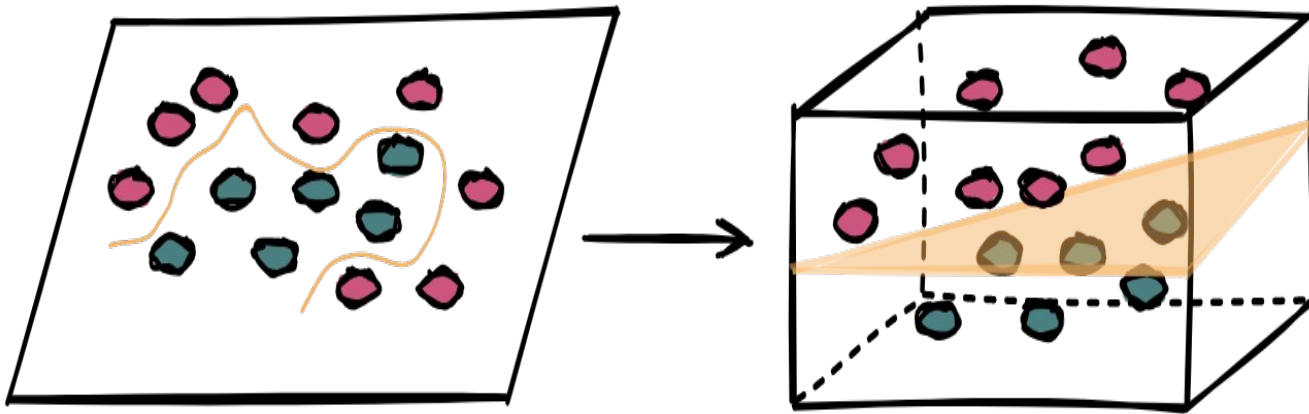


Support vector machines (SVM)



Week 12

Middlesex University Dubai;
CST4050; Instructor: Dr. Ivan Reznikov

SVM concept

Support vector machine (SVM) is a simple algorithm that produces significant accuracy with relatively low computation power. SVM can be used for regression problems, but more often used to solve classification tasks.

For the purpose of today's session, we'll slightly modify our usual dataset. Let's include a level column, that will be 1 for Overall Rating ≥ 70 , else 0.

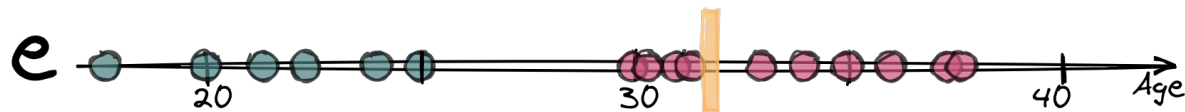
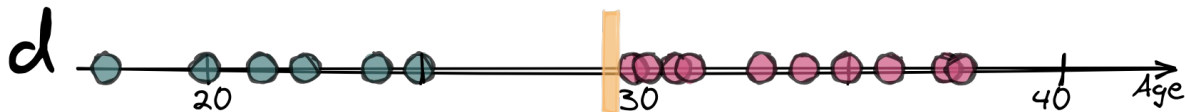
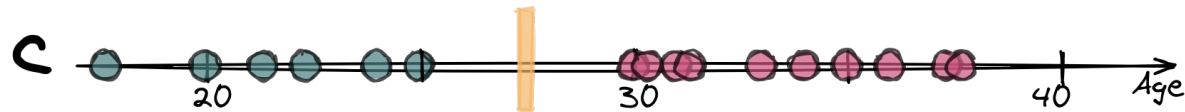
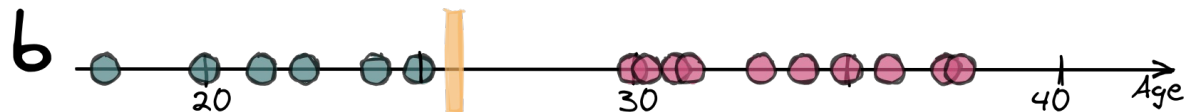
target		new target	
Overall Rating	Level	Age	
40	0	37	
50	0	30	
55	0	36	
55	0	31	
60	0	35	
60	0	31	
60	0	38	
60	0	34	
65	0	33	
65	0	30	
65	0	27	
65	0	23	
70	1	34	
70	1	29	
70	1	22	
75	1	25	
75	1	21	
80	1	24	
85	1	20	
90	1	18	

(a) 1Dimensional Data

target	new target	
Overall Rating	Level	Age
40	0	37
50	0	30
55	0	36
55	0	31
60	0	35
60	0	31
60	0	38
60	0	34
65	0	33
65	0	30
65	0	27
65	0	23
70	1	34
70	1	29
70	1	22
75	1	25
75	1	21
80	1	24
85	1	20
90	1	18

test

What classifier would you pick?

