

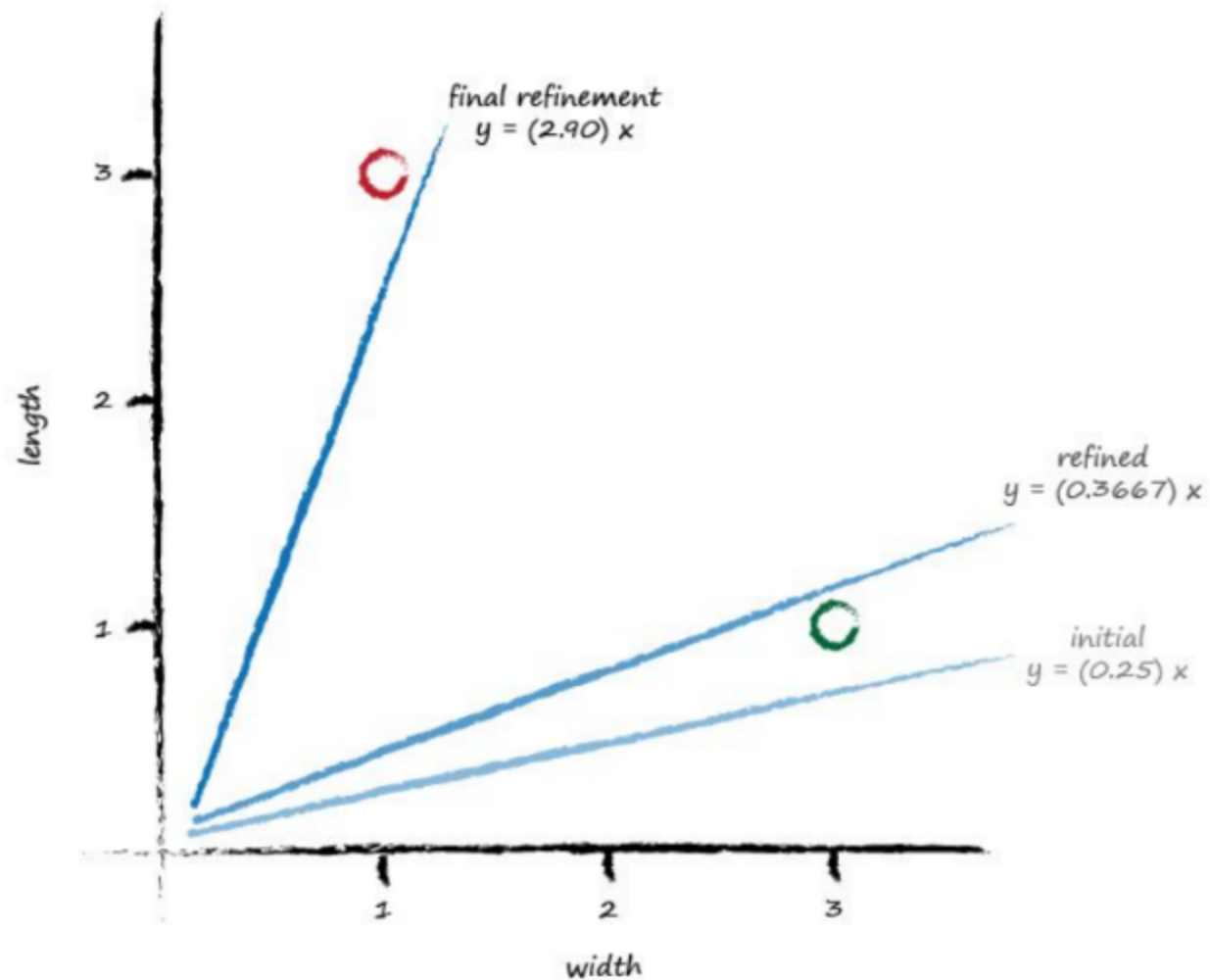
Learning Rate

lets take the $x = 3$ and $y = 1$ and then $x = 1$ and $y = 3$ and plot according to the equation

$$\Delta A = E / x$$

what is happening here?

