



Announcement

We will start **checking in** with each team tentatively next week before break!

If having issues communicating with your project partners, let us know and we will coordinate or reassign groups

Dec 2 presentation **will take attendance** (each group 5-10 min)



Final Project! Finish before DECEMBER 2

Each group will present 5-10 min on DEC 2 normal meeting time

Skeleton:

<https://colab.research.google.com/drive/1vFZVxKKk8gSJdfmtVf5z6RdkjmPcACfD?usp=sharing>

As always, if there are any questions or concerns about this week's lab or the project, please don't hesitate to let us know!



Final Project Presentation Examples

- Education, Linguistic Isolation, and Poverty:

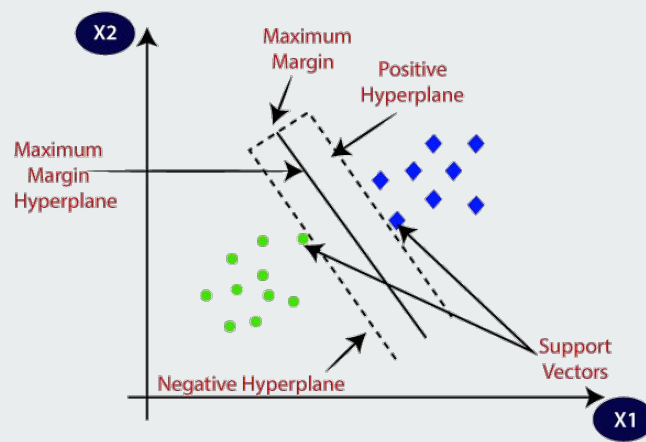
<https://docs.google.com/presentation/d/1QCzJjHykbj0NORZGNg6kHprJor2E2wiTiaA947XCCKI/edit?usp=sharing>

- ADIHP Healthcare Fraud Detection Project:

https://docs.google.com/presentation/d/1YGrR9sfcc76xVeWWqFRI_w_fR5crBvh1WHSMqijDZDk/edit?usp=sharing



SVM (Support Vector Machines)



