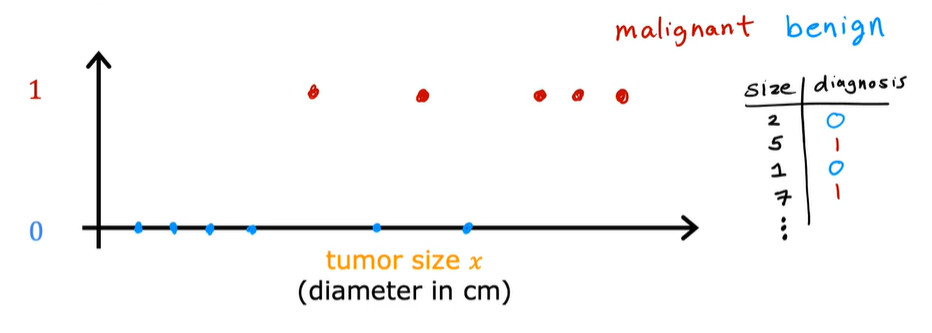
So, Supervised Learning algorithms learn to predict input, output or X to Y mapping. And, in the last lecture you saw that regression algorithms, which is a type of Supervised Learning algorithm learns to predict numbers out of infinitely possible numbers.

There’s a second major type of Supervised Learning algorithm called a Classification algorithm. Take breast cancer detection as an example of a classification problem. Say you’re building a machine learning system so that doctors can have a diagnostic tool to detect breast cancer. This is important because early detection could potentially save a patient’s life.

Using a patient’s medical records your machine learning system tries to figure out if a tumor that is a lump is malignant meaning cancerous or dangerous. Or if that tumor, that lump is benign, meaning that it’s just a lump that isn’t cancerous and isn’t that dangerous?

Maybe your dataset has tumors of various sizes and these tumors are labeled as either benign, which I will designate with a 0 or malignant, which I will designate with a 1. You can then plot your data on a graph like below, where the horizontal axis represents the size of the tumor and the vertical axis takes on only two values 0 or 1, depending on whether the tumor is benign 0 or malignant 1.



One reason that this is different from regression is that we’re trying to predict only a small number of possible outputs or categories. In this case, two possible outputs 0 or 1, benign or malignant. This is different from regression which tries to predict any number, all of the infinitely many number of possible numbers.

And so the fact that there are only two possible outputs is what makes this classification. Because there are only two possible outputs or two possible categories in this example, you can also plot this data set on a line. In this we can use two different symbols to denote the category using a circle an O to denote the benign examples and a cross to denote the malignant examples.

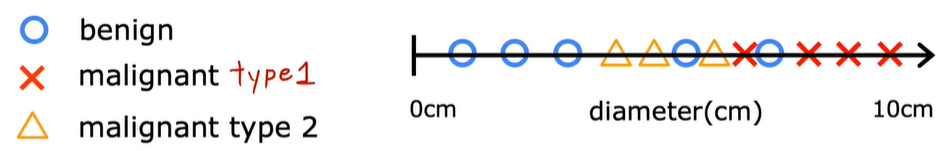


And if a new patient walks in for a diagnosis and they have a lump that is the size as shown below, then the question is will your system classify this tumor as benign or malignant?

Diagram

Description automatically generated

It turns out that in classification problems you can also have more than two possible output categories. Maybe your learning algorithm can output multiple types of cancer diagnosis if it turns out to be malignant. So, let’s call two different types of cancer type 1 and type 2. In this case the average would have three possible output categories it could predict.



In classification, the terms output classes and output categories are often used interchangeably.

**To summarize:**

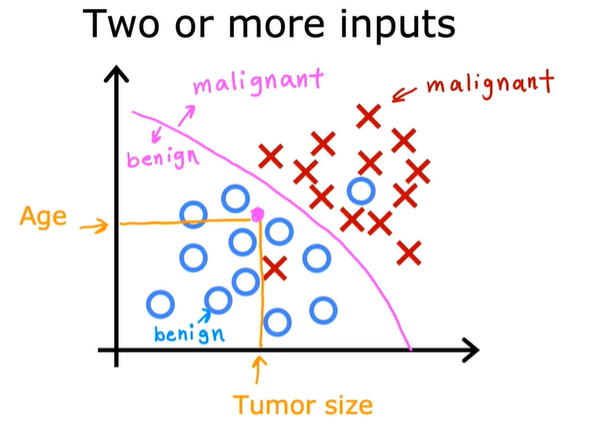
* Classification algorithms predict categories.
* Categories don’t have to be numbers. It could be non-numeric for example, it can predict whether a picture is of a cat or of a dog.
* Categories can also be numbers like 0, 1 or 2.

But what makes classification different from regression when you’re interpreting the numbers is that classification predicts a small finite limited set of possible output categories such as 0, 1 and 2 but not all possible numbers between like 0.5 or 1.7.

In the example of supervised learning that we’ve been looking at, we had only one input value the size of the tumor. But you can also use more than one input value to predict an output. Here’s an example, instead of just knowing the tumor size, say you also have each patient’s age in years. Your new dataset now has two inputs, age, and tumor size.

What in this new dataset we’re going to use circles to show patients whose tumors are benign and crosses to show the patients with a tumor that was malignant. So, when a new patient comes in, the doctor can measure the patient’s tumor size and also record the patient’s age. And so, given this, how can we predict if this patient’s tumor is benign or malignant?

Given the day like this, what the learning algorithm might do is find some boundary that separates out the malignant tumors from the benign ones. So, the learning algorithm has to decide how to fit a boundary line through this data. The boundary line found by the learning algorithm would help the doctor with the diagnosis. In this case the tumor is more likely to be benign.



From this example, we have seen how two inputs the patient’s age and tumor size can be used. In other Machine Learning problems often many more input values are required like the thickness of the tumor clump, uniformity of the cell size, uniformity of the cell shape and so on.

**To recap:**

* Supervised Learning maps input X to output Y, where the learning algorithm learns from the quote right answers.
* The two major types of Supervised Learning are Regression and Classification.
* In a regression application like predicting prices of houses, the learning algorithm has to predict numbers from infinitely many possible output numbers.
* Whereas in classification the learning algorithm has to make a prediction of a category, all of small set of possible outputs.

