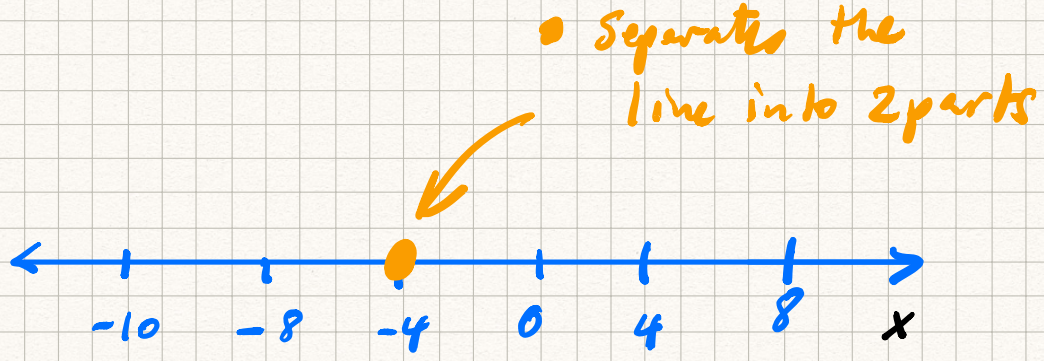
Separation is yet another guise in which classification problems arise.

Let's get an intuitive understanding of separation/classification.

1 Dimension

How can I separate this line (into 2 parts)?



if $x > -4$, $x \Rightarrow 1$; if $x \leq -4$, $x \Rightarrow 0$

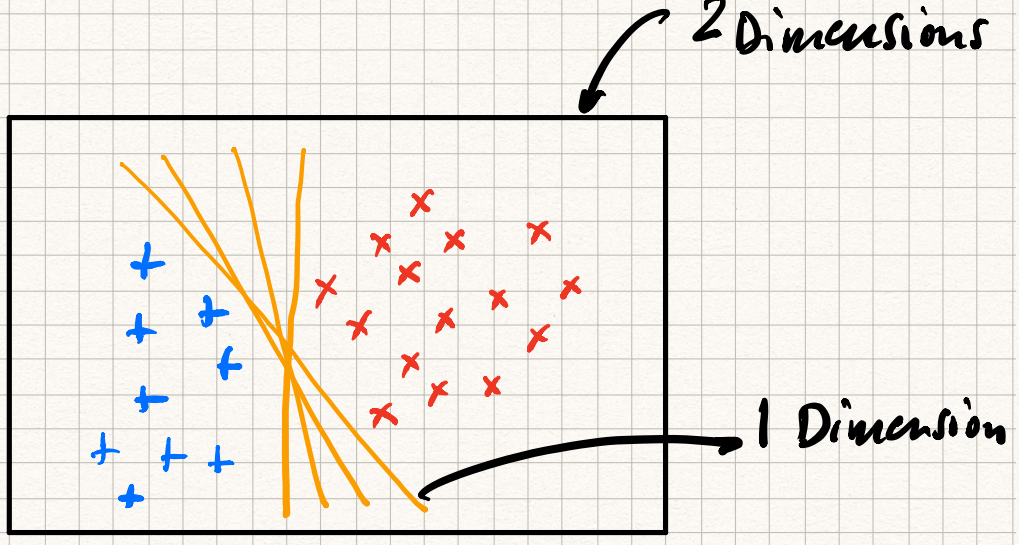
By placing a point somewhere on it.

