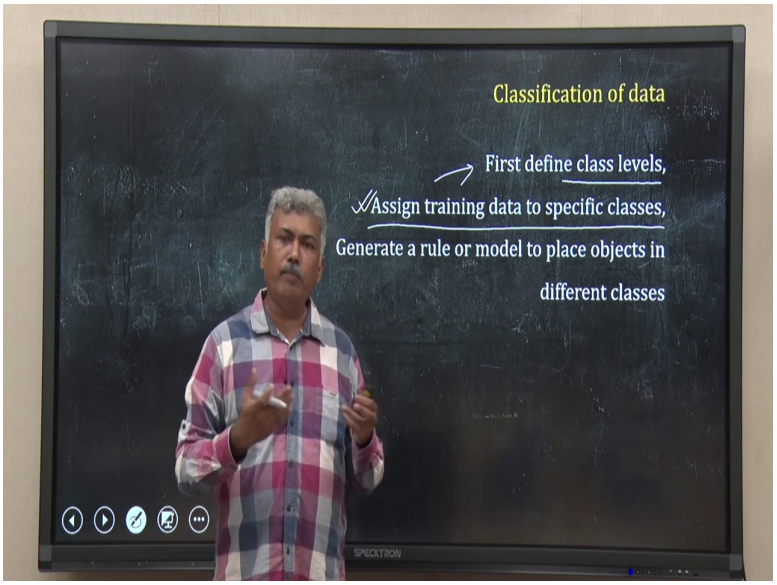
Data Analysis for Biologists
Professor Biplab Bose
Department of Bioscience & Bioengineering
Mehta Family School of Data Science & Artificial Intelligence
Indian Institute of Technology Guwahati
Lecture 35
Logistics Regression

(Refer Slide Time: 0:30)

Hello everyone, welcome back. In this lecture, we will discuss about logistic regression.

(Refer Slide Time: 0:48)



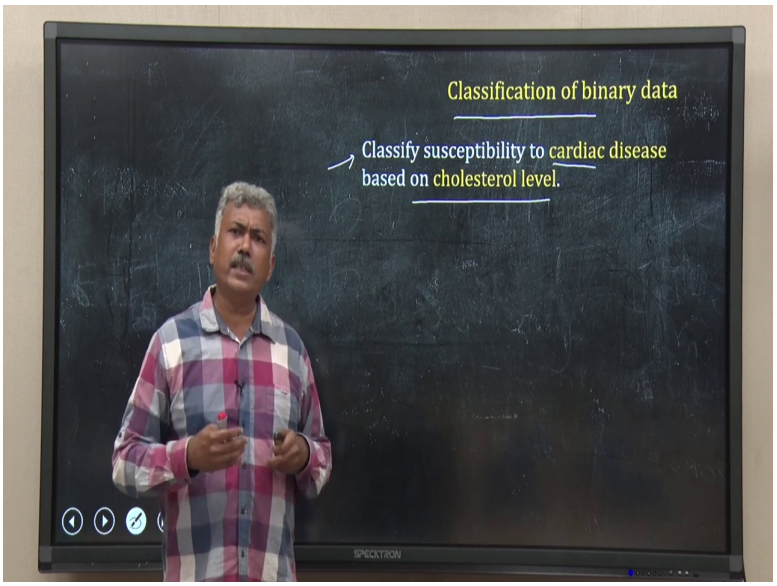
Logistic Regression is a type of classification technique. If you remember, we have discussed earlier the difference between classification and clustering. In classification, what we have,

we have an object or a data point, we want to put it in a particular class. So, to do that, we have to start with a training data set. And we should first define different classes, for example, we may define three different classes or two different classes, like for example, diseased person, non-diseased person, two classes.

So, we will have two class labels. So, we will define class labels. Then what we will do, we will take a data set which we will call training data set, which is labelled that means, for this diseased non diseased case, I know, who are the person in my training data set have the disease and I also know those persons in my training dataset who does not have the disease. So, we assign a training data to specific classes. So, we have labelled data sets.

And then using this labelled data set, what we do, we create a classifier, or a mathematical model, so that that model can predict the class of an unknown sample or an unknown object. So, taking the example of disease and non-disease case, if I can create a predictive classifier model from the training data set where the data is labelled, I should be able to predict a new person coming to my clinic, whether that person has the disease or not.

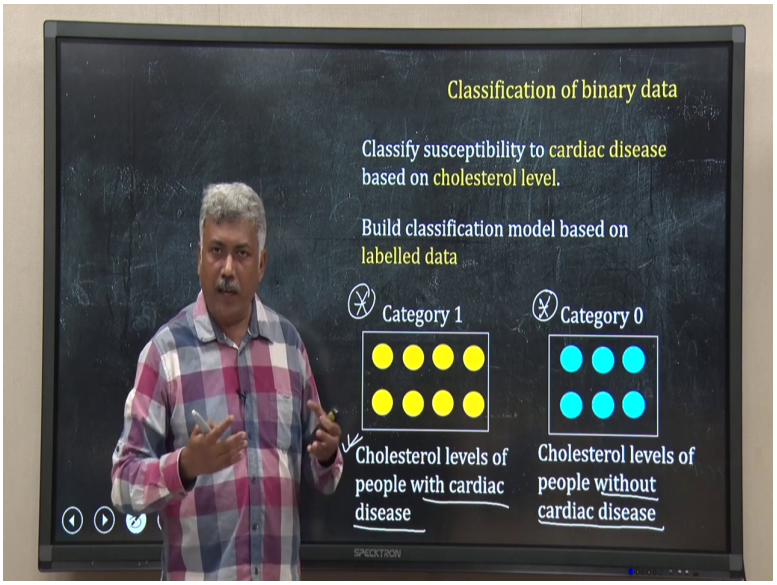
(Refer Slide Time: 2:26)



Now, data classification can be of different type. For example, the simplest one is that you have just two classes. Let us start with that, we call that binary classification, classification of binary data. For example, what I have written here, for example, consider that you want to classify susceptibility to cardiac diseases based on the cholesterol level of a person.

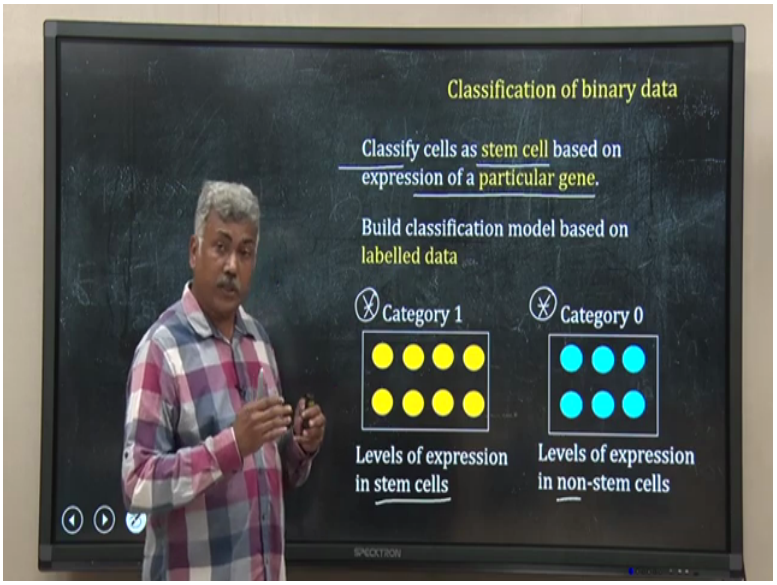
That means I want to create a classifier model or classifier, which will help me to decide when a person comes to my clinic just by measuring the cholesterol level of that person, I should be able to decide whether that person is susceptible to a cardiac problem or not, so, it is a binary problem, either the person belongs to the class of susceptible to cardiac problem, or he or she belongs to the class of not susceptible.

(Refer Slide Time: 3:19)



To create the classifier, we need labelled data. So, we need two type of labelled data. So, we need the data of category 1, in category 1, I will put all those persons who have cardiac disease. And I know the cholesterol level of all those persons. Its binary so, I have another category or class category 0. In category 0, I have all the persons who do not have that cardiac disease. And I also know the blood cholesterol level for these people. So, I will use this labelled data set to create my classifier model.