Figure 2.1



What this is not what we want? Each new observation the line is jumping directly to the desired value for that observation. This is what you would expect from our implementation of our previous equation

$$\Delta A = E / x$$

What we need is some moderation with our updates. This is the beginning of one of the core idea of machine learning the learning rate. We use it to slowly nudge our updates in the right direction while not losing the all the data from the previous observation.

$$\Delta A = L \times (E / x)$$

Real training examples can contain noise and errors, it helps us lessen the effect they have our training. Now we will have a classifier that takes in account all the observations and classifies ladybirds and caterpillars accordingly.

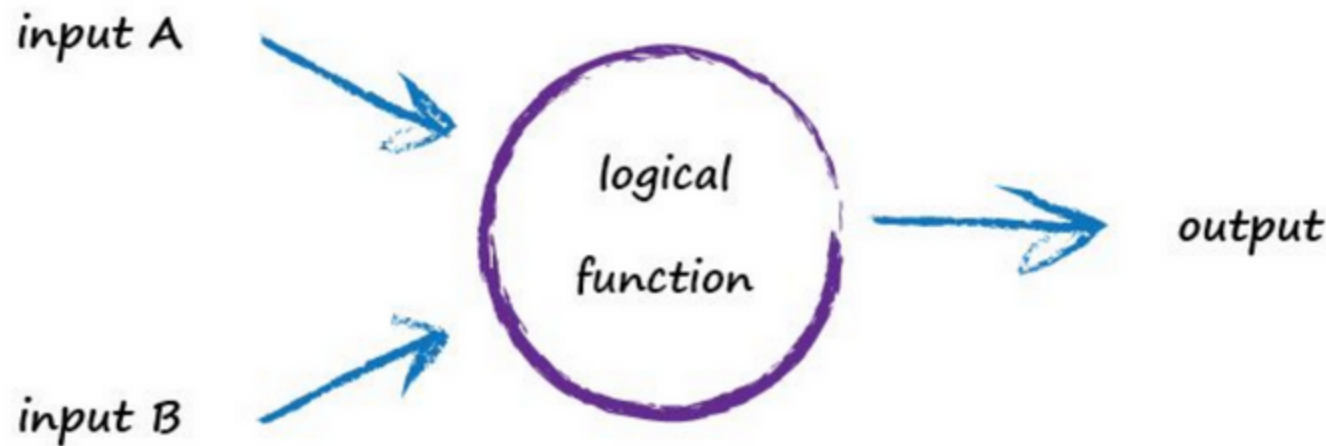
Conclusion:

- Doing each new update naively makes us discard the previous observation data.
- Training examples can contain noise and errors. Moderation lessens their impact on our data.

Sometimes One Classifier is not enough

Using the basic predictor as a classifier is all good and all but what happens when the data is not goverend by a single linear function. Lets look towards Boolean function to explain such a case.

Boolean functions take 2 inputs and one output.