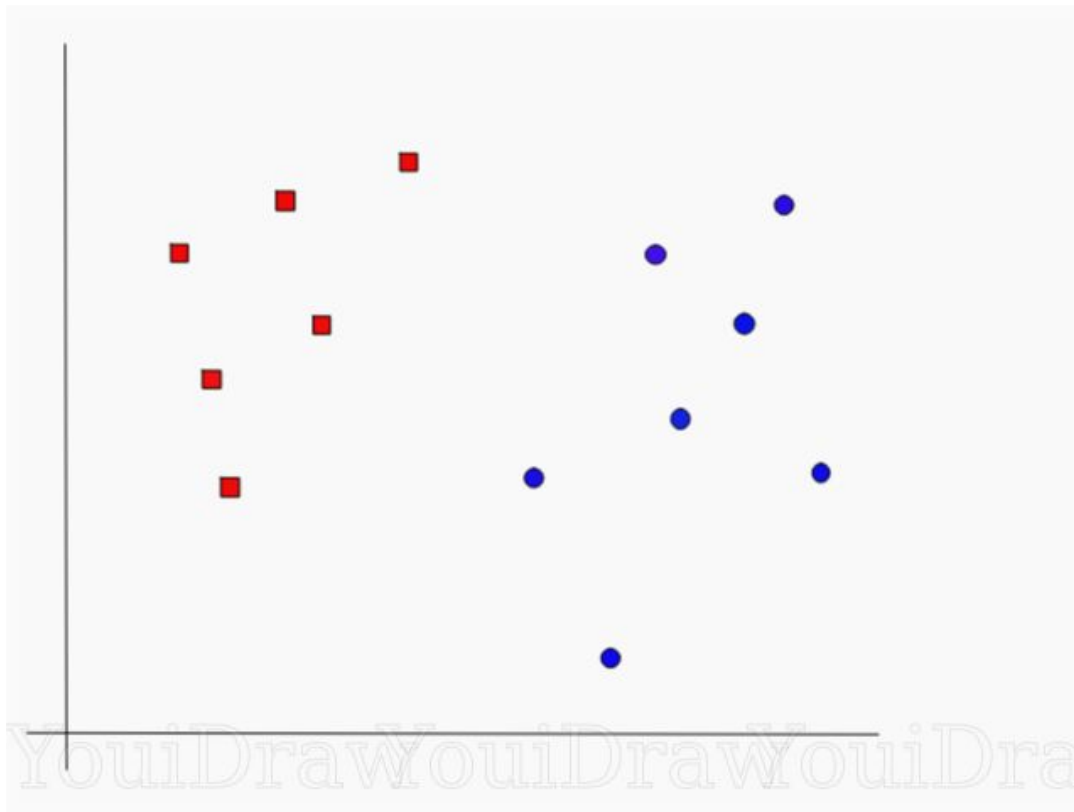


What is SVM?

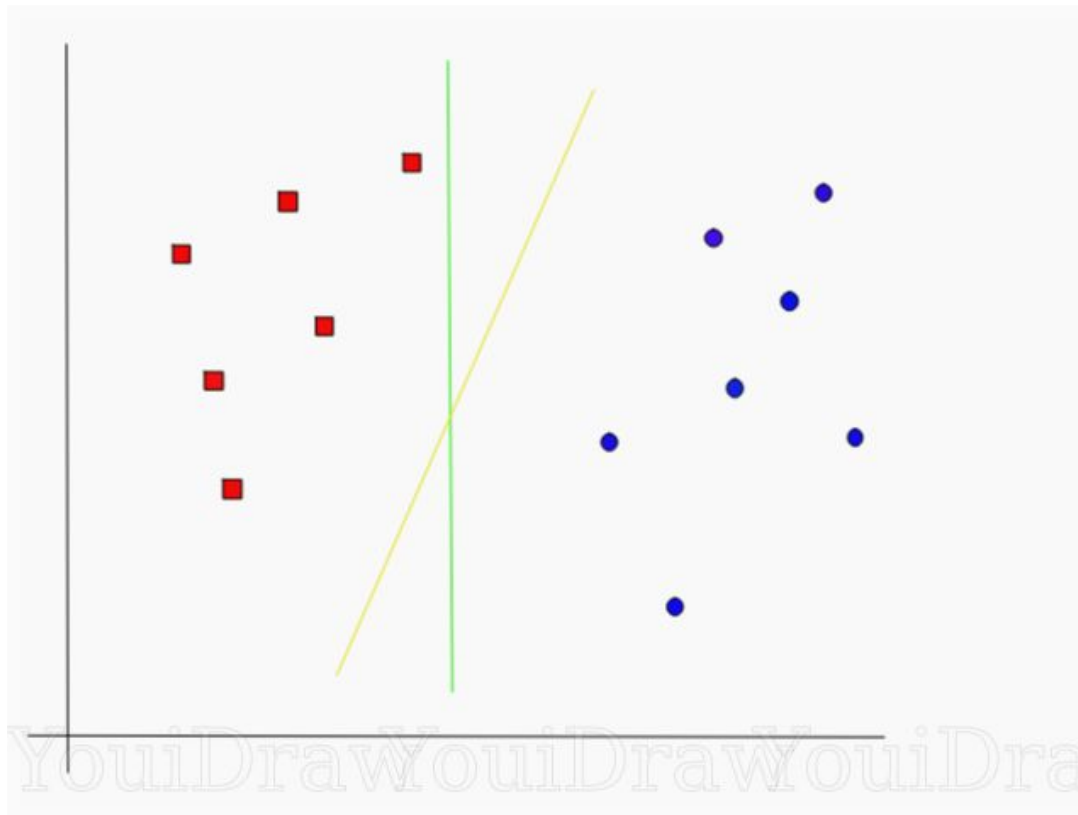
A **linear model** for **classification** and regression problems.

- Can solve linear and non-linear problems.
- Basic algorithm: creates a line or a hyperplane which **separates the data into classes**.

