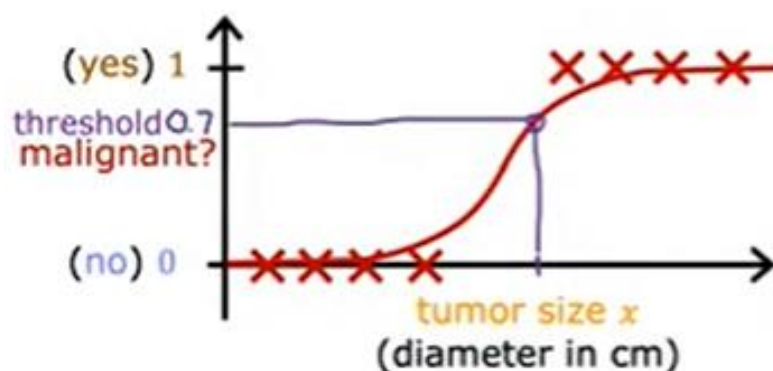# Régression logistique

Let's talk about logistic regression, which is probably the single most widely used classification algorithm in the world. This is something that I use all the time in my work. Let's continue with the example of classifying whether a tumor is malignant. Whereas before we're going to use the label 1 or yes to the positive class to represent malignant tumors, and zero or no and negative examples to represent benign tumors.

Here's a graph of the dataset where the horizontal axis is the tumor size and the vertical axis takes on only values of 0 and 1, because is a classification problem. You saw in the last video that linear regression is not a good algorithm for this problem. In contrast, what logistic regression we end up doing is fit a curve that looks like this, S-shaped curve to this dataset.

For this example, if a patient comes in with a tumor of this size, which I'm showing on the x-axis, then the algorithm will output 0.7 suggesting that is closer or maybe more likely to be malignant and benign. Will say more later what 0.7 actually means in this context. But the output label y is never 0.7 is only ever 0 or 1.
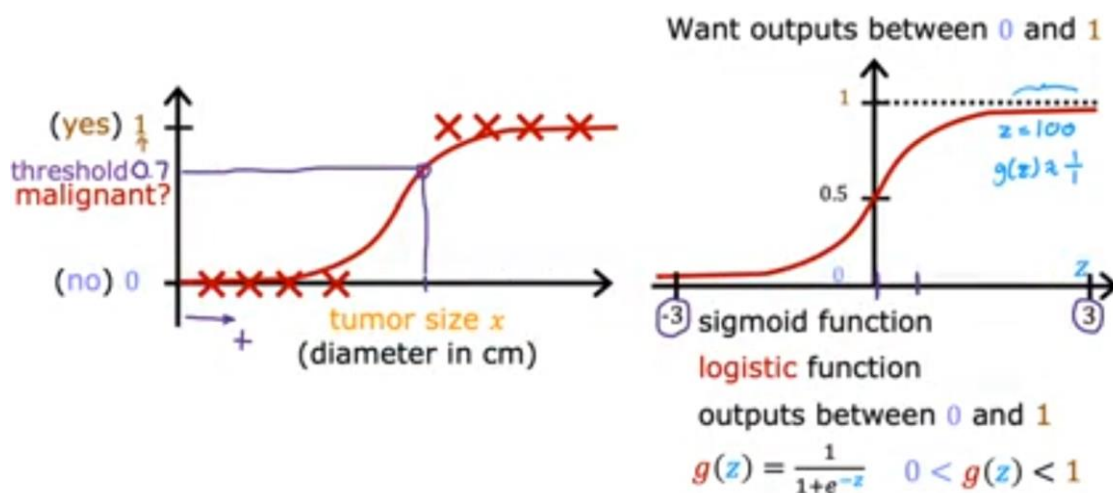

