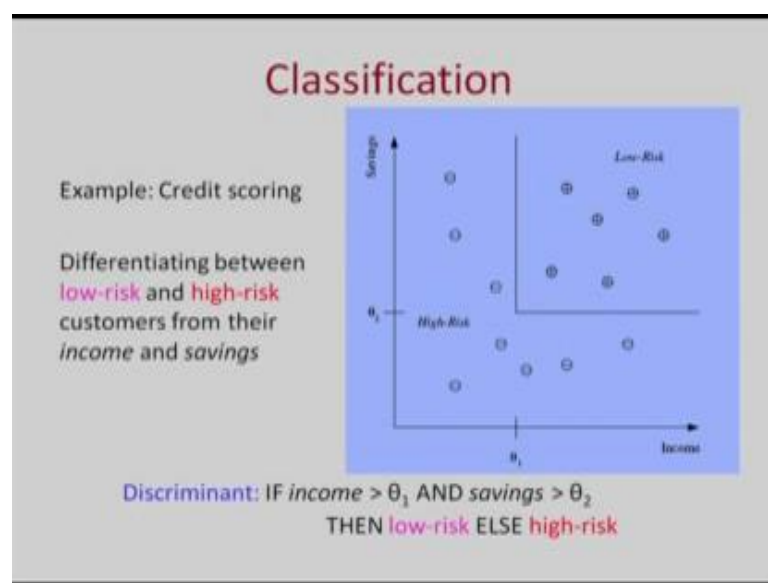
So, let us look at some examples of classification problem. One example is Credit scoring. Suppose, in this problem we are describing each input in terms of two features; income and savings. We have a person, we look at the income and the savings of a person and we want to predict whether this is a high risk person or a low risk person. And you seeing this picture, the high risk persons are labeled with minus and the low risk persons labeled with plus.

Now given this data you want to come up with a classifier which given the attributes of a new person will predict whether that person is high risk or low risk. And this is an example of a rule that you have found if income is greater than theta 1 and savings is greater than theta 2 then the person is low risk else the person is high risk. This rule you have come up with the rule based on the data. In this case you can visualize the data and come up with the rule.
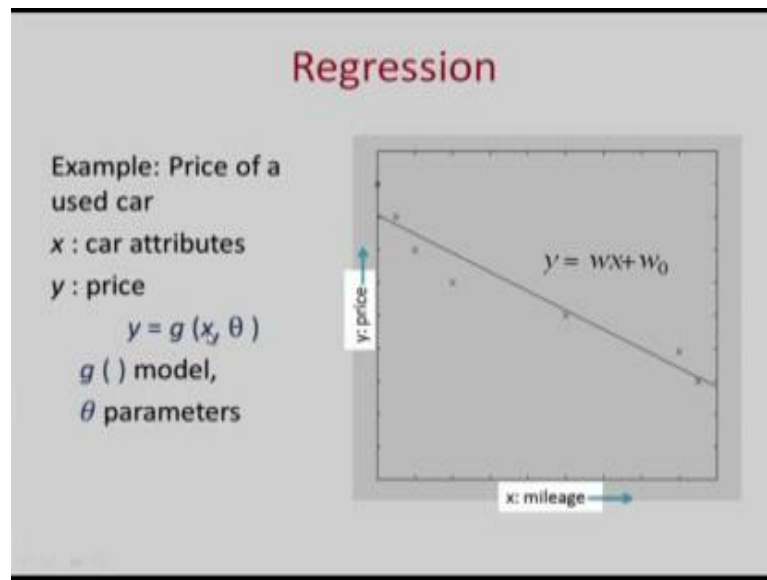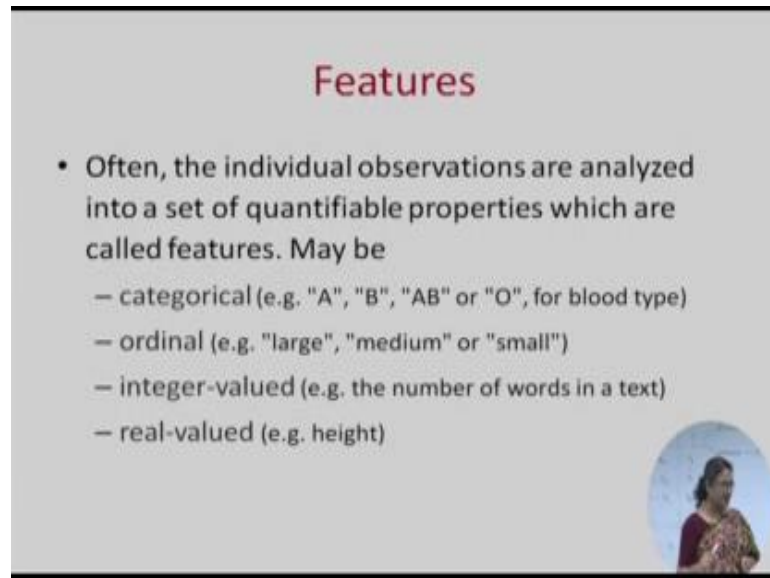
Now let us look at an example of a regression problem. You want to find out the price of a used car and you use certain attributes of the car to predict its price. For example, let us say that you have only looking at the mileage of the car and based on the mileage you have predicting the price. And these different points they correspond to a particular car show what is the mileage and what is the corresponding price.

