MACHINE LEARNING

CLASSIFICATION

What exactly classification is a process which predicts a class where the data points belongs to.
Here class can be a target or a label. This comes under supervised learning

Here some examples for classification problems :

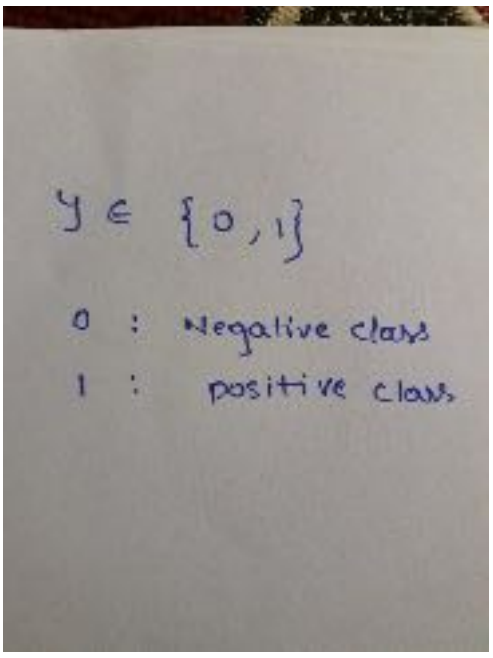
~ Online transaction : Fraudulent (yes/no) ?

So if you have a website that sells stuff and if you want to know if a particular transaction is fraudulent or not, whether someone is using a stolen credit card or has stolen the user's password.

~ Tumor : Malignant / benign ?

Here the another example for classification problem to classify whether the tumor is malignant or benign.

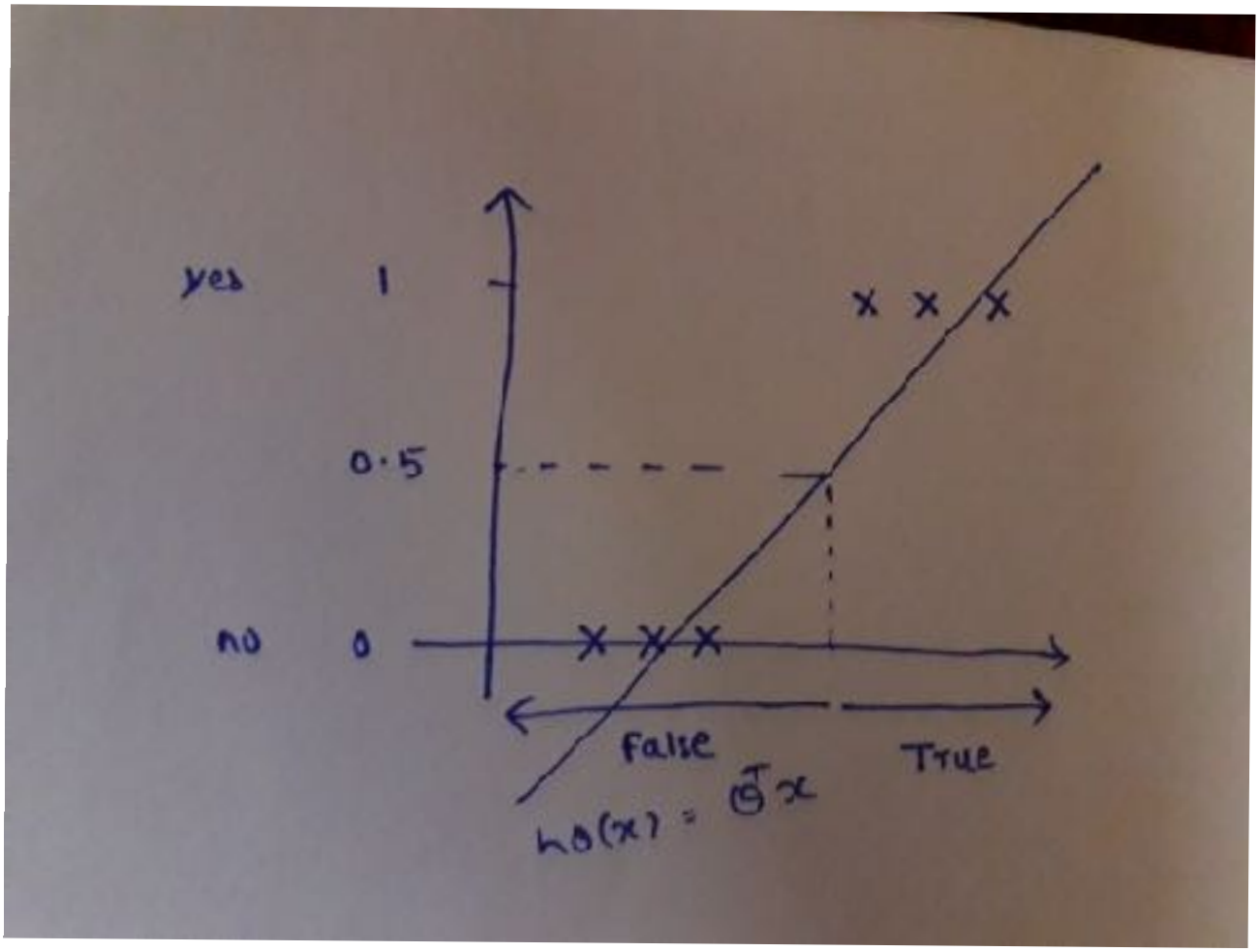
In all of these problems the variable that we're trying to predict is a variable y that we can think of as taking on two values either zero or one, either fraudulent or not fraudulent, related malignant or benign.