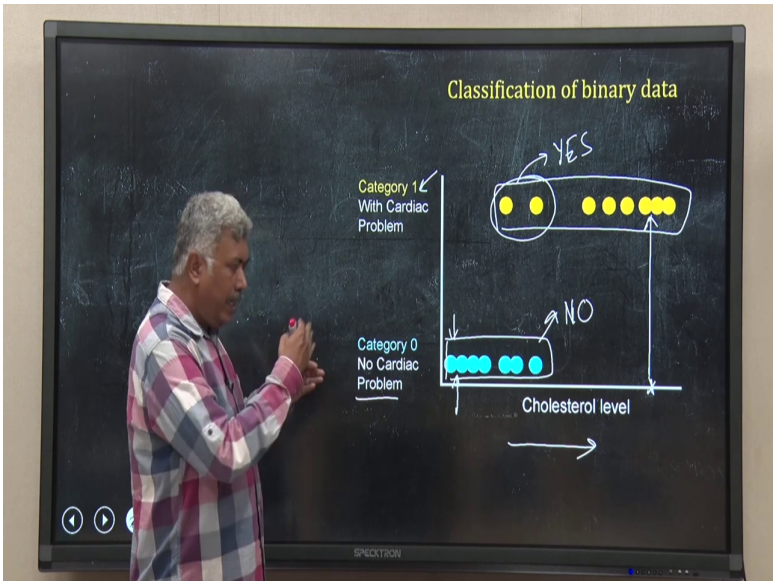
(Refer Slide Time: 4:03)



Take another example, to understand this issue. For example, I want to classify cells, whether they are stem cells or not based upon the expression of a gene. For example, a particular stem cell marker. So, my problem is to classify cells, either a stem cell or not based on expression of a particular gene, a stem cell marker. To solve this problem, to create the classifier, I need a labelled data set with two categories or two classes. Category 1, and category 0.

In category 1, I will put I will put all those cells which are stem cells. And in category 0, I will put non stem cells. So, in both this category 1 and category 0, I know the level of expression of the stem cell marker gene. So, now, let us move further with the example of that blood cholesterol level problem where we want to classify people whether they are susceptible to the disease or not, cardiac disease or not based on the cholesterol level. I will represent the data in a visual format.

(Refer Slide Time: 5:08)

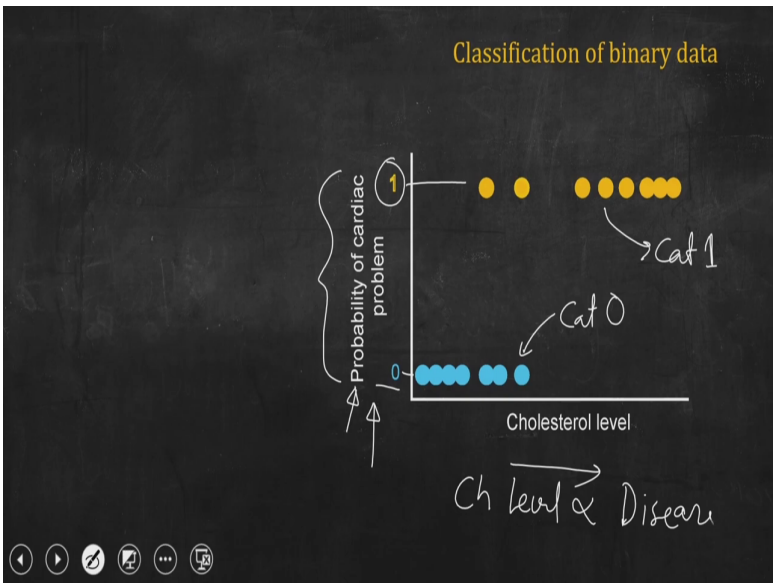


So, what I have done, I have taken the data of the training set. So, these are my people who do not have the have the disease, so, they are YES category 1 so, they have the disease whereas, these people the blue they do not have the disease, so, they are category 0, they do not have the cardiac problem. And I know their cholesterol level, so, cholesterol level is in the horizontal axis.

So, you can easily see people having a high cholesterol level they are already in the class or category of susceptible to the cardiac problem whereas, people having low cholesterol level are in the category 0 and they do not have the cardiac problem. There are some outliers obviously, for example, these two people have low cholesterol level, but still, they have cardiac problem. So, all data will have this type of outlier.

Now, we want to build a classifier model, so that will be numerical model. So, what we have done here, I have just categorized two data and visually represented, but to make it more quantitative numerical, I will just rearrange the data, same data in a different way.

(Refer Slide Time: 6:22)

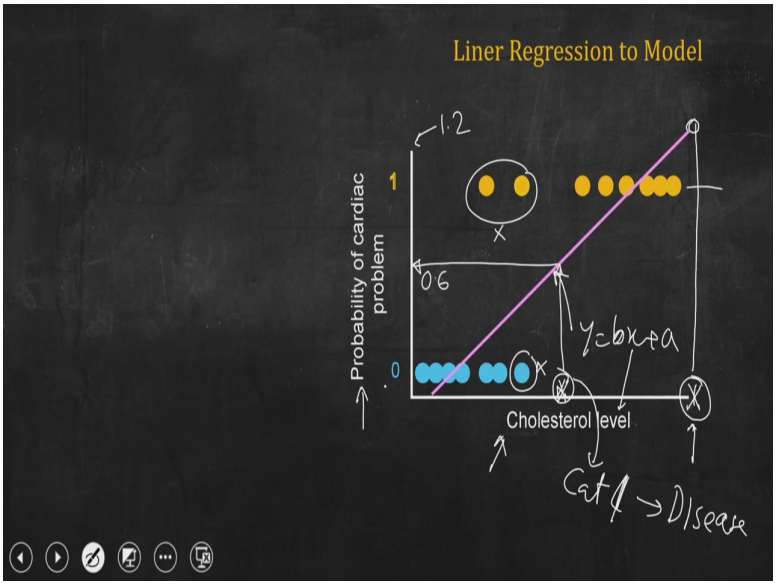


What I will do, I will change this vertical axis, in the horizontal axis I still have the cholesterol level. And what I have done, I have scaled this from 0 to 1. And what is there in the vertical axis, I have the probability of cardiac problem, probability of having cardiac problem, in other words probability or the susceptibility of having the cardiac problem.

So, this category 1 people we already know they have cardiac problems. So, they belong to probability 1. Whereas those people in category 0, I know they do not have the disease, so they all belong to the probability 0. So, I have just a scaled vertical axis. Now, once I have this representation, I can try to fit a model.

And when we talk about fitting a model, immediately it comes to our mind that we may do regression for example, linear regression. You may assume that okay the cholesterol level, the cholesterol level may have some linear relation with suppose the disease, cardiac disease. So, you want to fit a y equal to bx plus a type linear equation to this data set. And I have done that.

(Refer Slide Time: 7:51)



So, this is my linear regressed line, the pink one. If I leave these outlier, if I leave them and even if I leave this one, you can visually see its quite a good fit. So, now, if this is my classifier model, this is a linear classifier model, which gives a linear relation between cholesterol level and the probability of having the cardiac problem, suppose I have this model with me now, I know the equation that is y equal to bx plus a, where x is the cholesterol level, y is the probability.

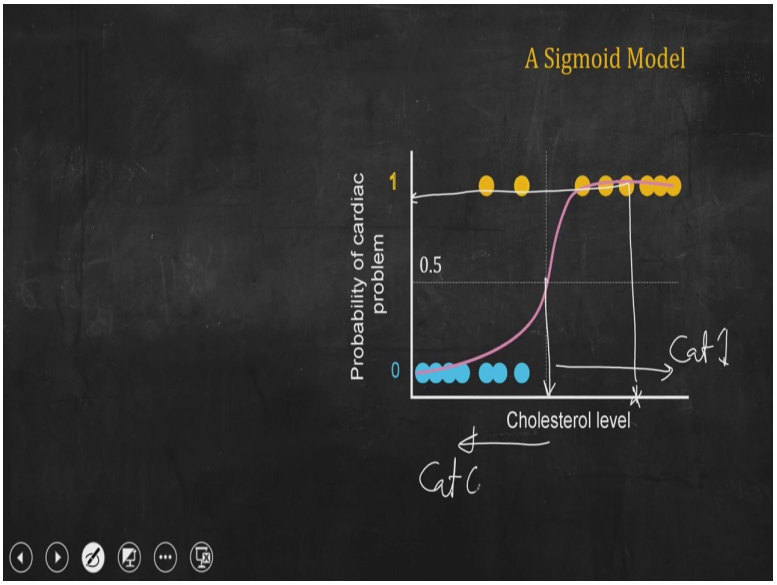
So, if I have this model now, and suppose a new person comes to my clinic, so that person has suppose the cholesterol level here, then I can use the straight line to predict whether that person belongs to disease category or not disease category. For example, if I consider this came to 0.6, so that means the probability that the person will have the cardiac disease is 0.6, it is about 0.5, I can consider 0.5 as a cutoff and I can say that person belongs to category 1 or diseased, this is so simple to do.