So given the mileage price of several cars you have the data points and you can come up with the function so that given a new car whose mileage is given you can predict the price. In regression you come up with a function which takes the input instance and the parameters of the model.

Now, we see that in order to describe an instance features play a very important role. Observations that you see are analyzed into a set of quantifiable properties which are called features. So if you want to predict the price of a car or the price of a house you have to have the right features which will help you to come up with the price. If you want to predict what disease a patient has you have to come up with the right tests and parameters which will enable you to make that prediction.

Now feature are of various types. Features can be categorical. For example, blood group is a feature which has four values "A", "B", "AB" and "O", or gender is feature which has certain values like, male, female etcetera. Or features can ordinal like, large, medium, small. Features can be integer valued like, the number of words in a text or a feature can be real valued like height.

Example Data

Training Examples:

	Action	Author	Thread	Length	Where
e1	skips	known	new	long	Home
e2	reads	unknown	new	short	Work
e3	skips	unknown	old	long	Work
e4	skips	known	old	long	home
e5	reads	known	new	short	home
e6	skips	known	old	long	work

New Examples:

	Action	Author	Thread	Length	Where
e7	???	known	new	short	work
e8	???	unknown	new	short	work

Now, let us look at this table which shows a sample training examples and for these training examples you have five features; action, author, thread, length, where. And the different rows are the different instances. So instance e 1 says; action is skips, author is known, thread is new, length is long, and the person is at home. These are the different instances based on the training examples comprising of this labeled instances you learned a model so that for a new instance you can predict the output feature.

Suppose, you have to predict the action of the user whether the user skips a link or reads a link and given the value of Author, Thread, Length and Where.