To build out to the logistic regression algorithm, there's an important mathematical function I like to describe which is called the Sigmoid function, sometimes also referred to as the logistic function. The Sigmoid function looks like this. Notice that the x-axis of the graph on the left and right are different. In the graph to the left on the x-axis is the tumor size, so is all positive numbers.
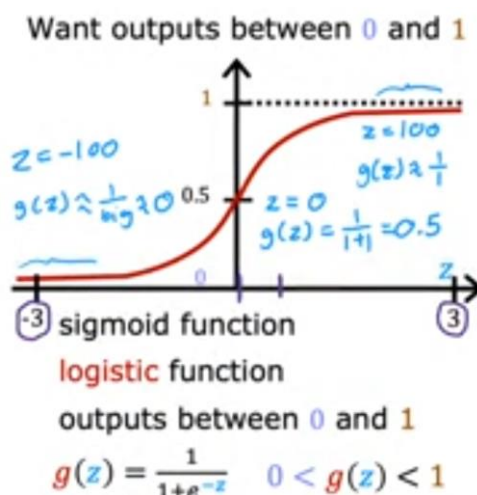
Whereas in the graph on the right, you have 0 down here, and the horizontal axis takes on both negative and positive values and have label the horizontal axis Z. I'm showing here just a range of negative 3 to plus 3. So the Sigmoid function outputs value is between 0 and 1. If I use g of z to denote this function, then the formula of g of z is equal to 1 over 1 plus e to the negative z.

Where here e is a mathematical constant that takes on a value of about 2.7, and so e to the negative z is that mathematical constant to the power of negative z. Notice if z where really be, say a 100, e to the negative z is e to the negative 100 which is a tiny number.

So this ends up being 1 over 1 plus a tiny little number, and so the denominator will be basically very close to 1. Which is why when z is large, g of z that is a Sigmoid function of z is going to be very close to 1.

Want outputs between 0 and 1

sigmoid function
logistic function
outputs between 0 and 1

$$g(z) = \frac{1}{1+e^{-z}} \qquad 0 < g(z) < 1$$

Conversely, you can also check for yourself that when z is a very large negative number, then g of z becomes 1 over a giant number, which is why g of z is very close to 0. That's why the sigmoid function has this shape where it starts very close to zero and slowly builds up or grows to the value of one. Also, in the Sigmoid function when z is equal to 0, then e to the negative z is e to the negative 0 which is equal to 1, and so g of z is equal to 1 over 1 plus 1 which is 0.5, so that's why it passes the vertical axis at 0.5.

Want outputs between 0 and 1

sigmoid function
logistic function
outputs between 0 and 1

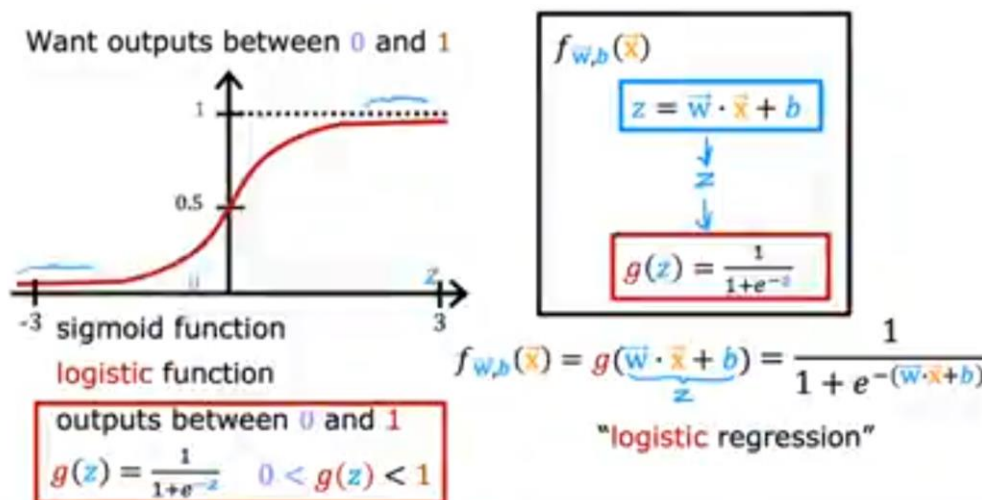$$g(z) = \frac{1}{1+e^{-z}} \qquad 0 < g(z) < 1$$

Now, let's use this to build up to the logistic regression algorithm. We're going to do this in two steps. In the first step, I hope you remember that a straight line function, like a linear regression function can be defined as w. product of x plus b. Let's store this value in a variable which I'm going to call z, and this will turn out to be the same z as the one you saw on the previous slide, but we'll get to that in a minute.