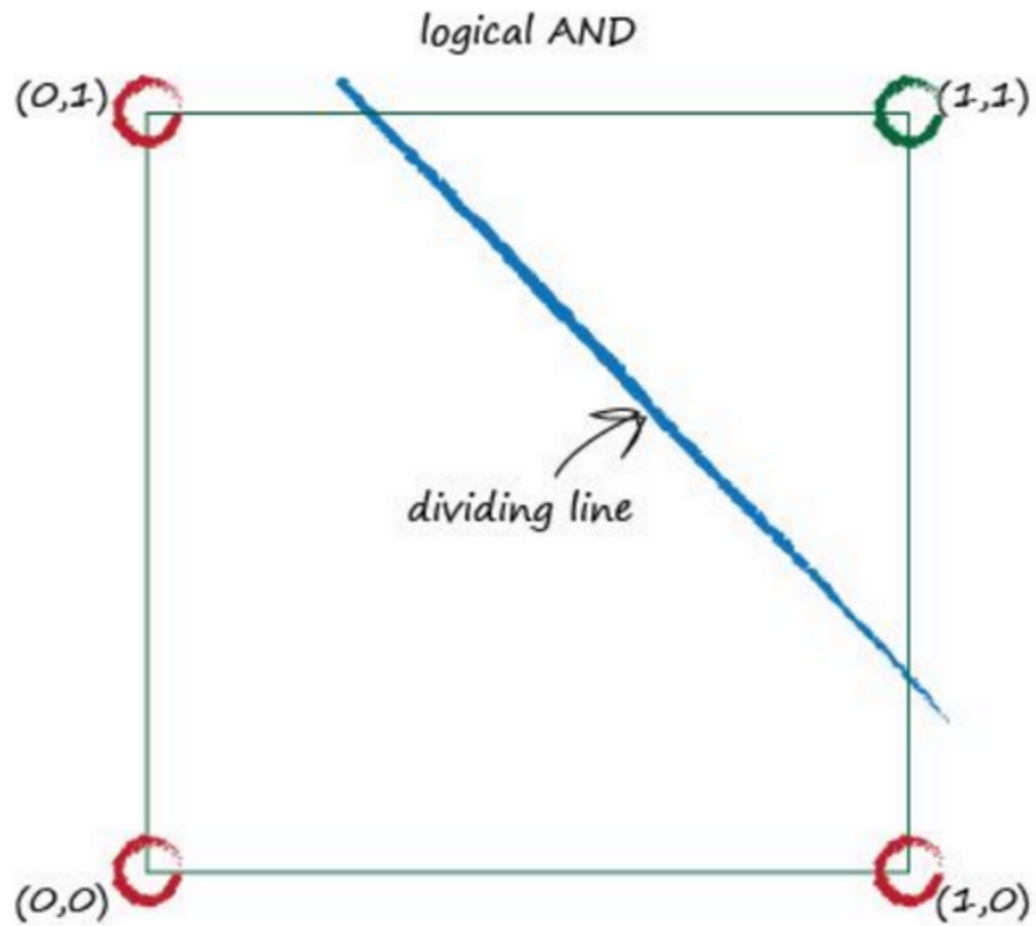


Computers often represent true as 1, and false as 0.

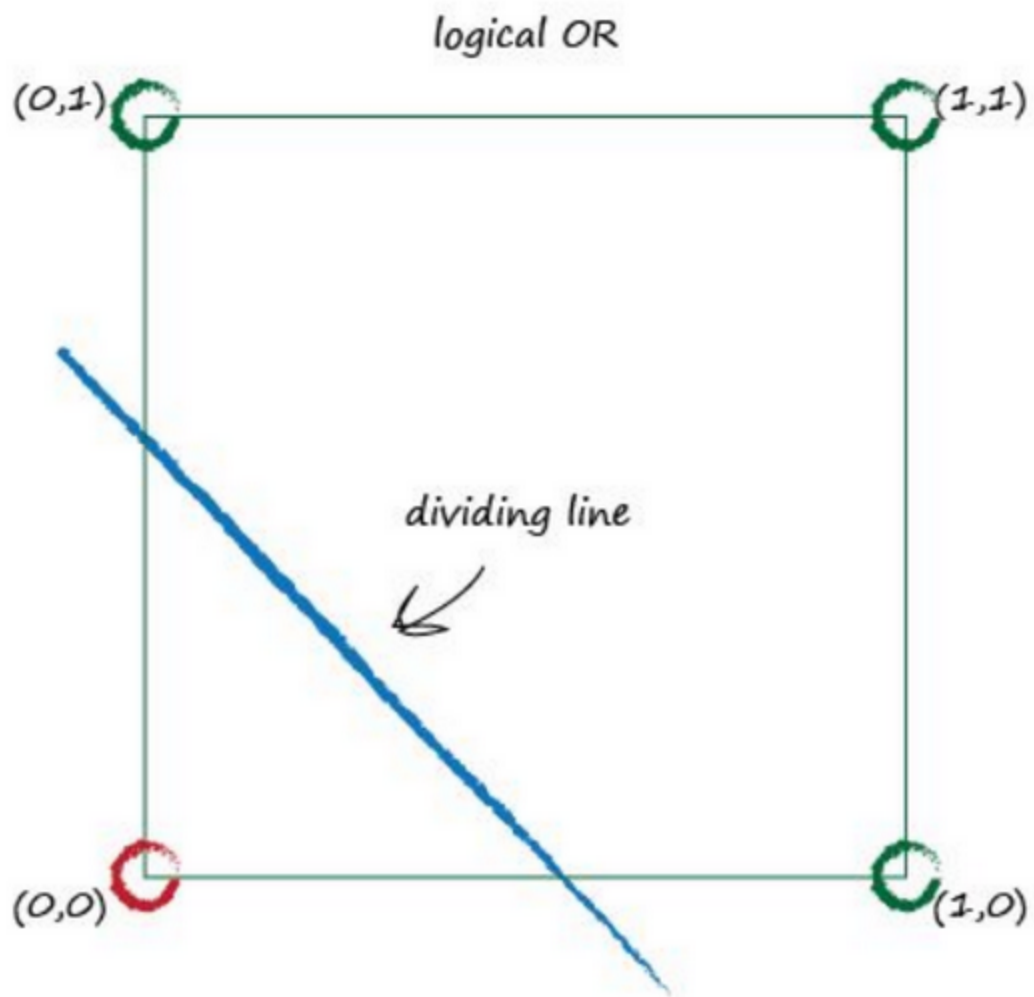
Input A	Input B	Logical AND	Logical OR
0	0	0	0