(Refer Slide Time: 15:25)
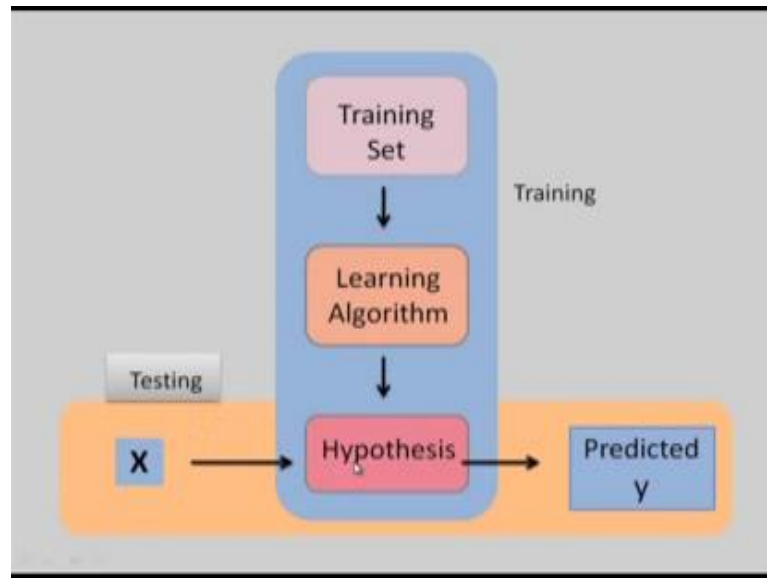
So, if you come up with the schematic diagram in supervised learning you have the training set, the learning algorithm uses the training set to come up with a model or hypothesis, will introduce this term in more detail later on. And in the testing phase given a new instance you use the hypothesis to predict the value of y.
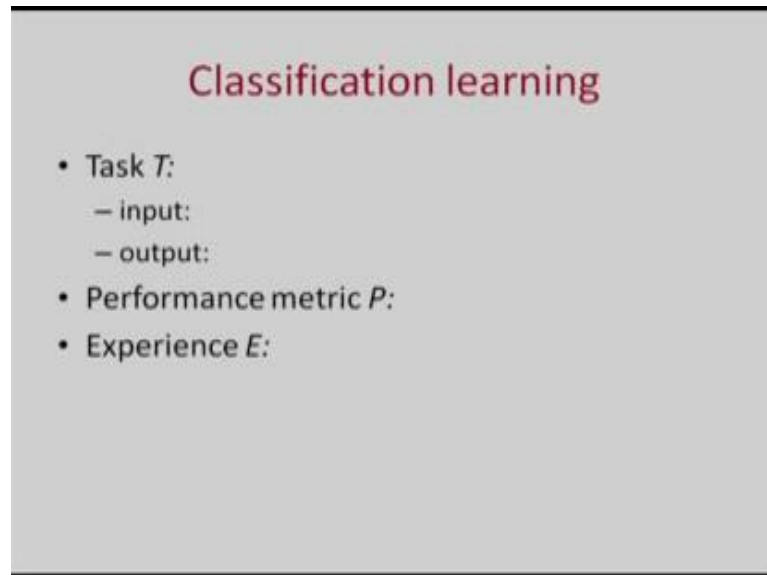
(Refer Slide Time: 15:49)
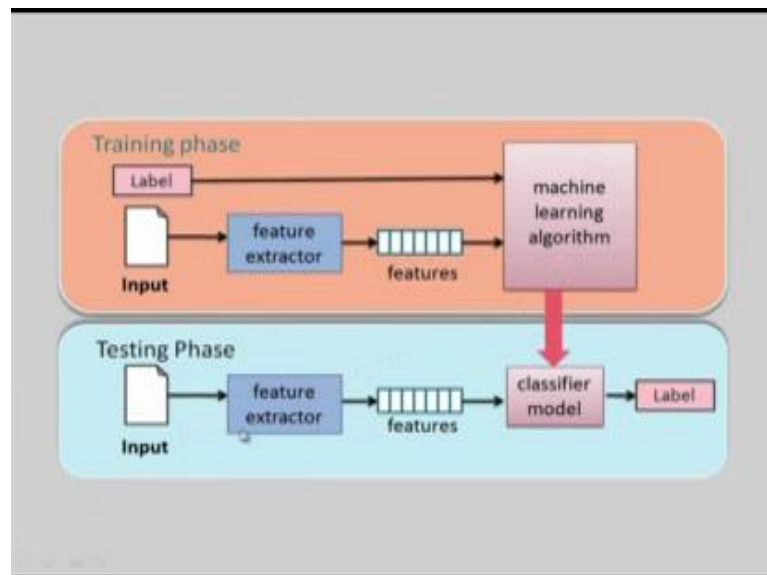
This can also be shown like this, in the training phase you get the input and the label. From the input you extract the features of the input and feed it to the machine learning