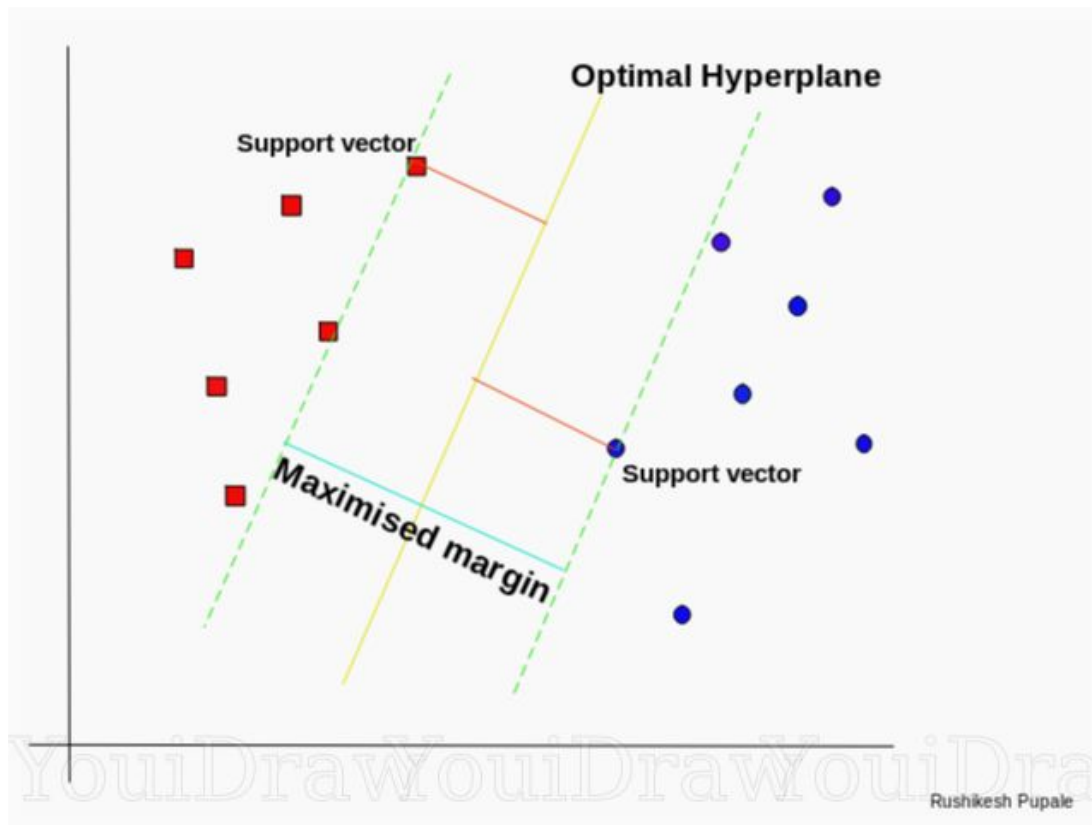
A classification example:



A classification example:



How SVM select better line?





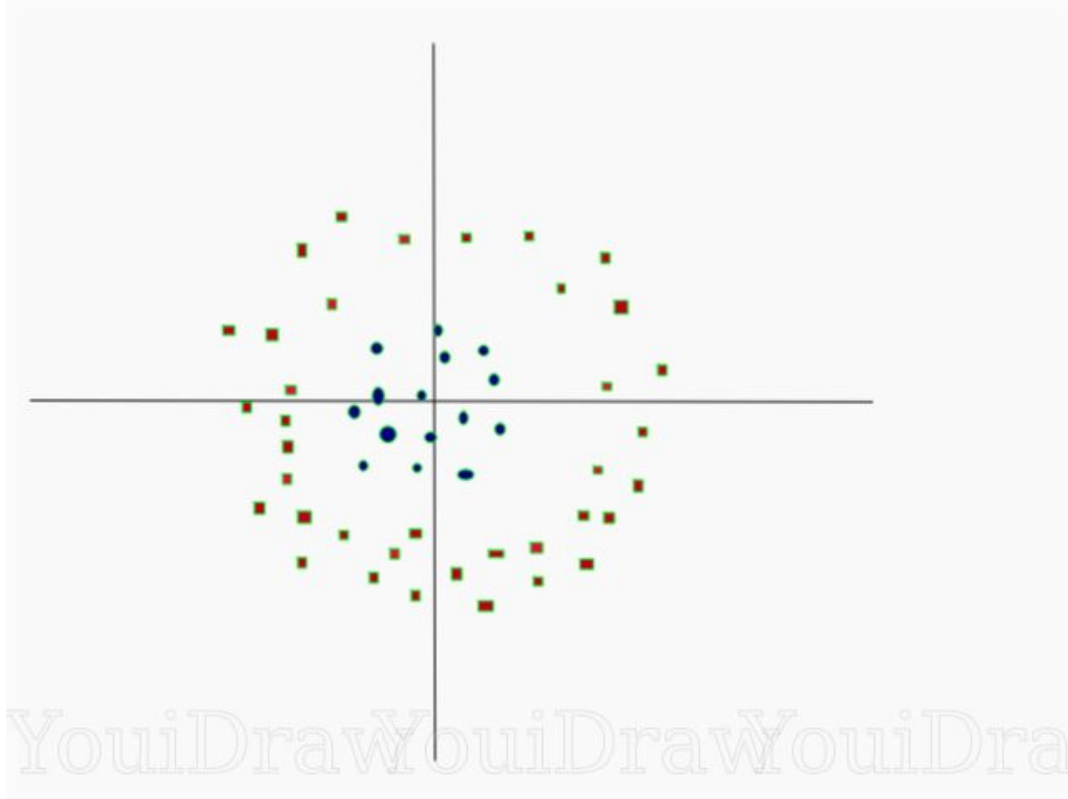
How it works?