0	1	0	1
1	0	0	1
1	1	1	1

For logical AND, if we want to classify it into two separate classes we can make a dividing line as such, classifying the true from false



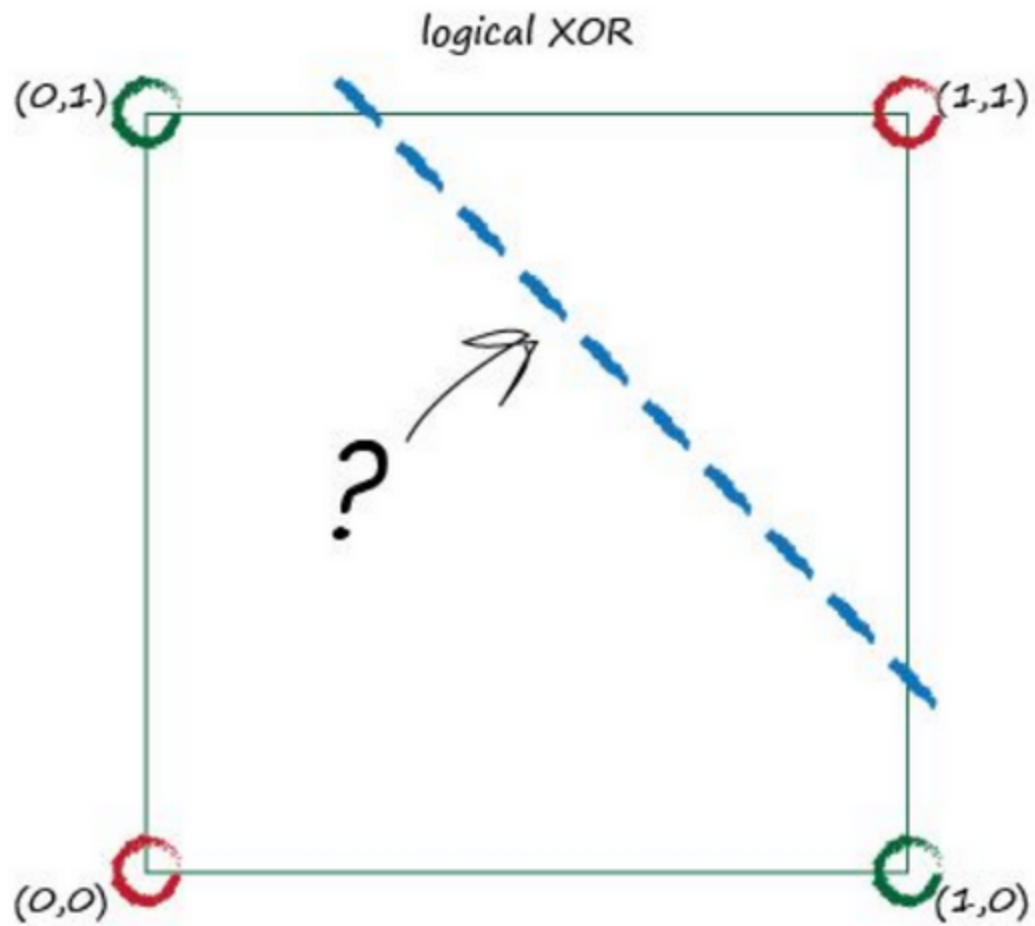
Same is the case with OR function



But what happens when XOR comes into the equation?