But if you look carefully this linear model, this linear classifier model that I have done by linear regression has a problem. Let me explain. Suppose I have one person in my clinic whose cholesterol level is very high here. Now, if I go up vertically, then I reach somewhere here and then if I go horizontally from there, I will reach somewhere here and this will be something like suppose 1.2.

Now probability cannot be bigger than 1. So, this is mathematically not possible, it does not make any sense that the probability of that person to have the disease is bigger than 1. So, this is one unique problem, if we fit a linear regression model to this data set. How can I avoid

this problem? Rather than fitting a linear equation, I can fit some other equation which will be bounded between these 1 and 0.

(Refer Slide Time: 10:32)

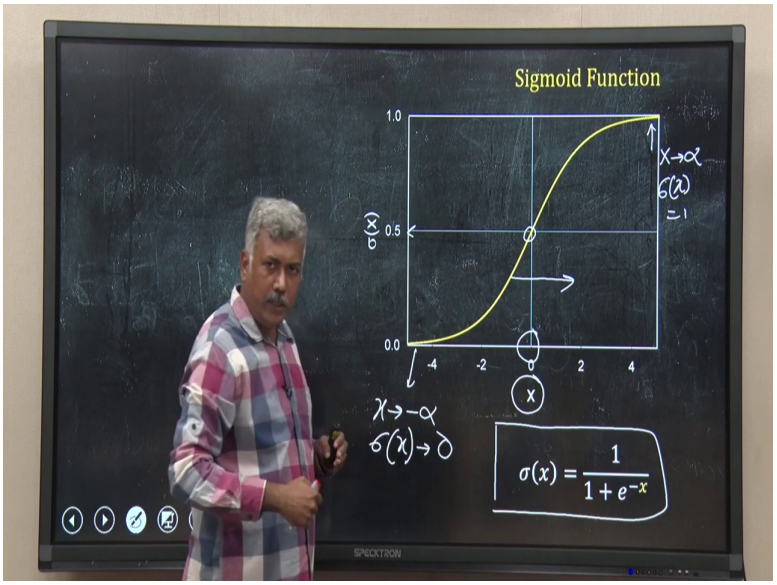


And one such way of doing that is to use a sigmoid function, something S shaped. So, I have a sigmoid function here, this is the sigmoid. So, now, I have fitted that sigmoid, I will come back on, we will come to the equation of that sigmoid, for the time being consider we know it. So, if this pink line the sigmoid S-shaped curve, if I say, then I can see this pink does not go beyond 1 and it does not go beyond 0 either.

So, now, if I have this model, then if somebody comes with a cholesterol level somewhere here, then I can use this pink line to predict the probability that the person will have the cardiac disease and based on the cutoff, for example, I can use 0.5 as the cutoff, I can say, his cholesterol level is this one, so the probability is this one, so that person has a probability above 0.5, so that person belongs to the category 1, that is the category of having cardiac disease.

So, any point, any cholesterol level, beyond this cutoff will belong to category 1. So, they all belong to category 1. As you can easily see, and all cholesterol level, all the cholesterol level below this cutoff will be category 0, that is no disease. So, in this case, we have solved the problem that the probability was going earlier for linear regression case beyond 1 and 0. Now, what type of sigmoid function I should use, there are many types of sigmoid function possible in logistic regression, we use a particular type of sigmoid function.

(Refer Slide Time: 12:20)



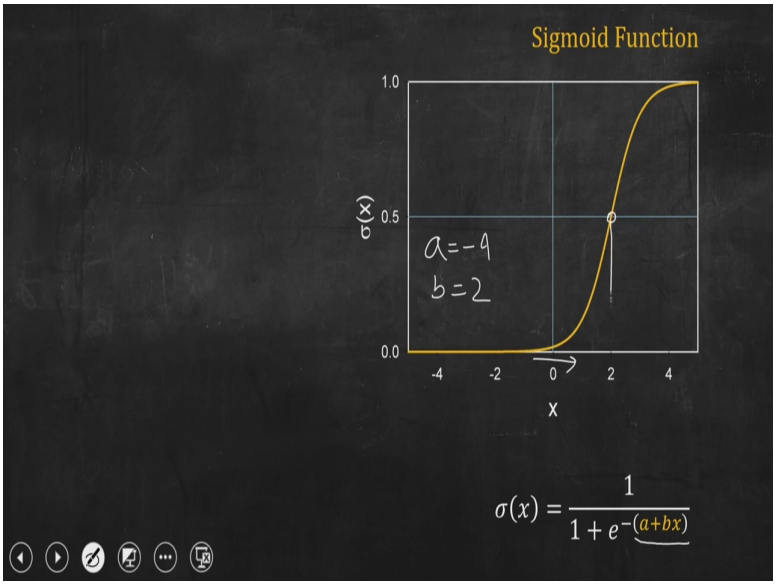
$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Let us look at that. So, the function that we use for logistic regression is a sigmoid and its form is given here, 1 divided by 1 plus e to the power minus x, and I have plotted the graph for the sigmoid. You can see when x becomes very big, so suppose x tends to infinity, then this sigmoid, sigma x will become 1. So, it saturates at 1. So, this function does not go above 1.

Whereas in this case, where x tends to minus infinity, that means it become very small, then this function becomes reaches is close to 0. So, it is bounded on the lower side also at 0. And in between it is centered at 0 and for that its found value is 0.5. So, we have solved this problem that my function is bounded to 1 and 0, but you must have noticed here that it is centered around 0.

Now, in most cases, in most real-life cases, our predictor, the predictor variable x, for example, the cholesterol level will not be negative, it will be always positive. So, if it is always positive, that means I have to shift this curve on the right and on that side, positive side, because I cannot have x negative. So, how can I do that? Doing that is very simple, I have not to change much of this function, actually, I had to write it in a different fashion.

(Refer Slide Time: 13:56)

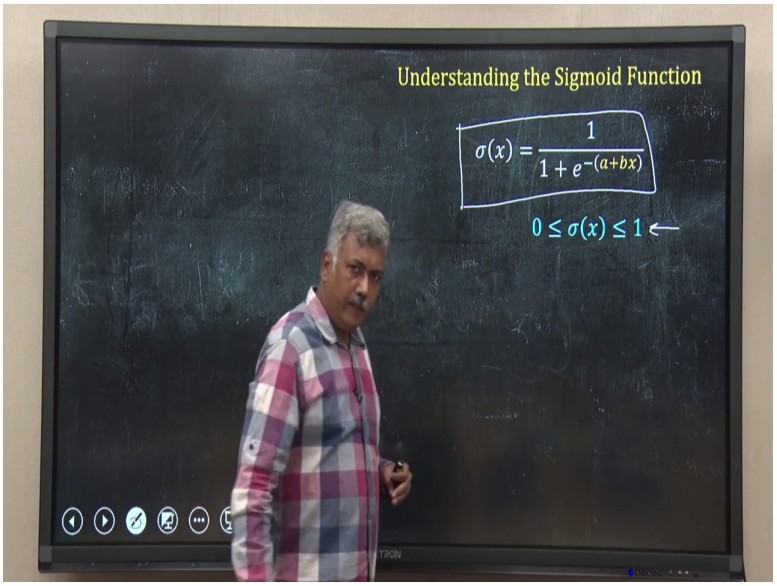


$$\sigma(x) = \frac{1}{1+e^{-(a+bx)}}$$

So, what I will do, rather than using x, I will use a plus b x, just like a linear equation. So, my function will be now, 1 divided by 1 plus e to the power minus a plus b x, and in this case, what I have done in this diagram, I have considered a equal to minus 4 and b equal to 2.

And you can easily see that the curve has shifted towards the positive side and it is centered around this point 2. So, in this way by choosing the right value of a and b, I can actually easily fit the sigmoid function to my data set, which I am using as a training data to create that classifier.

