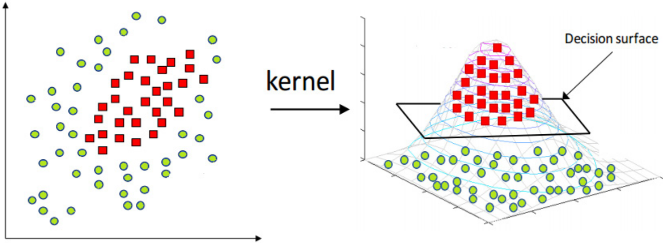
The "Kernel" trick

Going up makes your data linear



credits: https://medium.com/@zxr.nju

swipe right →

TOY EXAMPLE

Let's start with a simple classification problem.

You need to seperate 2 types of toys: bottle caps and lily (toy girl with red hat).

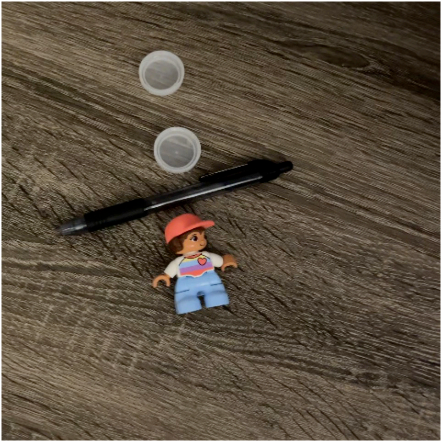
Let's assume that they are in 2D space.

swipe right →

You can easily seperate them using a pen (or a line).

You can place any straight object in between and the caps are on one side and Lily on the other.

This data is linearly seperable.

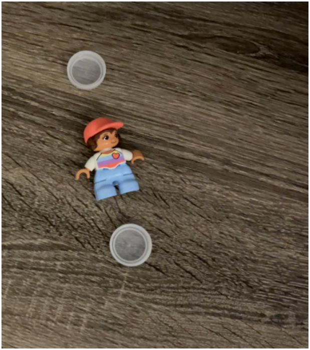


swipe right ⟶

Let's change the positions of the toys

Let's bring Lily to the middle.

Can you now seperate them with a pen?

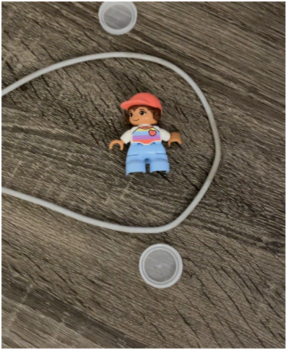


The answer is No!

swipe right →

Let's use a wire to seperate them.

A wire can easily seperate them because it is non-linear. This data right now is non-linear and cannot be seperated by a linear classifier unless...



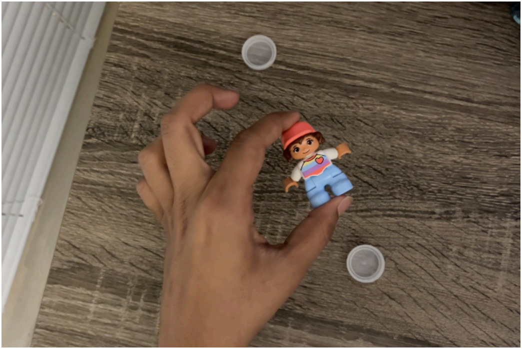
swipe right ⟶

You add a dimension

Let's make this 3D.

Let's add a dimension which depends on distance from center.

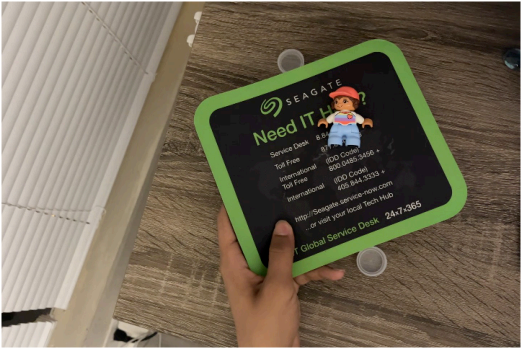
Because Lily is closer to the center than the caps, it is lifted higher than the caps.



swipe right ⟶

Are they linearly seperable now?

Yes. Because we are in 3D we can't use a pen anymore but we can use any 2D plane like a paper or mouse pad.


SEAGATE
Need IT H ?
Service Desk 8.84
Toll Free 87
International Toll Free (IDD Code) 800.0485.3456 +
International (IDD Code) 405.844.3333 +
http://Seagate.service-now.com
...or visit your local Tech Hub
24x7x365
T Global Service Desk

The pad is linear but the data is seperated. Lily on top and the caps on bottom.

swipe right ⟶

This is the intuition for "kernel" trick which SVM uses to seperate non-linear data.

This example has only 3 data points. Real datasets will have thousands. How do you find a plane that seperates them nicely?

That is where SVM comes in.

SVM is maximum margin classifier. It doesn't try to find a plane. It finds the plane that seperates the data optimally.

swipe right ⟶

Without going into the detail, SVM has the mathematical convinience to easily add dimensions to the data and find a fitting hyper-plane.

Of course finding a fitting plane in higher dimensions is not easy. It becomes harder as the dataset size increases.

But with the right kernels, SVMs does a good job and has been proven emperically to work in higher dimensions.

swipe right ⟶

The first SVM blog is out (1 of 3). It goes over the basic intuition, terminolgy and math.

The next blogs will focus on the kernel trick and using SVMs for NLP (with code).

You can read the blog in my newsletter.
https://convergence-point.beehiiv.com/

Subscribe so you don't miss out the later blogs of the NLP series.

We started with Naive Bayes Classifier and will cover everything in between until we can make our own Transfomer model. SEE YOU THERE.

swipe right →

THANK YOU FOR READING TILL THE END

SEE YOU IN THE NEXT ONE!

FOLLOW ME FOR MORE CONTENT LIKE THIS