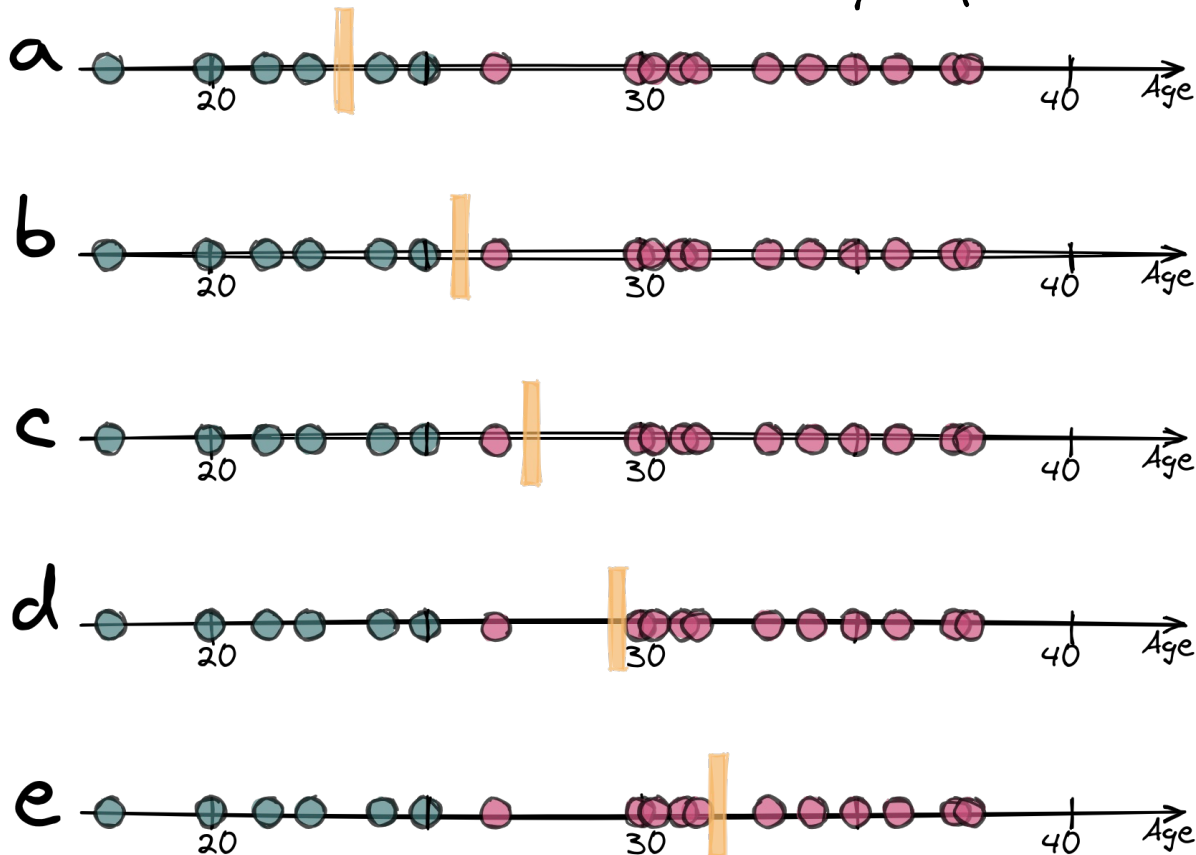


(b) 1Dimensional Data

What classifier would you pick now?