The next step then is to take this value of z and pass it to the Sigmoid function, also called the logistic function, g. Now, g of z then outputs a value computed by this formula, 1 over 1 plus e to the negative z. There's going to be between 0 and 1. When you take these two equations and put them together, they then give you the logistic regression model f of x, which is equal to g of wx plus b.

Or equivalently g of z, which is equal to this formula over here. This is the logistic regression model, and what it does is it inputs feature or set of features X and outputs a number between 0 and 1.

Next, let's take a look at how to interpret the output of logistic regression.of logistic regression.

Want outputs between 0 and 1

$$z = \vec{w} \cdot \vec{x} + b$$

$$g(z) = \frac{1}{1+e^{-z}}$$

sigmoid function
logistic function

outputs between 0 and 1

$$g(z) = \frac{1}{1+e^{-z}} \quad 0 < g(z) < 1$$

$$f_{\vec{w},b}(\vec{x}) = g(\underbrace{\vec{w} \cdot \vec{x} + b}_{z}) = \frac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}}$$

"logistic regression"

We'll return to the tumor classification example. The way I encourage you to think of logistic regressions output is to think of it as outputting the probability that the class or the label y will be equal to 1 given a certain input x.

For example, in this application, where x is the tumor size and y is either 0 or 1, if you have a patient come in and she has a tumor of a certain size x, and if based on this input x, the model I'll plus 0.7, then what that means is that the model is predicting or the model thinks there's a 70 percent chance that the true label y would be equal to 1 for this patient.

In other words, the model is telling us that it thinks the patient has a 70 percent chance of the tumor turning out to be malignant. Now, let me ask you a question. See if you can get this right. We know that y has to be either 0 or 1, so if y has a 70 percent chance of being 1, what is the chance that it is 0?

So y has got to be either 0 or 1, and thus the probability of it being 0 or 1 these two numbers have to add up to one or to a 100 percent chance. That's why if the chance of y being 1 is 0.7 or 70 percent chance, then the chance of it being 0 has got to be 0.3 or 30 percent chance.

Interpretation of logistic regression output

$$f_{\vec{w},b}(\vec{x}) = \frac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}}$$

"probability" that class is 1

Example:
x is "tumor size"
y is 0 (not malignant)
or 1 (malignant)

$$P(y = 0) + P(y = 1) = 1$$

$$f_{\vec{w},b}(\vec{x}) = 0.7$$
70% chance that y is 1

If someday you read research papers or blog pulls of all logistic regression, sometimes you see this notation that f of x is equal to p of y equals 1 given the input features x and with parameters w and b. What the semicolon here is used to denote is just that w and b are parameters that affect this computation of what is the probability of y being equal to 1 given the input feature x? For the purpose of this class, don't worry too much about what this vertical line and what the semicolon mean. You

don't need to remember or follow any of this mathematical notation for this class. I'm mentioning this only because you may see this in other places.

Interpretation of logistic regression output

$$f_{\vec{w},b}(\vec{x}) = \frac{1}{1 + e^{-(\vec{w}\cdot\vec{x}+b)}}$$

"probability" that class is 1

Example:
x is "tumor size"
y is 0 (not malignant)
or 1 (malignant)

$$f_{\vec{w},b}(\vec{x}) = 0.7$$
70% chance that y is 1

$$\overset{P(1)}{f_{\vec{w},b}(\vec{x}) = P(y = 1|\vec{x}; \vec{w},b)}$$

Probability that y is 1,
given input x, parameters $\vec{w}, b$

$$P(y = 0) + P(y = 1) = 1$$