Another name for the class that we denote with zero is the negative class, and another name for the class that we denote with one is the positive class. So zero we denote as the benign tumor, and one, positive class we denote a malignant tumor.

So how do we develop a classification algorithm?

Let us suppose a training set

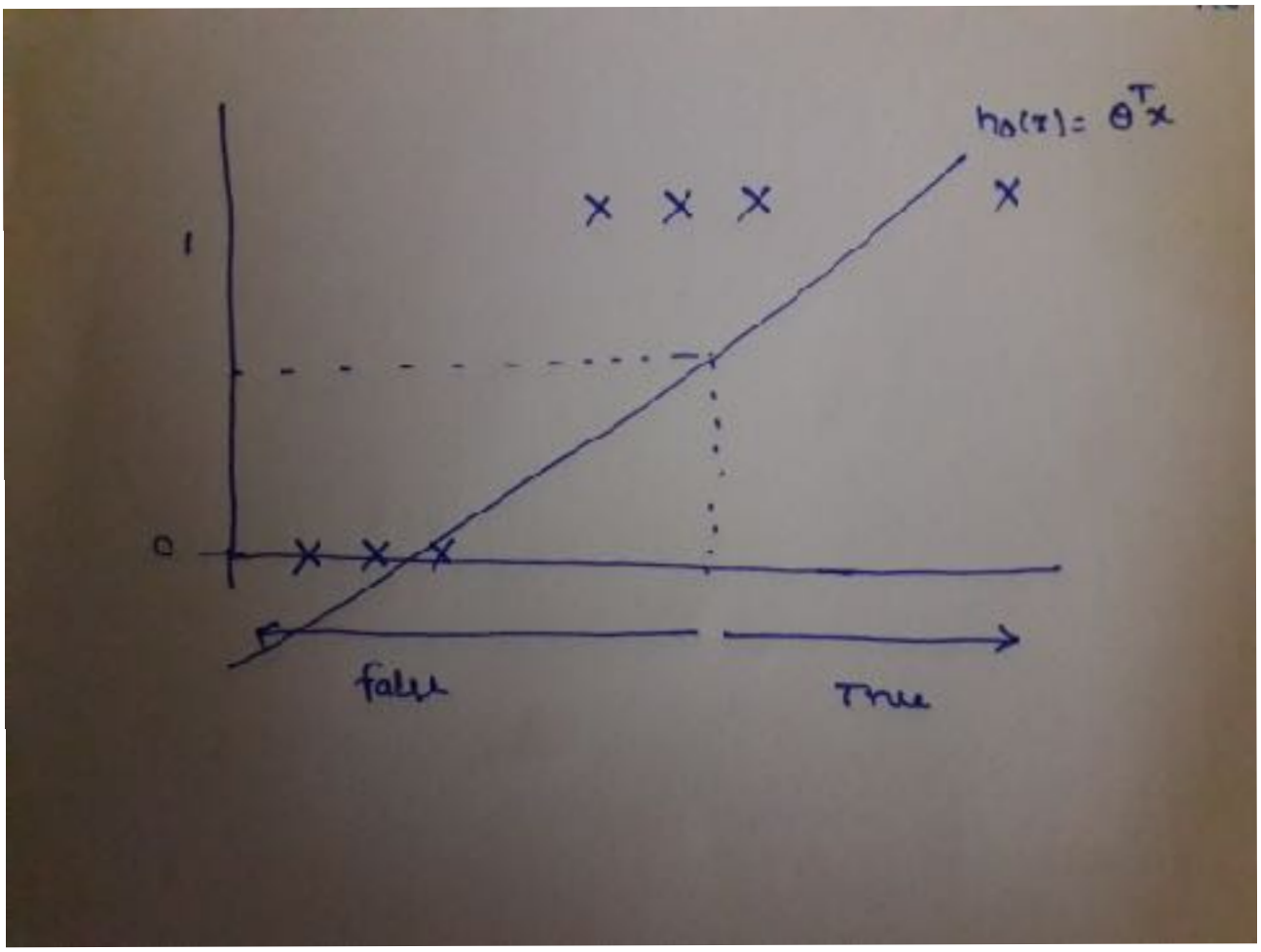


To this training set if we apply the algorithm we know that is linear regression produce a line something this.
We can predict using threshold value 0.5.
if h(x)>=0.5 is consider as “1”

If h(x)<0.5 is consider as “0”
To that point the left would be false (i.e 0) and the right would be true(i.e 1)

So In this particular example, it looks like linear regression is actually doing something reasonable. Even though this is a classification toss we're interested in. But now let's try changing the problem

Consider a new data is added to training set and apply the same liner regression to the updated training set, thus produce a line something like this



Now if we observe it carefully the point has changed and if we consider left to the point is false and right to it is true its not so

good liner regression is not giving the exact we are looking for if we add a new data.