target	new target	
Overall Rating	Level	Age
40	0	37
50	0	30
55	0	36
55	0	31
60	0	35
60	0	31
60	0	38
60	0	34
65	0	33
65	0	30
65	0	27
65	0	23
70	1	34
70	1	29
70	1	22
75	1	25
75	1	21
80	1	24
85	1	20
90	1	18

test

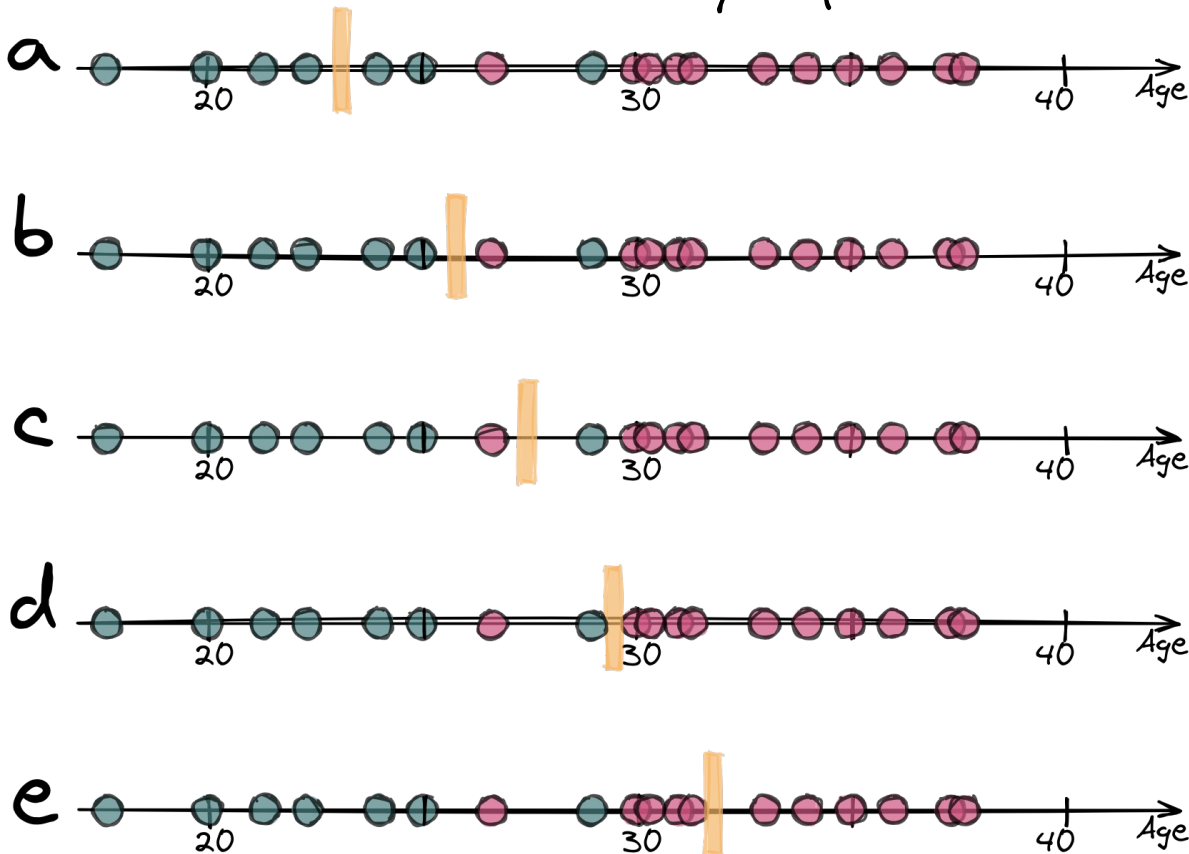


(c) 1Dimensional Data

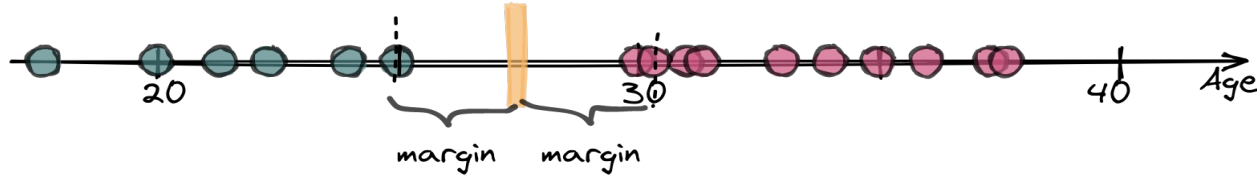
target	new target	
Overall Rating	Level	Age
40	0	37
50	0	30
55	0	36
55	0	31
60	0	35
60	0	31
60	0	38
60	0	34
65	0	33
65	0	30
65	0	27
65	0	23
70	1	34
70	1	29
70	1	22
75	1	25
75	1	21
80	1	24
85	1	20
90	1	18