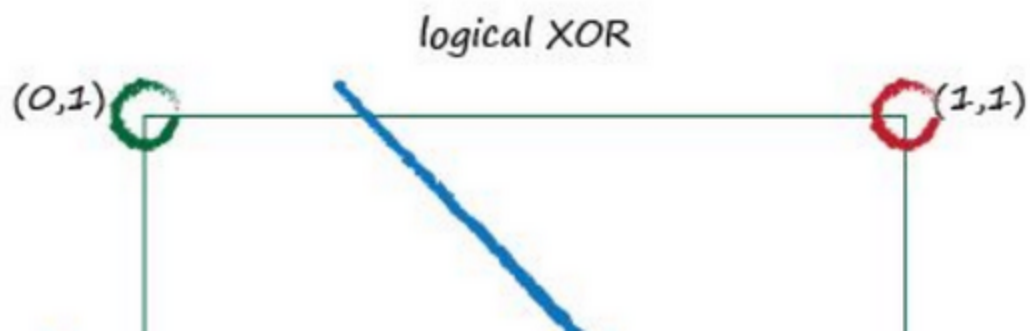
Input A	Input B	Logical XOR
0	0	0
0	1	1
1	0	1
1	1	0

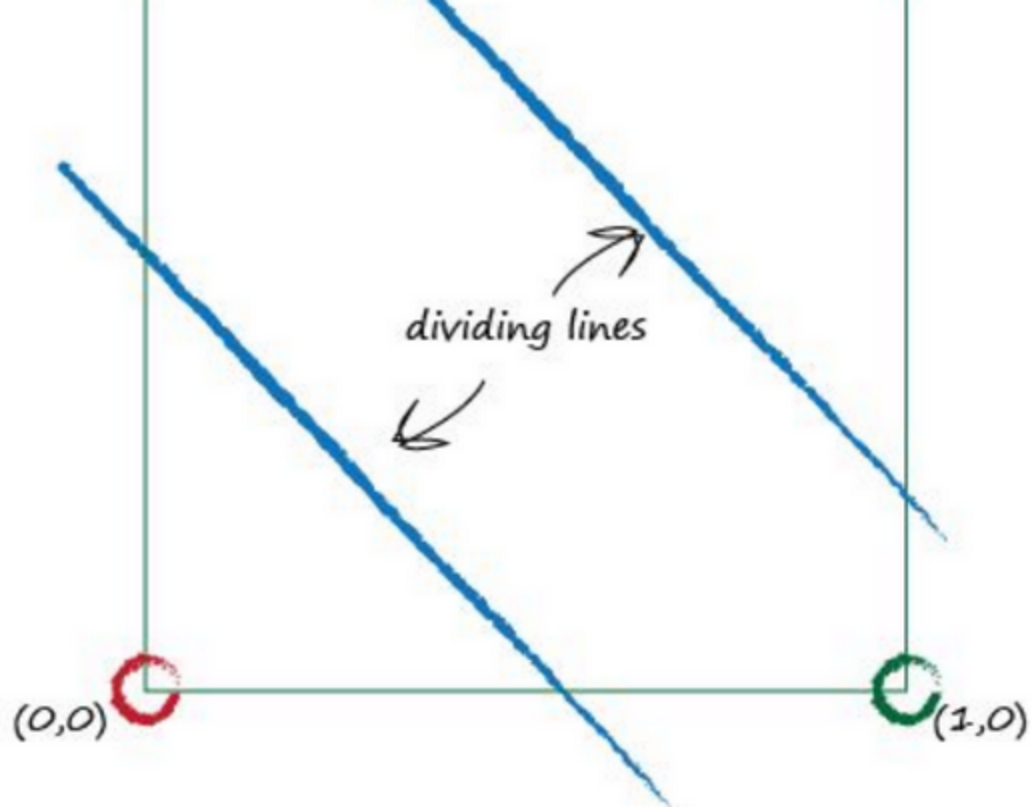
The figure can't be classified using a single line



In such a case, we need a fix. What would be the fix?

It would be to use two separate lines to classify different regions as shown below.





This is where the idea of using multiple classifier comes in. This idea is also central to neural networks.

Conclusion:

- A single classifier cant classify the data if it is not governed by a single linear process.
- The fix is easy just use multiple classifiers.