Goal: to maximize the **margin**, so that:

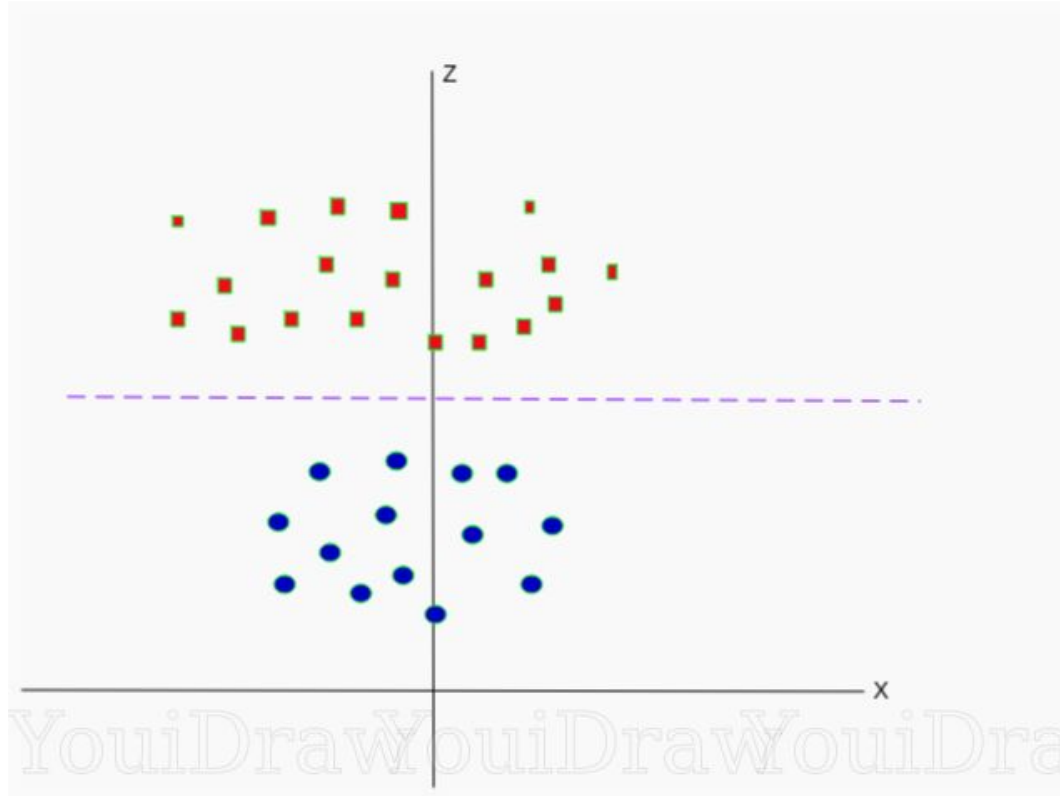
the separation between the two classes(that street) is as wide as possible

- **Support vectors:** we find the points **closest to the line** from both the classes. These points are called
- **Margin:** compute the distance between the line and the support vectors

More complex example (higher dimension):



More complex example (higher dimension):





How it works?