test

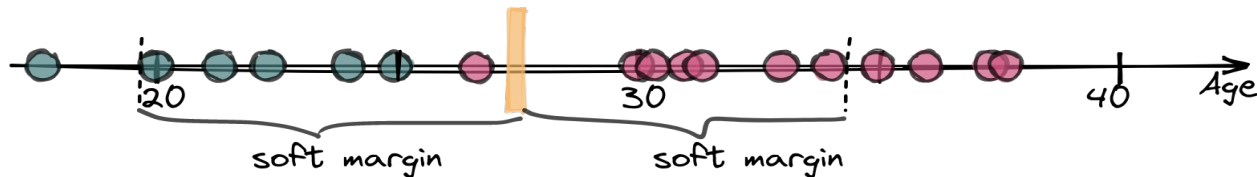
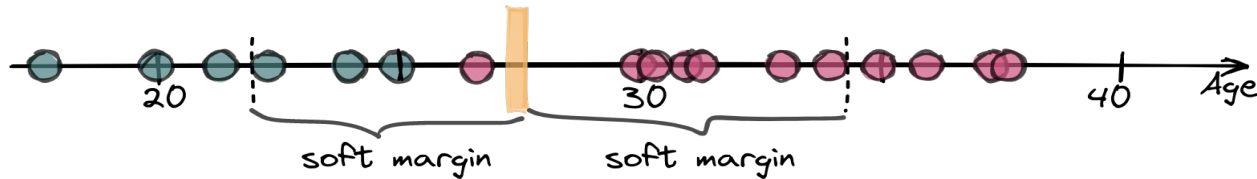
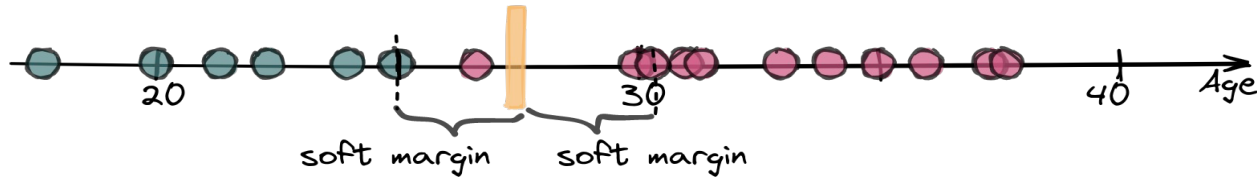
What classifier would you pick this time?