1 dimension to separate

Separator = point = 0 dimension

2 Dimensions

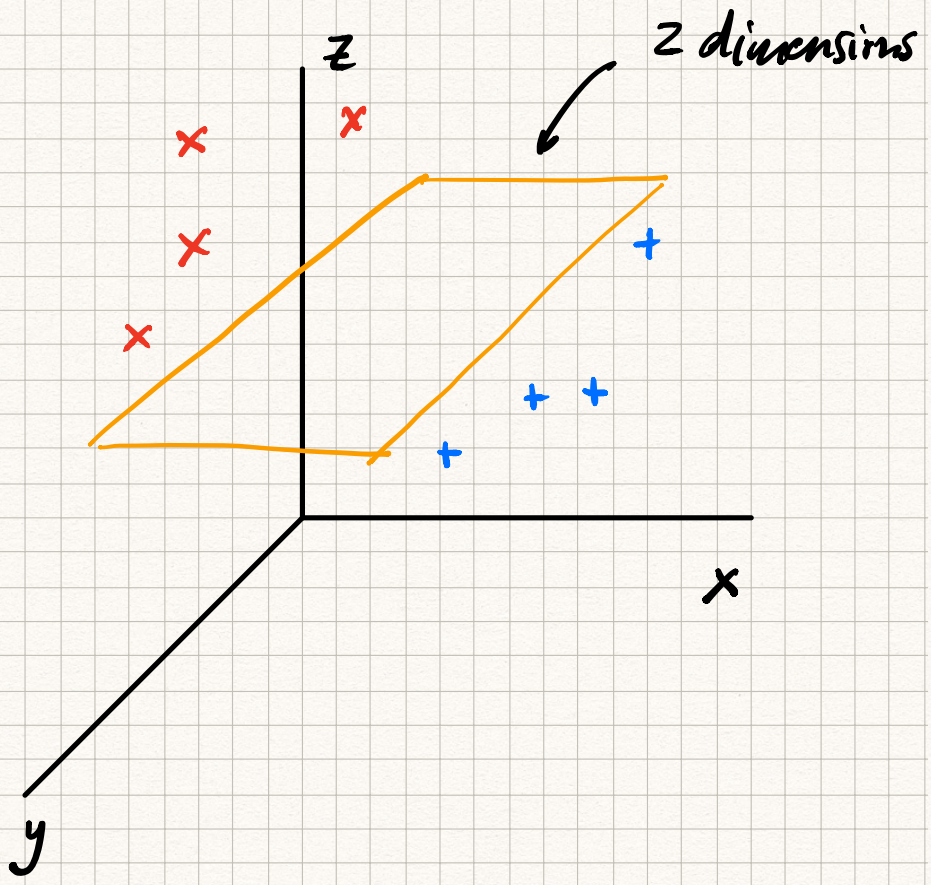
How to separate points on a plane?



By drawing a line that separates the points.

Seperator = line = 1 dimension

3 Dimensions



Separator = plane = 2 dimensions

42 Dimensions (why 42?)

Visualization

??

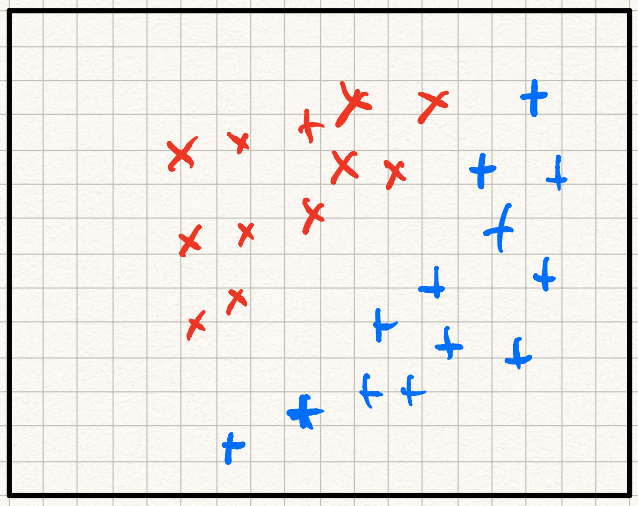
But we know that

Separator = 41-dimensional hyperplane

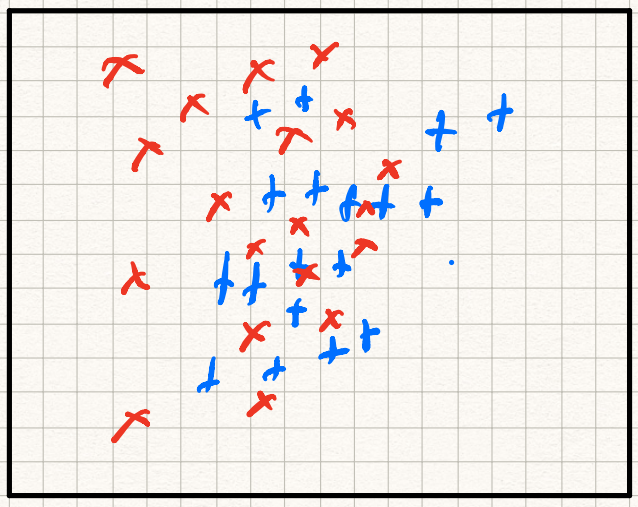
What's this got to do with our classification problem?

Thinking in terms of dimensions can help us deal with messy (i.e., real world) classification/separation problems.

simple



real world



How to separate this data into 2 classes / categories ?

Planes over Orlando

How do air traffic controllers keep planes separated?

By adding a dimension -

altitude.

Adding dimensions is a

standard "trick" that

mathematicians and physicists

love!

THE IDEA

Things that are not cleanly separable in n-dimensions could become cleanly separable in $n+1$ dimensions!!

That's what happened when we went from using just V_1 and V_2 to all 4 dimensions of the bank notes data set.

The data became linearly separable.

But what if the data are not linearly separable at this higher dimension?

3 options:

- Try a higher dimension (add another column of data)
- Use a non-linear separator (construct new features)
- Do both.

Questions

How many dimensions do I need?

What features should I construct?

Linear?

Non-Linear?

These questions are,

(roughly speaking)

automatically taken care

of by Support Vector

Machines (SVMs) and

Neural Networks.