algorithm. Similarly in the testing phase given the input use a feature extractor to extract the features and you feed it to the classifier model to get the label.
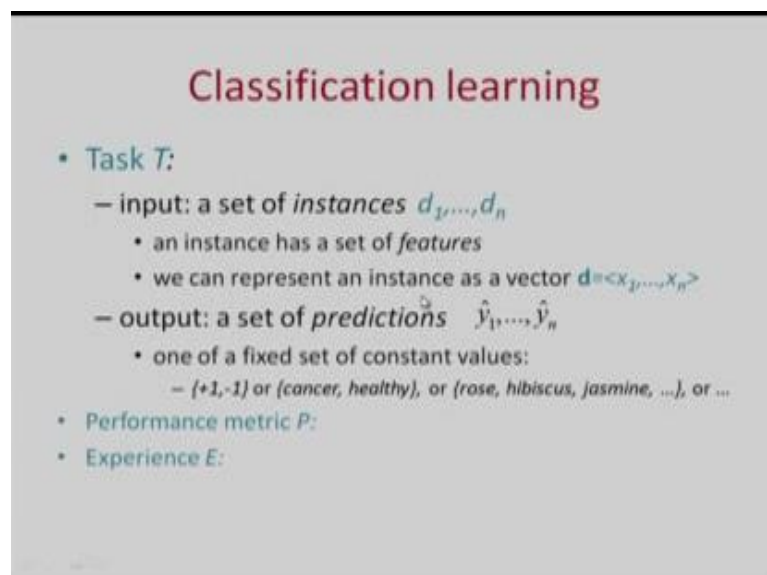
(Refer Slide Time: 16:12)

So in classification learning these are the components. You have a task T which has input and output, there is a performance metric P, and there is experience E.

(Refer Slide Time: 16:23)

The task T comprises of input which is a set of instances d 1, d 2, d n, each instance has a set of features we can represent an instance as a vector d equal to x 1 x 2 x n and you have the output a set of predictions for the inputs.

(Refer Slide Time: 16:44)

Classification Learning

| Task	Instance	Labels
medical diagnosis	**patient record:** blood pressure diastolic, blood pressure systolic, age, sex (0 or 1), BMI, cholesterol	{-1,+1} = low, high risk of heart disease
finding company names in text	**a word in context:** capitalized (0,1), word-after-this-equals-Inc, bigram-before-this-equals-acquired-by, ...	{first,later,outside} = first word in name, second or later word in name, not in a name
brain-human-interface	**brain state:** neural activity over the last 100ms of 96 neurons	{n,s,e,w,ne,se,nw,sw} = direction you intend to move the cursor
image recognition	**image:** 1920*1080 pixels, each with a code for color	{0,1} = no house, house

Now, what we will do is that we will look at some sample classification tasks. Let us say the task is medical diagnosis, instance are patient record, comprising of the features of blood pressure, diastolic and systolic, age of the patient the sex of the patient, the BMI of the patient, cholesterol of the patient. And the labels are whether the person has is low risk of heart disease or high risk of heart disease, there are two classes we call them positive class and negative class. One is the positive class; the other is a negative class.

Then there are some other examples finding company names in text. So given a particular word you want to find out is it the name of a company. This is the entity recognition task in text (Refer Time: 17:42). So, the instance is a word and the context of the word. What words, which words come before the word, which words come after the word, and may be certain features of the word.