(Refer Slide Time: 14:48)

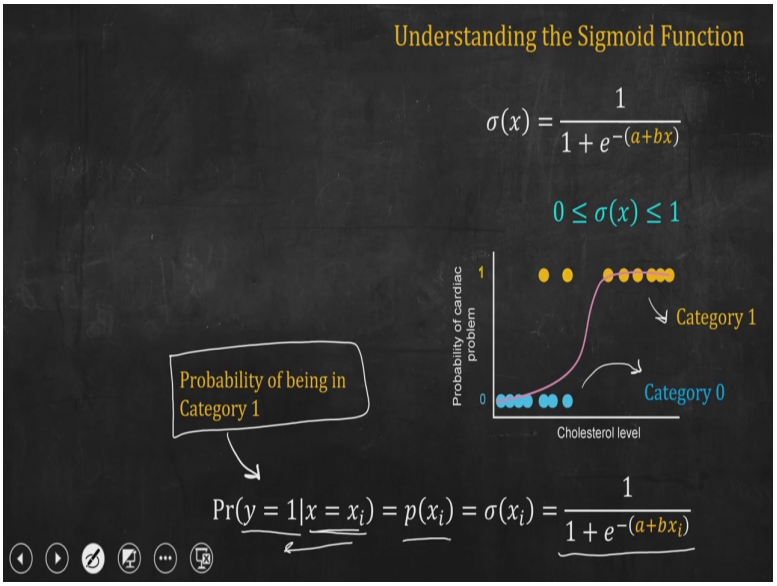


$$\sigma(x) = \frac{1}{1+e^{-(a+bx)}}$$

$$0 \leq \sigma(x) \leq 1$$

Now, let us dig into more into the sigmoid function, there are some interesting features of that. So, what I have, I have this generalized sigmoid function, 1 divided by 1 plus e to the power minus a plus b x. And this sigmoid function, sigma x, varies from 1 to 0, it is bounded from 0 to 1.

(Refer Slide Time: 15:08)



$$\sigma(x) = \frac{1}{1+e^{-(a+bx)}}$$

$$0 \leq \sigma(x) \leq 1$$

$$\Pr Pr\left(\frac{y=1}{x=x_i}\right) = p(x_i) = \sigma(x_i) = \frac{1}{1+e^{-(a+bx_i)}}$$

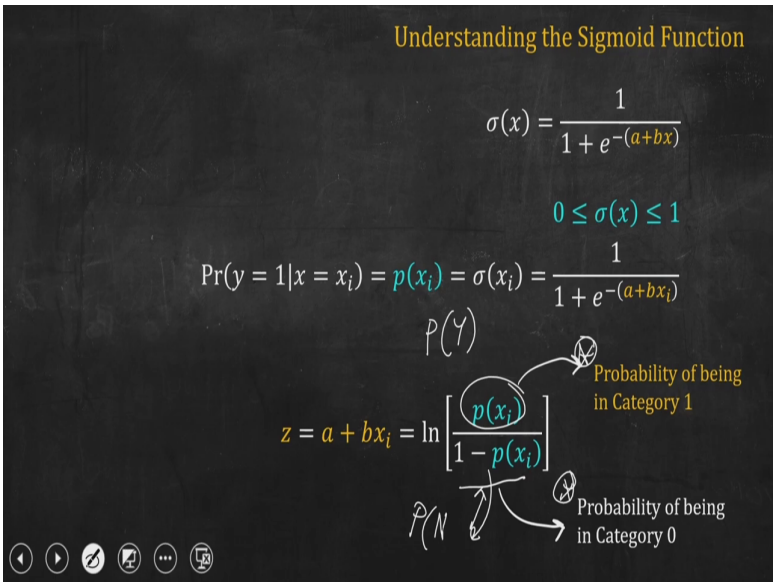
Now, if you carefully look into my data, what I have? I have 2 data set, with category 1 and category 0. And on the vertical axis, I have probability of having the disease and the vertical axis varies from 0 to 1. So, my sigmoid function also varies from 0 to 1. Now probability cannot be bigger than 1 less than 0, so, as the sigmoid function. So, in a way, this sigmoid function itself is giving me the probability, probability of what?

Let us check, if somebody comes with a cholesterol level this one, then using the sigmoid function, I will be able to calculate the probability of having the cardiac problem. So, I can write using mathematical notation, I can write the probability that y equal to 1, because remember, y is the category and I have two categories 1 and 0, 1 mean having the disease. So,

the probability of y equal to 1, given x equal to x i, x i is the value of the cholesterol level for that person.

Probability of y equal to 1, given x equal to x i, is given by 1 divided by 1 plus e to the power minus a plus b, x i, the sigmoid function. So, this is essentially nothing, but this relation is nothing but the probability of being in category 1. And as this conditional probability I have to write repeatedly, so, it is written in a short form like p, x i. So, p, x i, is the probability that the data point x i, belongs to category 1.

(Refer Slide Time: 16:54)



$$\sigma(x) = \frac{1}{1+e^{-(a+bx)}}$$

$$0 \le \sigma(x) \le 1$$

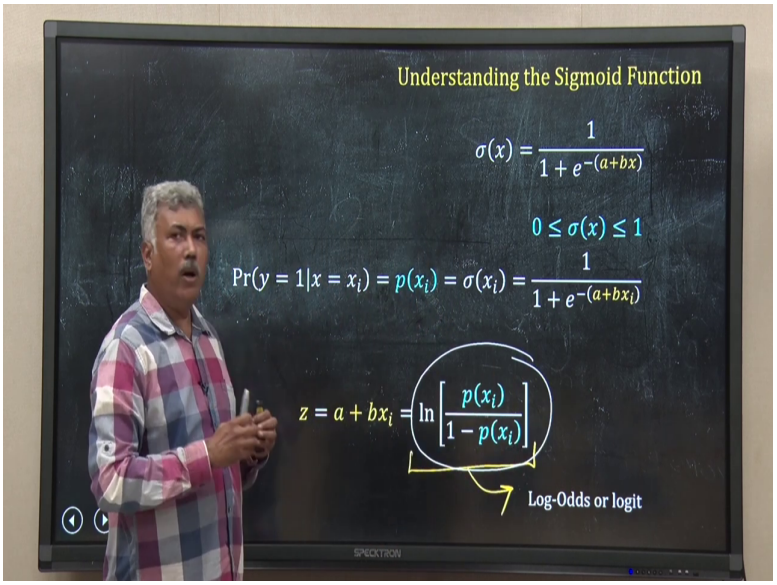
$$\Pr Pr\left(\frac{y=1}{x=x_i}\right) = p(x_i) = \sigma(x_i) = \frac{1}{1+e^{-(a+bx_i)}}$$

$$z = a + bx_i = \ln ln\left[\frac{p(x_i)}{1-p(x_i)}\right]$$

Now, let us rearrange this algebraic relationship. So, this is my relation p, x i, is equal to the sigmoid function with the exponential term. If you do some algebraic rearrangement, you will reach this a plus b, x i, is equal to l n of p x i, divided by 1 minus p x i. Now, what is p x i? p x i, is the probability of a data being in category 1, a person is in category 1, whereas 1 minus p x i, is the probability that the same data is belonging to category 0.

So, this is, these two terms in the ratio form written in the inside the square bracket is the ratio of two probabilities. And if you carefully look it, this is the probability of Yes, and this is the probability of No. If I ask whether the person belongs to the disease case or not. So, the upper one, this p x i, is the probability of yes and the denominator is probability of no.

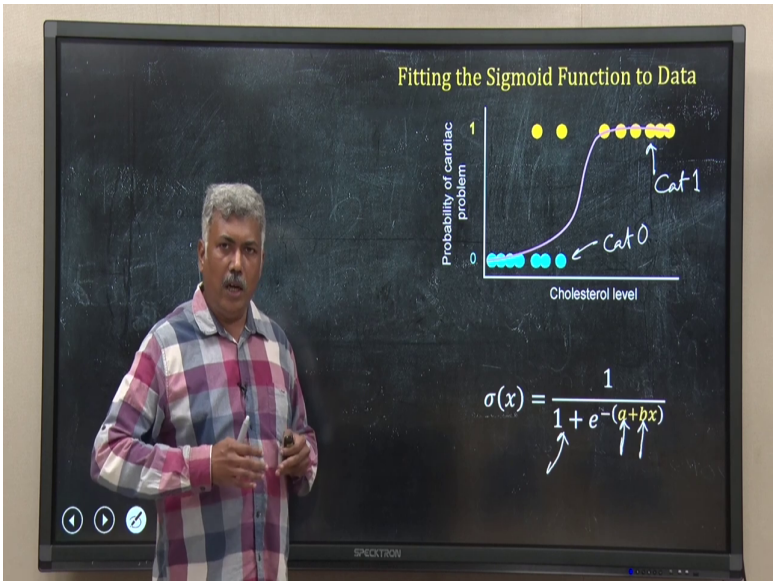
(Refer Slide Time: 18:07)



$$z = a + bx_i = \ln ln\left[\frac{p(x_i)}{1-p(x_i)}\right]$$

So, if I have in gambling, probability, the ratio between probability of winning, yes, and the probability of losing, no, that is called odds. So, in the square bracket, I have the odds of being in category 1. Now, I am taking the log of that. So, this whole thing we call log odds function or log it function. And that is why this regression using this function is called logistic regression.