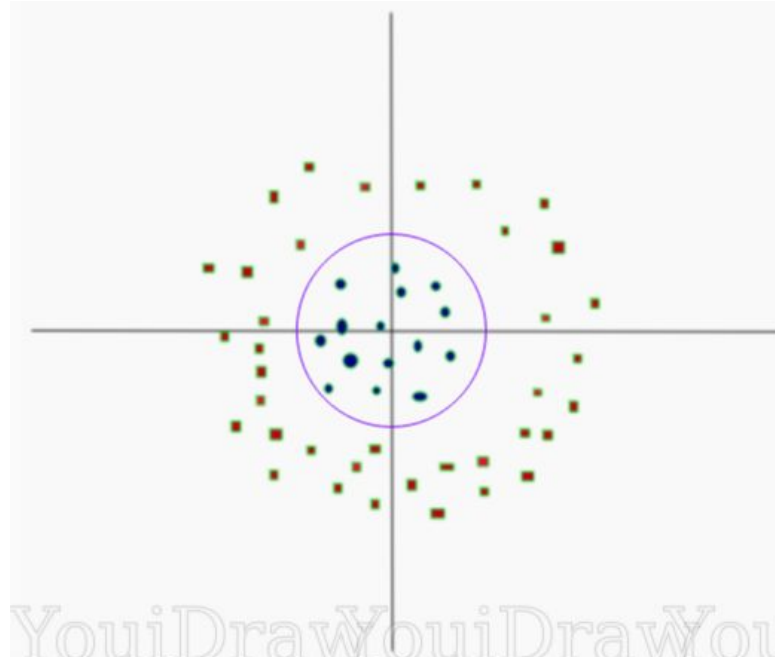
$$z = x^2 + y^2$$

Let $z = k$

then $x^2 + y^2 = k$

And then we can project the

separation back to original dimension using this transformation:



We use kernel functions to do the classification



1. Linear Function

$$k(x_i, x_j) = x_i * x_j$$

2. Polynomial Function

$$k(x_i, x_j) = (1 + x_i * x_j)^d$$

We use kernel functions to do the classification



3. Radial Basis Function (RBF)

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

4. Sigmoid Function

$$k(x_i, x_j) = \tanh(\alpha x^T y + c)$$



What if we want to deal with multiple classes?

i.e. we want to see how diamond depth and carat impacts color:

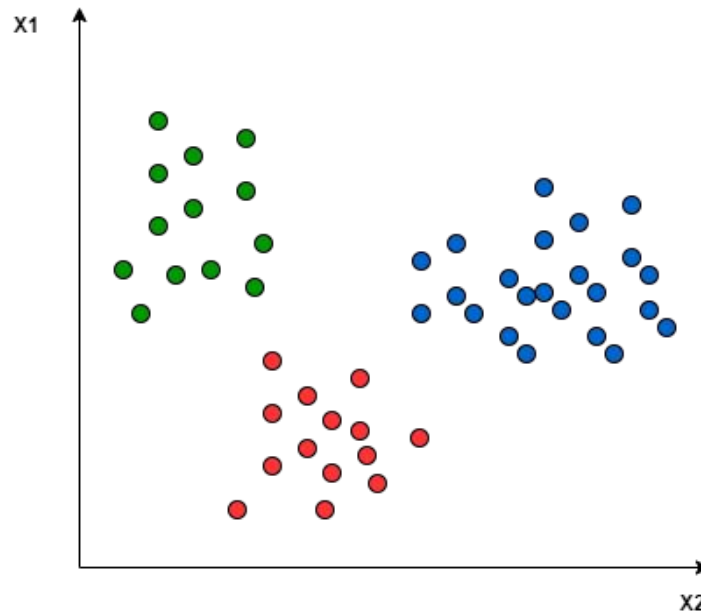
	carat	cut	color	clarity	depth	table	price	x	y	z
0	0.23	Ideal	E	SI2	61.5	55.0	326	3.95	3.98	2.43
1	0.21	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
2	0.23	Good	E	VS1	56.9	65.0	327	4.05	4.07	2.31
3	0.29	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63
4	0.31	Good	J	SI2	63.3	58.0	335	4.34	4.35	2.75



What if we want to deal with multiple classes?