For example, is the word capitalized, is the word following this word are i and c, is the bigram or two words previous this word context acquired by and so on. And the output labels are; first, later, outside. That is, if it is a company name is it the first word in that name, is it inside a company name, is it an outside a company name.
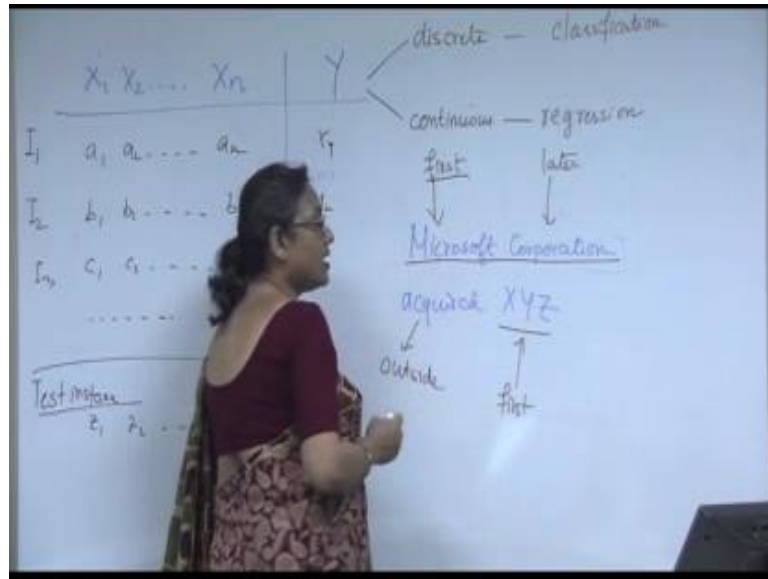
For example, in a running text you may have Microsoft corporation acquired x y z. Let us say Microsoft corporation together is the name of a company, and this is the first word in the name, this is a later word in the name. And let us say xyz is also a company name and this is also the first word in the name and the word acquired is not a company name so it is outside. We may use this sort of convention in order to label company names in text, and we can have a machine learning algorithm finding these labels.

Similarly, let us look at the fourth row which is image recognition, the input are image which is 1920 by 1080 pixels and each pixel has a color. And your output is whether the image is the image of a house or not.

So the performance metric in classification learning can be, what is the probability of wrong prediction, what is the probability or it can be the performance metric accuracy or the probability of wrong prediction of the error on examples from the distribution from which the instances are drawn. In classification learning the task comprises as we said a set of input instances d 1, d 2 d n and the output is a set of predictions y 1, y 2 y n.

The performance metric is what is the probability of wrong prediction and the experience is a set of labeled examples where y are the true labels for x for those training examples we have the ground truth data. And these examples come from some fixed distribution. During the training this examples are drawn from a distribution and what is expected when we get text instances, those examples also come from the same distribution.

Classification Learning

| Task	Instance	Labels	Getting data
medical diagnosis	**patient record:** lab readings	risk of heart disease	wait and look for heart disease
finding company names in text	**a word in context:** capitalized, nearby words, …	{first,later,outside}	text with company names highlighted (say by hand)
brain-human-interface	**brain state:** neural activity over the last 100ms of 96 neurons	{n,s,e,w,ne,se,nw,sw }	recordings of someone doing known tasks (so direction can be inferred)
image recognition	**image:** pixels	no house, house	hand-labeled images

can you get examples some other way?

Now, how do you get data for the learning problem? For example, in medical diagnosis problem which we have seen, in order to get data when you have a patient you find out the different parameters of the patient and we have to wait for the next several years to find out did the patient get heart disease or not. So, we have to wait to look further heart disease.