(Refer Slide Time: 18:52)

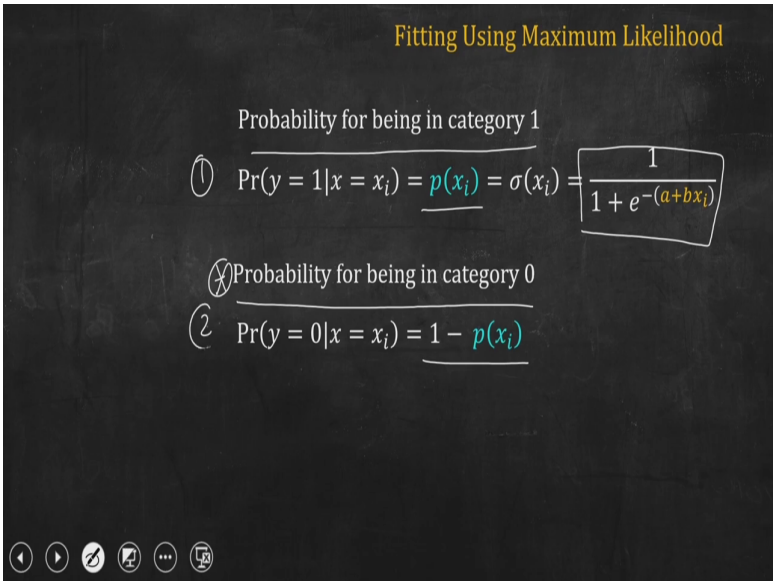


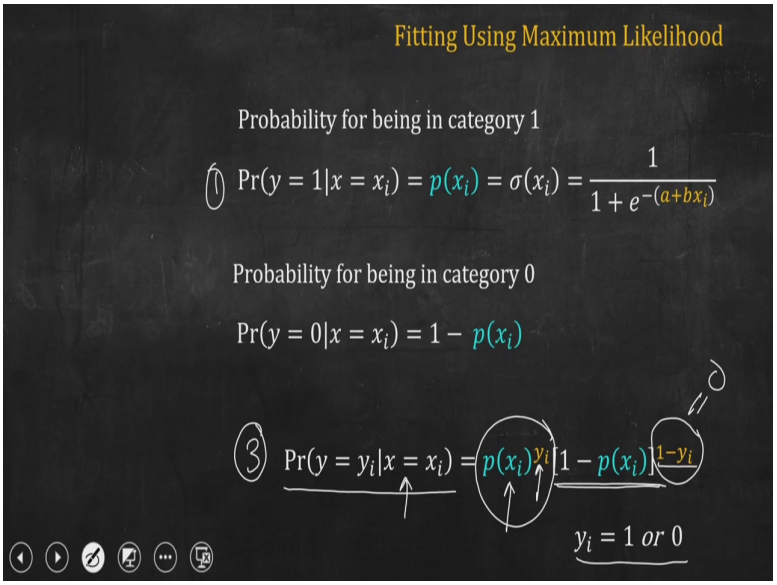
$$\sigma(x) = \frac{1}{1+e^{-(a+bx)}}$$

Now, let us go back to our problem in hand, we have to use this function to create a classifier, that means I have to fit this function to my data. What is my data, in the horizontal axis, I have cholesterol level, in the vertical axis, I have the probability of having cardiac problem. I have the label data, so I include these people, these yellow data points belong to category 1, and I know, these blue people or blue data point, belongs to category 0, so I have arranged them.

And now I want to fit that sigmoid function to this data, and I should get something like that pink curve. When I say I have to fit, that means I have to estimate the value of a and have to estimate the value of b. To achieve this to fit this model to the data, usually we use maximum likelihood approach. There are many methods based on maximum likelihood approach. And I will not go in detail of that, but I will briefly explain what is done in this case.

(Refer Slide Time: 19:48)





$$\Pr Pr\left(\frac{y=1}{x=x_i}\right) = p(x_i) = \sigma(x_i) = \frac{1}{1+e^{-(a+bx_i)}}$$

$$\Pr Pr\left(\frac{y=0}{x=x_i}\right) = 1 - p(x_i)$$

$$\Pr Pr\left(\frac{y=y_i}{x=x_i}\right) = p(x_i)^{y_i}\left[1 - p(x_i)\right]^{1-y_i}$$

$$y_i = 1 \text{ or } 0$$

Let us look into the probability terms first. p xi is the probability for being in category 1 and that is given by this sigmoid function. So, the probability for being in category 0, the other it

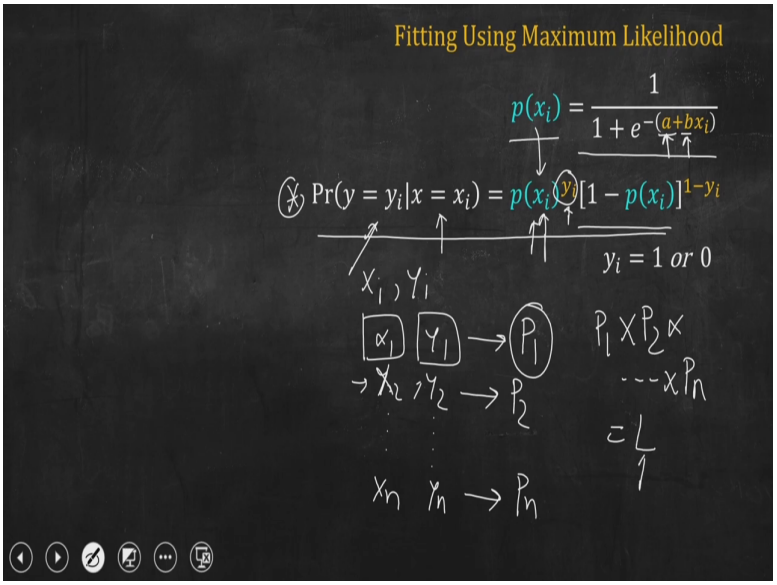
is binary case like for category 1 and category 0. So, the probability for being in category 0 is given by 1 minus p x i, because summation of these two probabilities must be equal to 1.

Now, I have two equations, equation 1 and equation 2, I can actually merge these two equations together and write in a integrated fashion, what is that? I can write the probability that y is equal to y i, where y i, is the categories 1 or 0, it can be either 1 or it can be 0, category 1 or category 0. So, the probability y equal to y i, given x equal to x i, x i is a particular value or for example, the cholesterol level of a person is equal to p x i to the power y i into 1 minus p x i to the power 1 minus y i. It may look complicated, but do not get confused simply consider y i equal to 1.

So, if I consider y i equal to 1 that means the sample belongs to category 1. So, y i equal to 1 means, 1 minus 1 this will become equal to 0, then this whole thing is something to the power 0, that means 1, then I will get only this part. So, I will get equation 1. So, when yi is equal to 1, this equation, the third equation gives me equation 1.

Similarly, why do not you put y i equal to 0, the second case, and you will find then this third equation will become same as equation 2. So, we have clubbed these two equations, equation 1 and equation 2, in one generalized form, and that helps us to calculate the likelihood and perform the maximum likelihood method. So, how do I formulate the likelihood? Let us look into that.

(Refer Slide Time: 22:05)



$$\sigma(x_i) = \frac{1}{1+e^{-(a+bx_i)}}$$

$$\Pr Pr\left(\frac{y=y_i}{x=x_i}\right) = p(x_i)^{y_i}\left[1 - p(x_i)\right]^{1-y_i}$$

$$y_i = 1 \; or \; 0$$

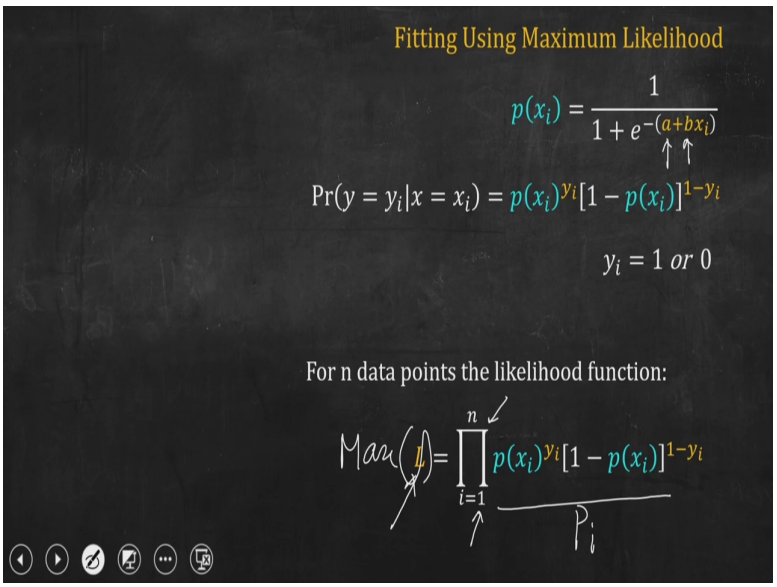
So, my probability of being in category 1, p x i is equal to the sigmoid function, I do not know a and b. What do I know? For some training data set if x i is known for a particular data, I know y i also. And I have this formula, this relationship, probability that y equal to y i, given x equal to x i, and is equal to probability of x i to the power y i into 1 minus p x i to the power 1 minus y i, I have this relation.