Soft margins



margin \rightarrow max \Rightarrow Maximum Margin Classifier



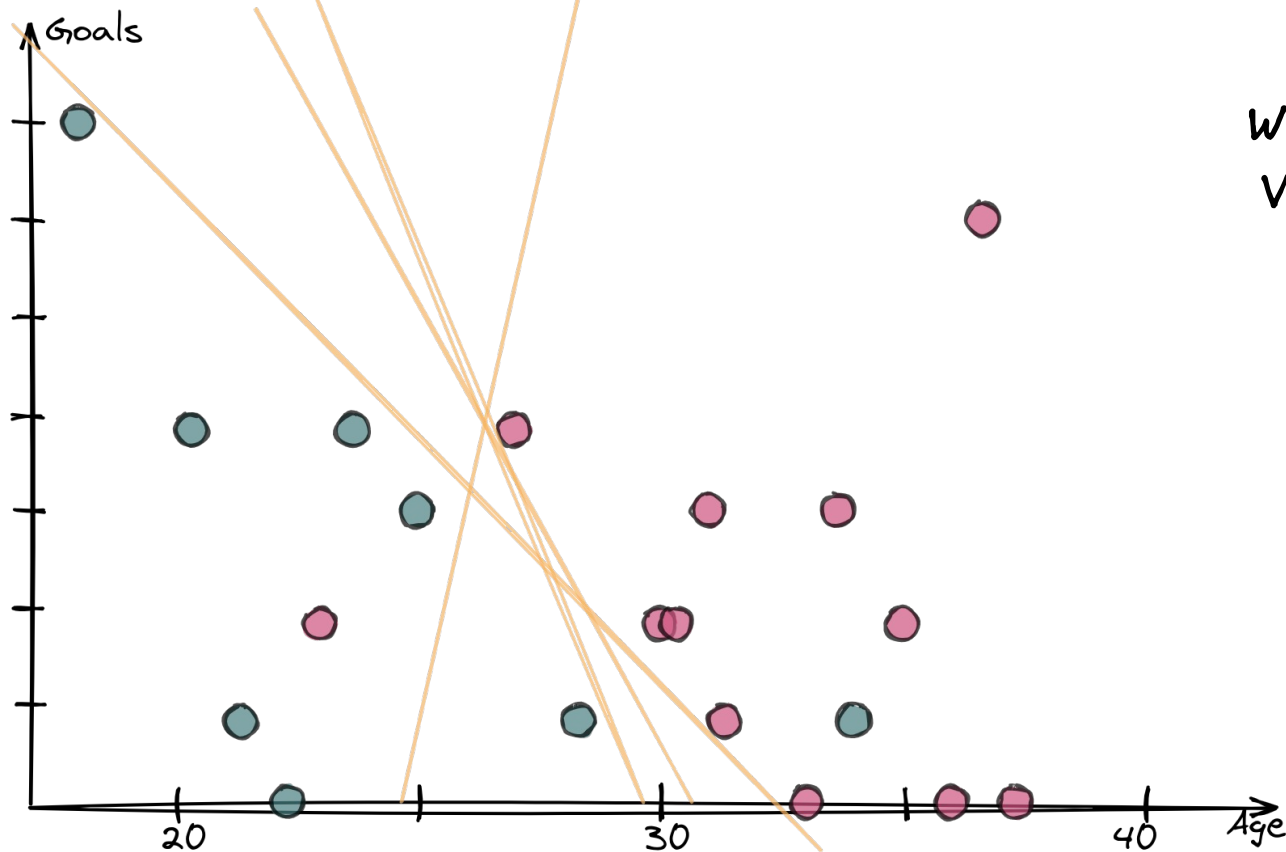
For 1D we only need a point to set threshold. It is considered as 0-dimensional hyperplane

Soft margins are more flexible and allow misclassification.

The thresh may be found empirically using cross-validation

2Dimensional Data

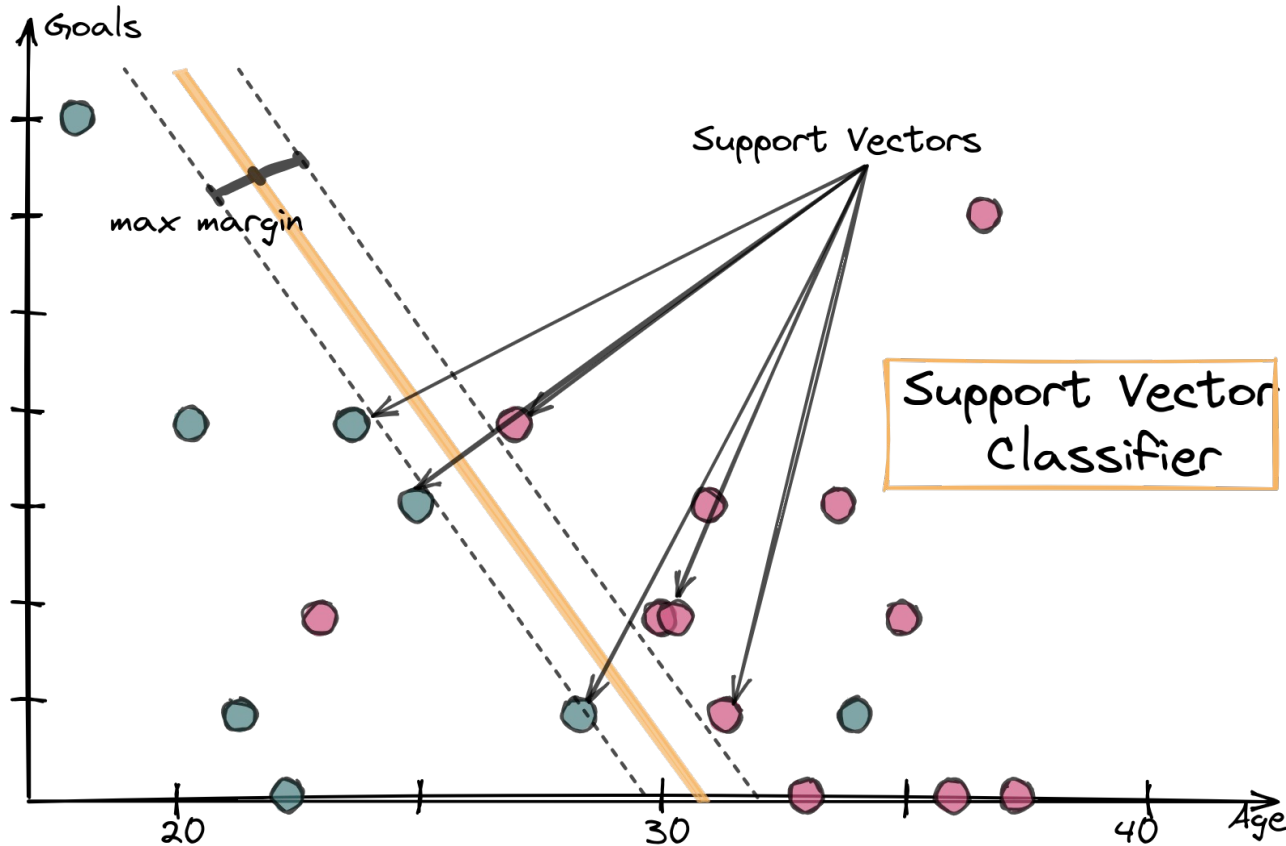
What classifier would you pick?