We might probably have a
bunch of classes like this: 🙌

Each color represents a class
of diamond



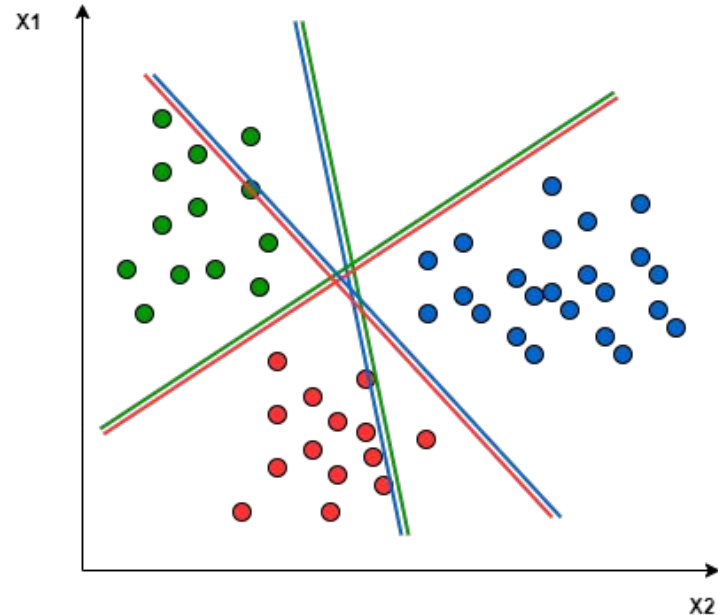


What if we want to deal with multiple classes?

One-to-one approach

separate between every two classes

i.e. red-blue line tries to maximize the separation only between blue and red points



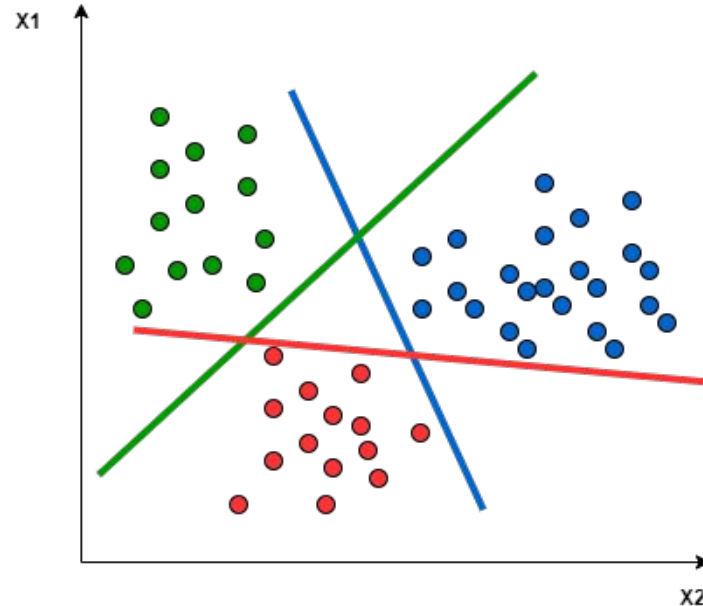


What if we want to deal with multiple classes?

One-to-rest approach

separation takes all points into account, dividing them into **two groups**

one group for **the class points**;
one group for **all other points**





Where to use SVM?

- Use ^{X x-coordinate} **sepal length** and ^{X y-coordinate} **petal length** to determine the **flower's**
classes
possible species
- Classification of news articles into “business” and “Movies”
- Classification of web pages into personal home pages and others
- Classification of genes, patients on the basis of their genes



Tips of real-life usage

- The x y data should be numerical
- Probably need to scale your data to be the ranges of $[0, 1]$ or $[-1, 1]$.
- Real dataset example: [link](#)
 - Feel free to use it as sample code for final project



How do you implement SVM?

Colab demo!!!

https://colab.research.google.com/drive/1U5QxXT_Ni1NtZyBd-yjLa7zNGtNl8DS6?usp=sharing

As always, if there are any questions or concerns about this week's lab or the project, please don't hesitate to let us know!