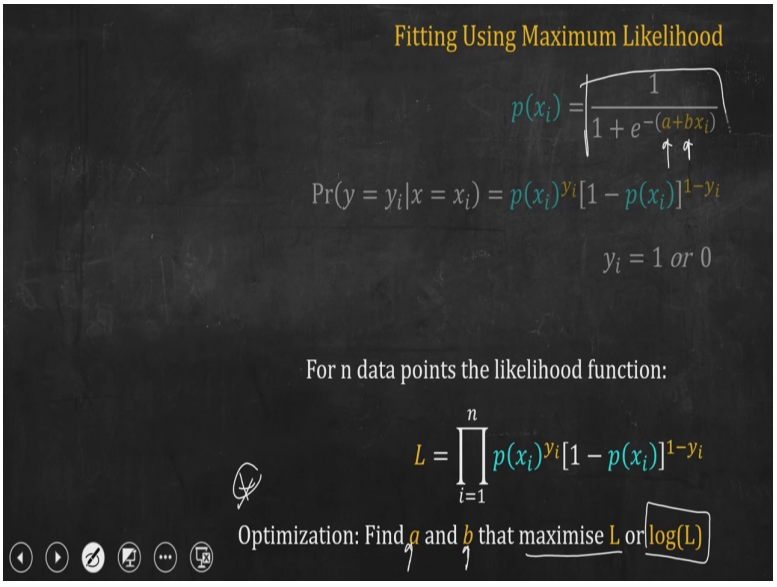
So now suppose I have n data points in my training data set. So, I have X 1 corresponding we have Y 1, X 2, correspondingly Y 2. So, X 1 is the cholesterol level of the person, and I know his or her category, so Y 1, that Y 1, will be either equal to 1 or equal to 0. So, in this way, I know X n, the cholesterol level of n-th person and his or her category, disease or not disease, 1 or 0. So, assume to start this algorithm, you can assume some particular value of a and b, consider some value.

Now, considering some value of a and b, using this X 1 and Y 1, value, you should be able to calculate the probability P 1, using this formula, you plug this one, put y i, equal to Y 1, x i, equal to X 1 and use this equation, you will get the probability, we call it P 1. Similarly, for the second data point in my training data set, using the same value of a and b, I can calculate the probability, so that will be P 2. Again, I am using the same equation, only the value of X and Y has changed.

In this way, I keep on calculating the probability for all the n data. So, for the n-th data, I have P n. So, each of these data in my training data set has a associated probability P 1, P 2, up to P n. So, what is the total probability? If I consider that these data points are independent, then the total probability will be equal to P 1 into P 2 and you multiply all of them up to P n. This total probability is the likelihood of our model. We write it as L.

(Refer Slide Time: 25:03)





$$L = \prod_{i=1}^{n} p\left(x_i\right)^{y_i}\left[1 - p\left(x_i\right)\right]^{1-y_i}$$

So, if I write it using mathematical notation, what do I get? I get the likelihood, which is called the likelihood function is nothing but the multiplication of this one for n data point. Remember, this is nothing but P, P i, So, I have n data points, so I have n P's and I am multiplying this, this symbol is a representation of multiplication from i equal to 1 to n. So, this is my likelihood function.

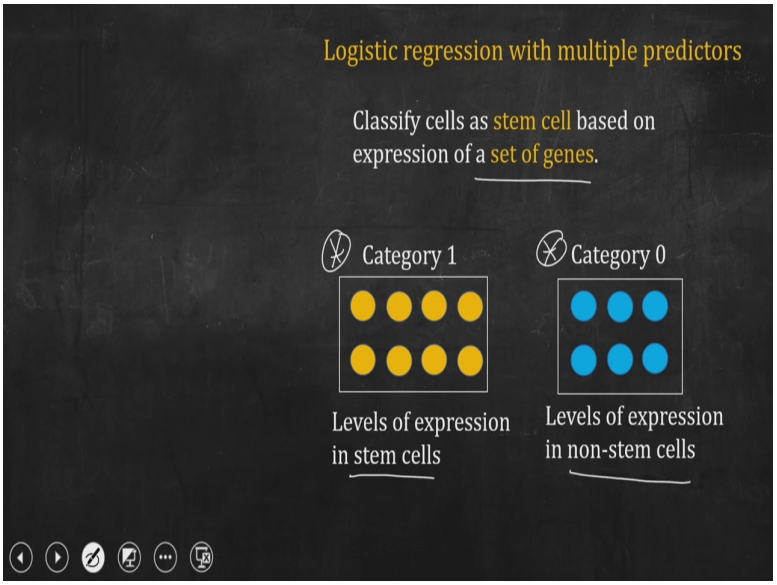
Now, I have calculated this likelihood value using an assumed value of a and b. Now, what you can do? You can change the value of a and b, and again calculate the same likelihood, new result will come, and if a new result, new likelihood is better, means bigger than the

previous one, that means, the probability is higher. So, I will discard the previous value of a and b, I will take the new value of a and b.

So, in this way using some algorithm, you have to find out the value of a and b which will maximize my L. So, that is an optimization problem. So, what we are doing, we are finding a and b that will maximize L, the likelihood function. In general, when we have the maximum of L, the log of that L is also maximum, but doing the calculation using log of L is much easier that is why your optimizing algorithm will maximize log of L.

And there are many algorithms to do that, gradient ascent, gradient descent type algorithms can be used to maximize log of the likelihood function, so, that you get the optimum value of a and b and once you have got it, you have got your model. The model is nothing but this function, this is your classifier model.

(Refer Slide Time: 27:12)

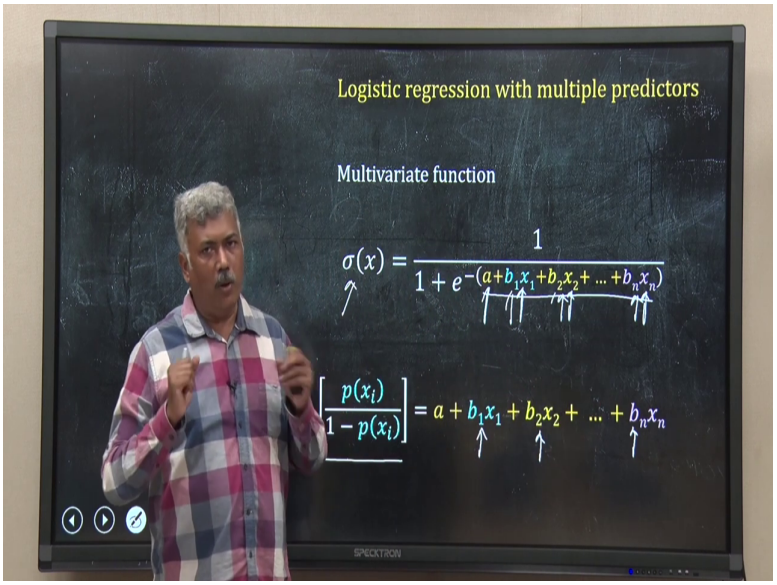


So, what we have discussed till now, we have taken a binary problem, binary classification problem, where I have two categories, category 1 and category 0. Category 1, belongs to people who have the cardiac problem, whereas category 0 belongs to people who does not have the disease. So, it is the binary case. And what we have done? We have only one predictor, the predictor is the cholesterol level.

Now, imagine in most of the real-life cases, actually, when you will do classification, the predictor will not be one there can be more than one predictor. For example, you can imagine the stem cell case. So, when you classify cell as stem cells, usually you will have a set of genes, which we will call markers for stem cell, not a unique one gene. So, you may have five different genes and their expression level decides whether that particular cell is stem cell or not. So, it is still a binary problem, but we have more number of predictors.