We can look for Support Vectors – border points of certain class.

After, we look for the maximum margin size

2Dimensional Data

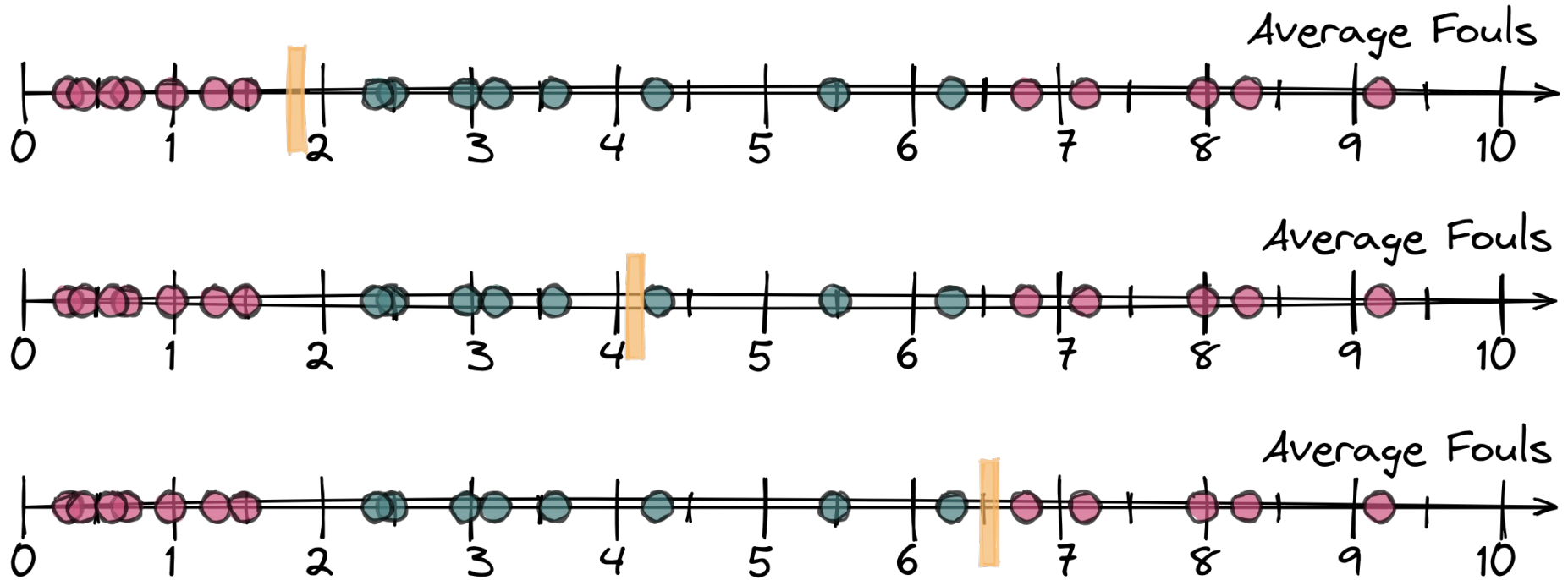


For 2Dimensional Data we need a line to set threshold. It is considered as 1-dimensional hyperplane.