Finding company name in text, how do you get the data? You get some text and you get the manually annotated for company names this forms the training set. For image recognition the data is a set of images for which some human or some people have labeled those images. So, this is how you get data.
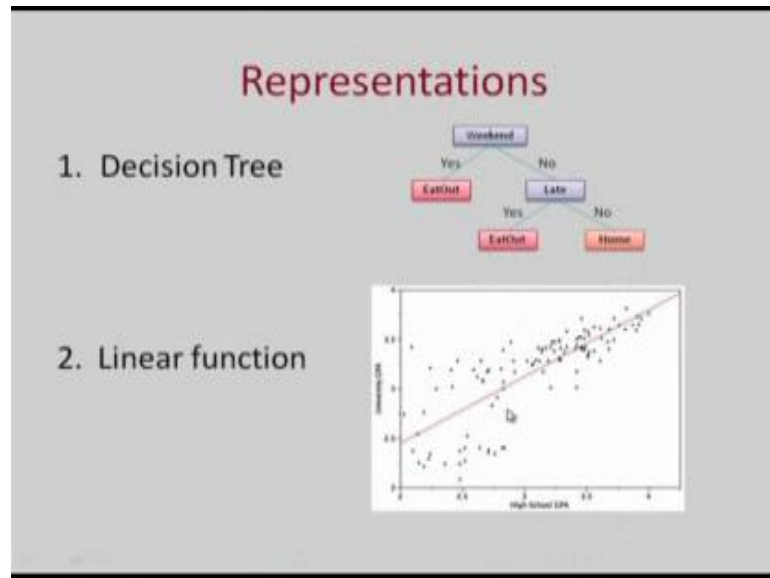
As we said that when we want to look at the representation of the function there are two things; one is the features, the other is the function class. And we will talk about this in greater detail later. The function class can be a linear function in terms of the attributes; the function can be a decision tree which we will talk about in week 2. It can be a linear function; it can be a decision tree.
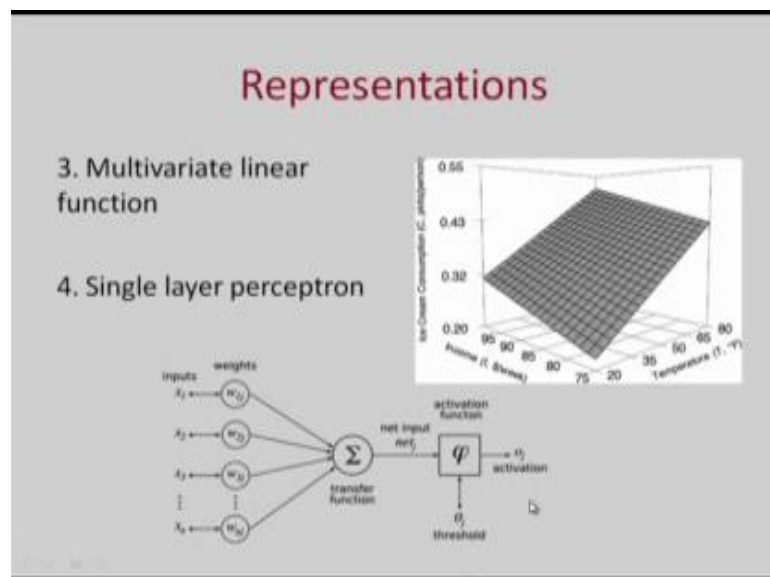
Or, it can be other functions like multivariate linear function as shown in the picture or single layer perceptron which is a unit of a neural network which we will talk about later in this class.

(Refer Slide Time: 22:52)
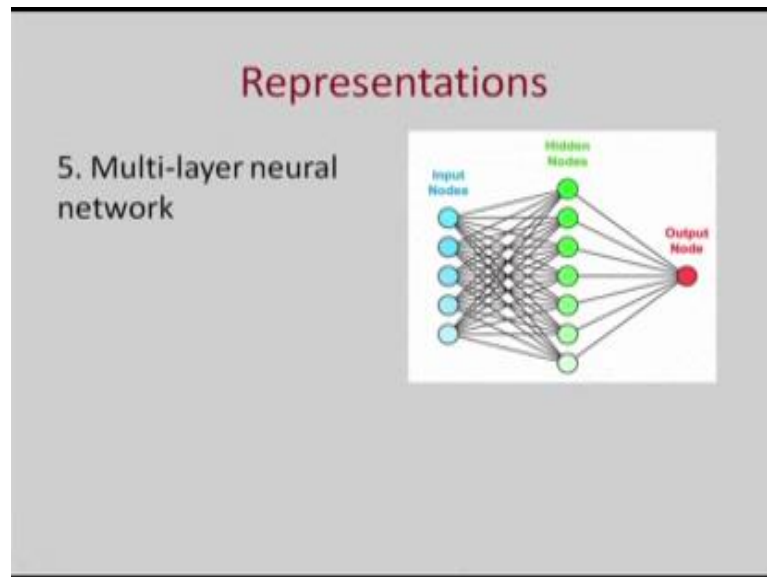
Or it can be multilayer neural network which we will also cover later.