So, again, our training data set will be labelled in two categories, category 1, and category 0. And in category 1, we will put stem cells and in category 0 we put non stem cell, but now we have more than one predictor variable. But in the sigmoid function that I have used, I have only one predictor variable x. So, how should I accommodate other predictors in my regression model, so, that can be done very easily.

(Refer Slide Time: 28:42)



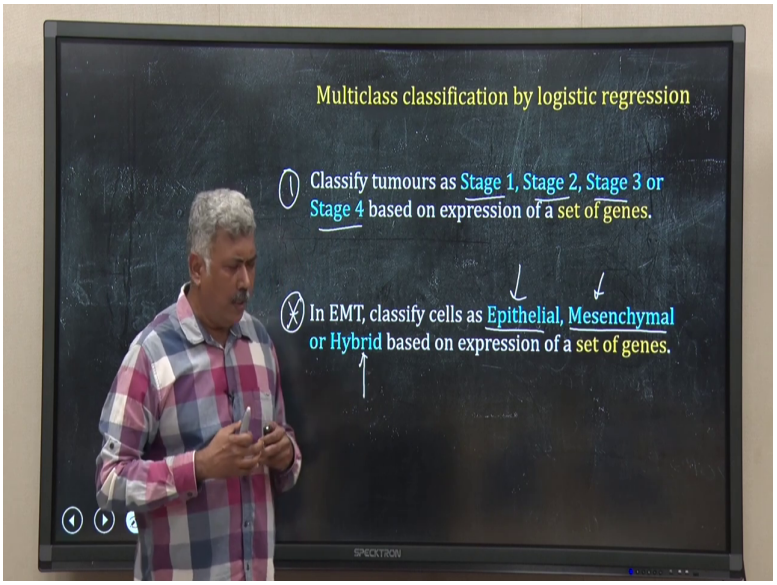
$$\sigma(x) = \frac{1}{1+e^{-(a+b_1x_1+b_2x_2+\cdots+b_nx_n)}}$$

$$\ln ln\left[\frac{p(x_i)}{1-p(x_i)}\right] = a + b_1x_1 + b_2x_2 + \cdots + b_nx_n$$

Actually, I can write a multivariate sigmoid. So, I will write a multivariate function where in place of x, I will have a plus b 1, x 1 plus b 2, x 2, up to b n, x n. So, suppose I have five genes, which are used to predict whether a cell belongs to stem cell class or non-stem cell class. So, this is my first gene, second gene, in this way upto the fifth gene. So, if you rearrange the whole thing, you can actually see this log odds ratio is equal to a plus b 1, x 1 plus and b 2, x 2, up to b n, x n.

Now, the rest of the algorithm is same, you will use the maximum likelihood-based method to fit this equation to your data, you have to use still the binary categorization 1 and 0 and that fitting that maximum likelihood-based model fitting will give you the value of a, b 1, b 2 and b n, and you will get your classifier model. So, I have dealt with now with where we have multiple predictors.

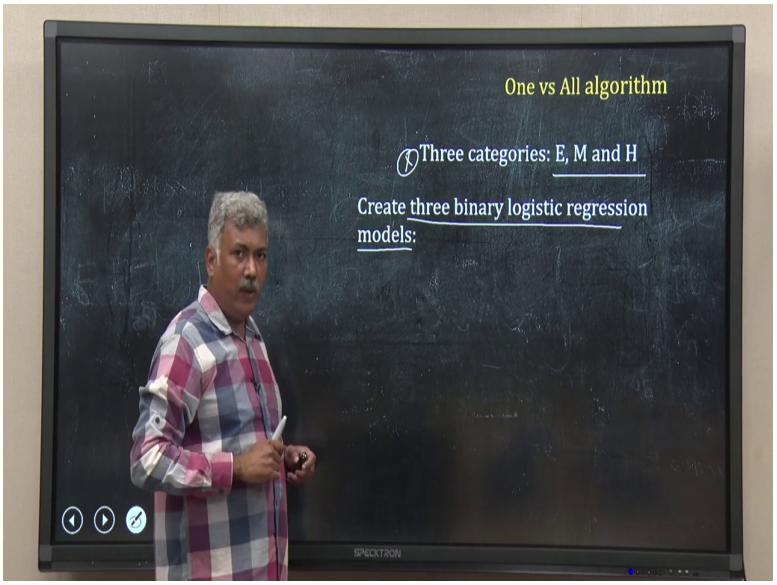
(Refer Slide Time: 29:57)



Now, if we have a problem where we have multiple categories more than two categories, that means we are moving from binary to more than two. Take an example, for example, if you are classifying tumours, that is a first example here. If you want to classify tumour, the tumour can be classification classified in stage 1, stage 2, stage 3, stage 4, and so on, different stages, more than two stages. So, I do not have any more binary problem, it has four classes.

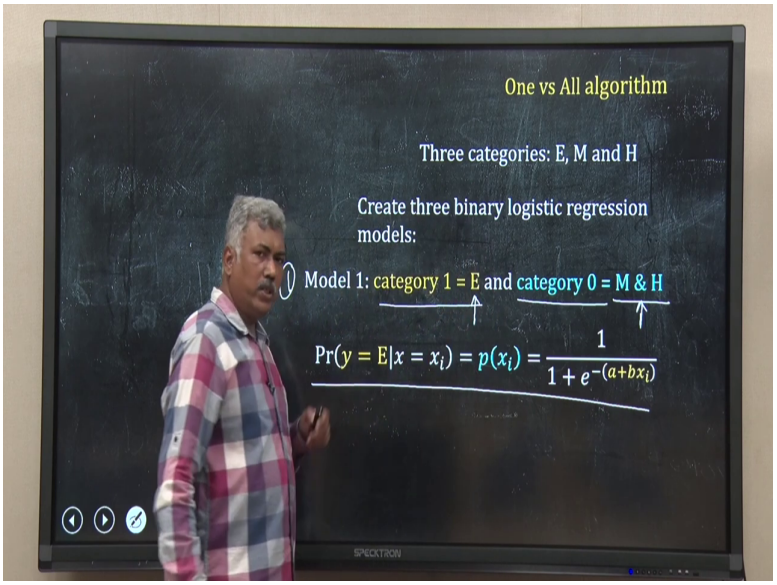
Whereas, suppose take the example of EMT, epithelial to mesenchymal transition. In epithelial to mesenchymal transition what happens, that epithelial cell becomes mesenchymal, but at the same time some cells are there which are hybrid between epithelial and mesenchymal. So, if I have to classify cell during EMT, I should have three classes epithelial, mesenchymal and hybrid. So, it is no more a binary classification problem. So, how should I handle this, there are many algorithms to do that.

(Refer Slide Time: 30:59)



I will discuss one, one particular algorithm that is called one versus all or one versus rest algorithm. So, what I have? Take the EMT case, I have three categories, I mark them as E, M and H. Now, I have three classes now, what I do? I create three binary logistic regression model, not one, but three. I will explain what these three are.

(Refer Slide Time: 31:23)



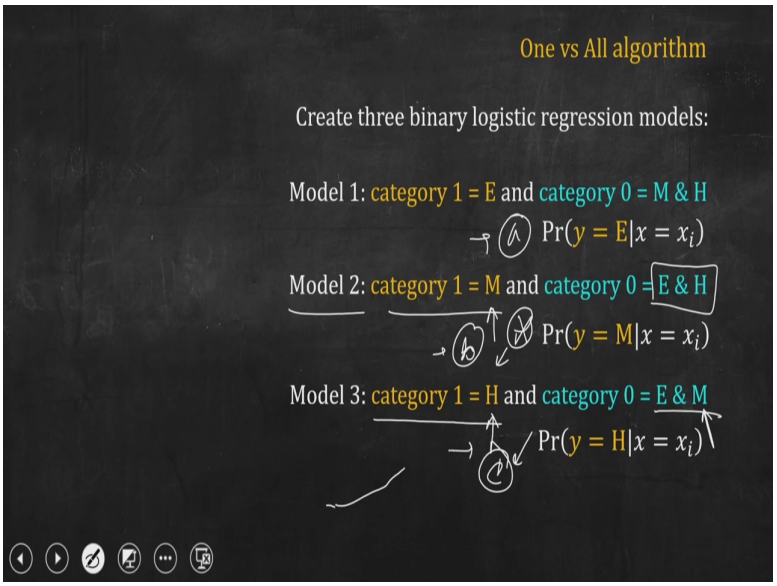
$$\Pr Pr\left(\frac{y=E}{x=x_i}\right) = p(x_i) = \frac{1}{1+e^{-(a+bx_i)}}$$

Let that the first model. In the model one what I do, category 1 is the cells, which are epithelial type, that means all cells which are epithelial are considered in category 1, rest of the cells that mean cells, which are either M or H type, mesenchymal or hybrid type, are

considered into categories 0. What I am doing? I have three class classification problem, but I am converted into two class binary case.

So, I have taken category 1, where I have all the E type cells epithelial cell and the rest of the cell are put in the bin of category 0, non-epithelial cells. Now, if you have this binary problem, now, you create a binary classifier the way we have done just now. So, you create a binary classifier using that sigmoid function of logistic regression and you do maximum likelihood-based method and create the classifier model.

(Refer Slide Time: 32:22)



$$Model\ 1: \Pr Pr\left(\frac{y=E}{x=x_i}\right)$$

$$Model\ 2: \Pr Pr\left(\frac{y=M}{x=x_i}\right)$$

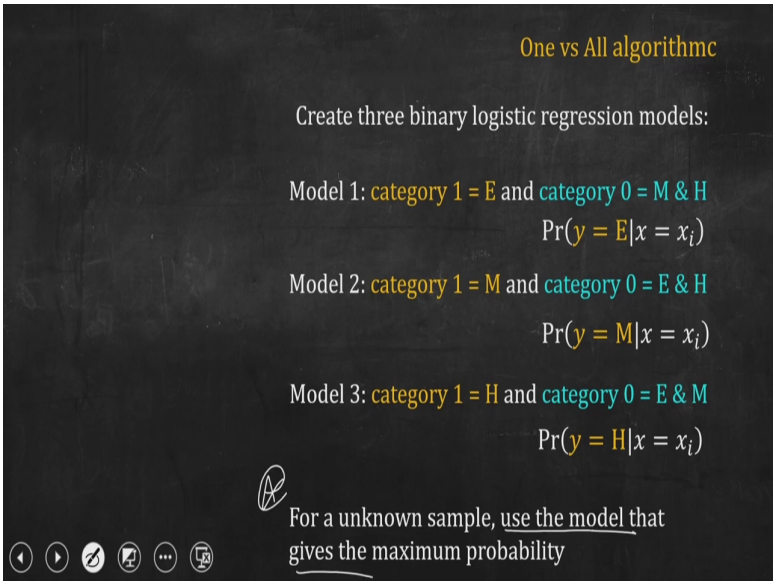
$$Model\ 3: \Pr Pr\left(\frac{y=H}{x=x_i}\right)$$

Now, create the second model. What is the second model? In second model in category 1, you put M, the mesenchymal cells, whereas, in category 0, you put rest of the cell E and H type. Again, you create a classifier model. So, you get this classifier model. For the third model, obviously, category 1 will have the hybrid type cell and category 0 will have epithelial and mesenchymal cells. So, one and rest of the type.

So, I have converted this three-class problem into three binary classifier problem. And for each of cases, I have developed the classifier using binary classification using the logistic

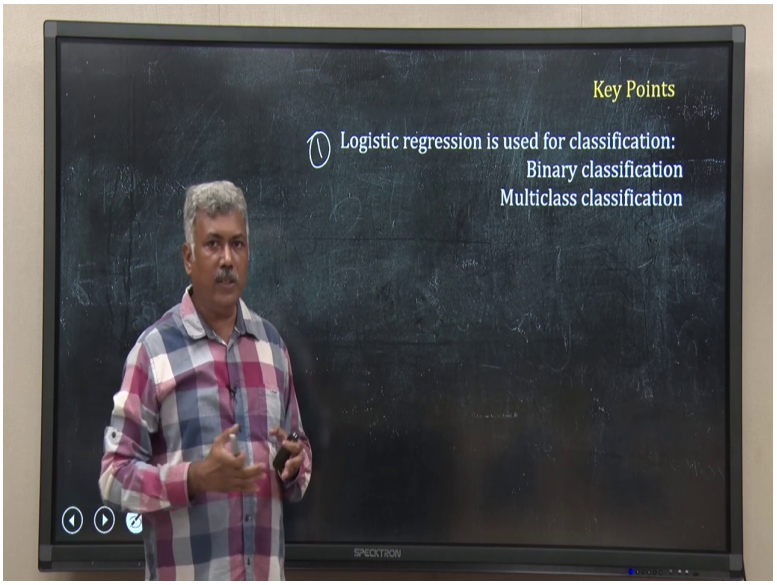
regression. So, I have three probability-based model. Now, you have a new cell, you do not know whether it is epithelial, mesenchymal or hybrid, you know the expression of certain marker genes in that. So, using the level of expression of those marker gene, you use this three-classifier model A, B and C and calculate the probabilities and you take the model which gives you the maximum probability.

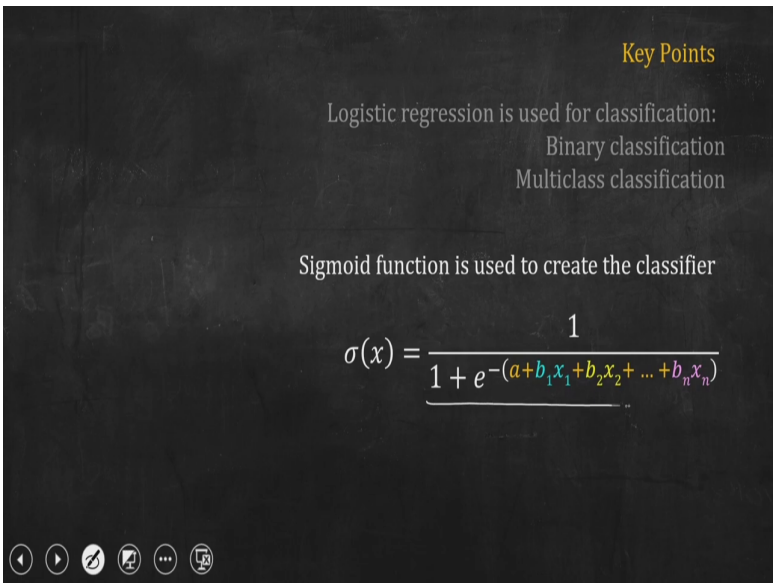
(Refer Slide Time: 33:40)

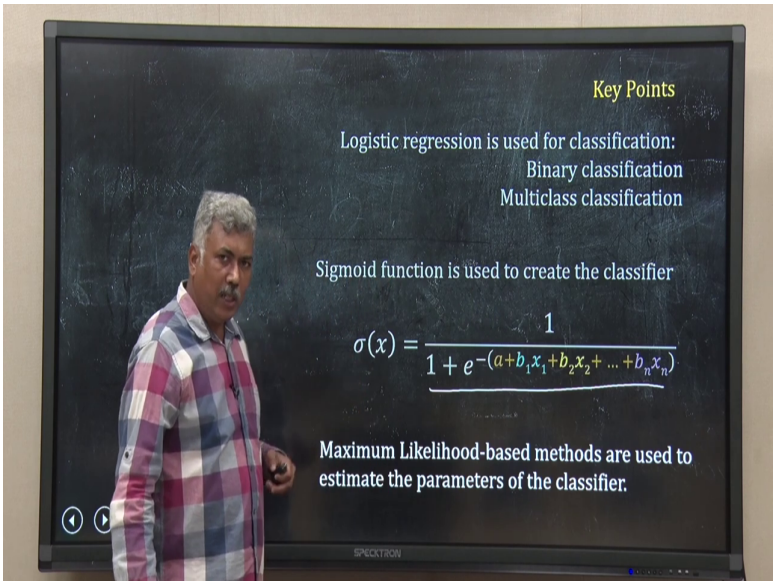


So, for a unknown sample, use the model that gives the maximum probability. So, in this way, one versus all algorithm or one versus rest algorithm, convert a multi class problem into multiple binary problem and then compare their probabilities.

(Refer Slide Time: 34:05)







$$\sigma(x) = \frac{1}{1+e^{-(a+b_1x_1+b_2x_2+\cdots+b_nx_n)}}$$

Let me jot down what we have learned in this lecture. The first thing that we have learned is that a logistic regression is used for classification, both binary classification as well as multi class classification. For logistic regression, we use a logistic function which is a sigmoid and that is given here, I have written the generalized one with multiple variable, 1 divided by 1 plus e to the power minus a plus b 1 x, plus b 2 x, up to b n, x n.

And we use a maximum likelihood-based approach to create this classifier model that means to calculate the parameters of this model, a, b 1, b 2 and b n. That is all for this lecture. See you in the next lecture. Till then happy learning.