(Refer Slide Time: 22:57)

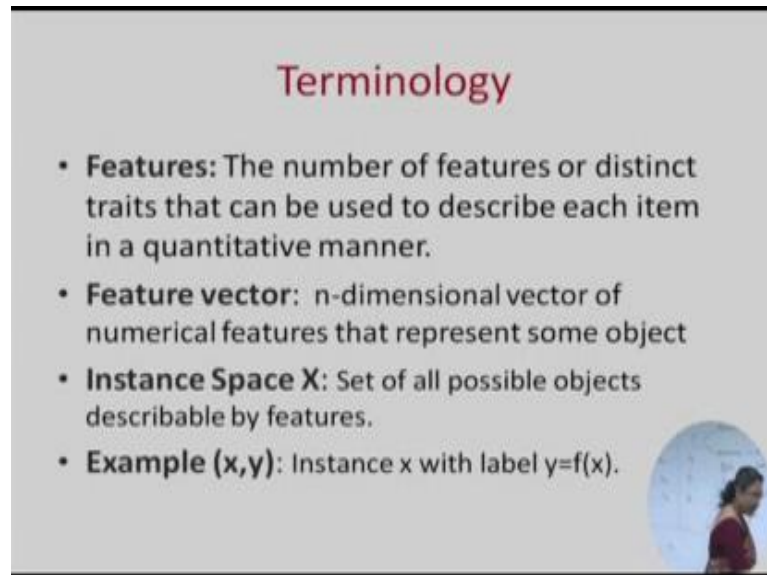And given the type of function, the hypothesis space is the set of candidate outputs that you can get, candidate functions that you get. So supervised learning you can think of, there is a set of functions which comprise the hypothesis space and you want to find out

that function from the hypothesis space which is most probable given you a training example. Again we will talk about this in detail later.
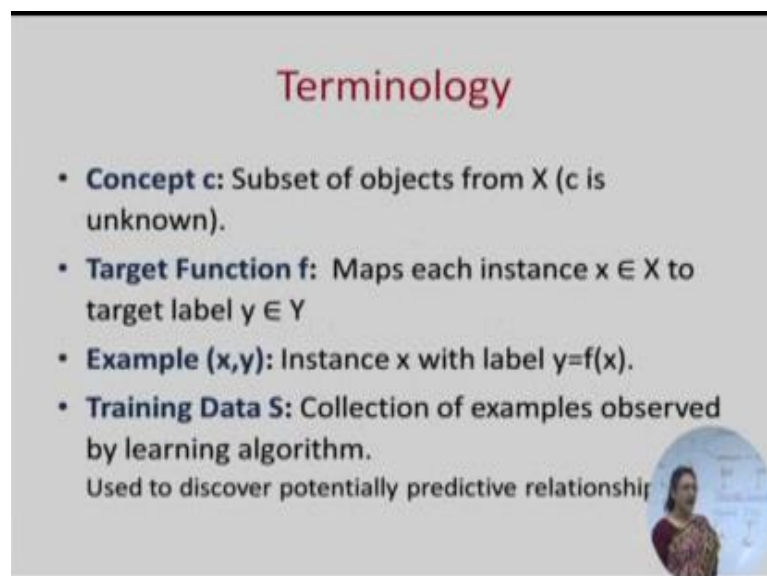
(Refer Slide Time: 23:36)

So, if you just look at the basic terminology we have some features which are distinct traits or attributes that are used to describe each instance. And the set of features they comprise the feature vector. The instance space is a set of possible objects that you can describe by the feature, and each example has the value of the input and the value of the output.

(Refer Slide Time: 24:04)

Suppose you are try to find out is this the object at image of a phase. So, the concept is a subset of objects and the instance space, among all possible images some images a subset of the images and images of phase. The target function is a function that you are trying to learn it will map each instance to its label. That is it a phase or is it not a phase. We have already talked about what is example and what is training data.

So with this we end today's module and we will star with the next module in the next class.

Thank you very much.