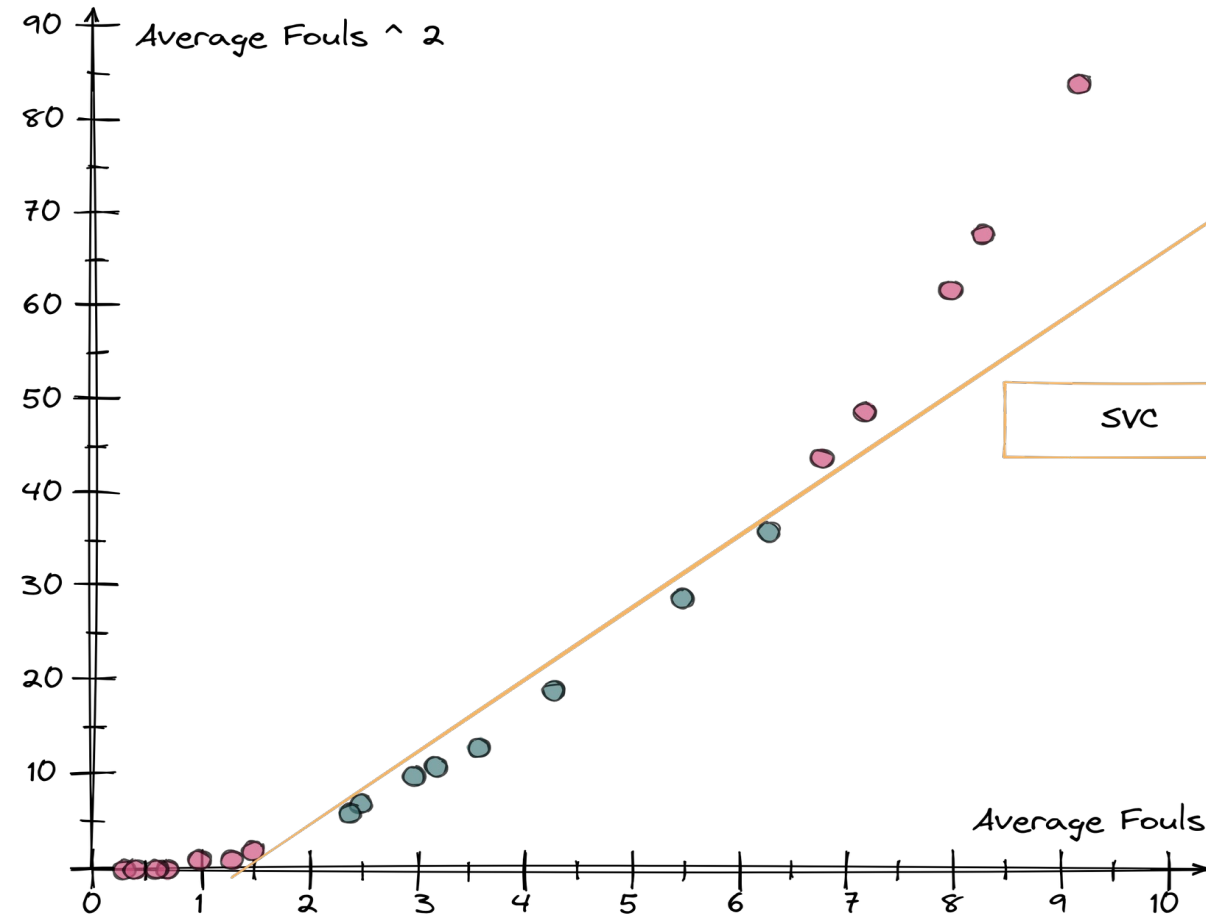
Similar situation we can imagine for higher dimensions.

SVM: Linear Kernel

What classifier would you pick?



SVM: Polynomial Kernel



We can plot our data in the following coordinates:

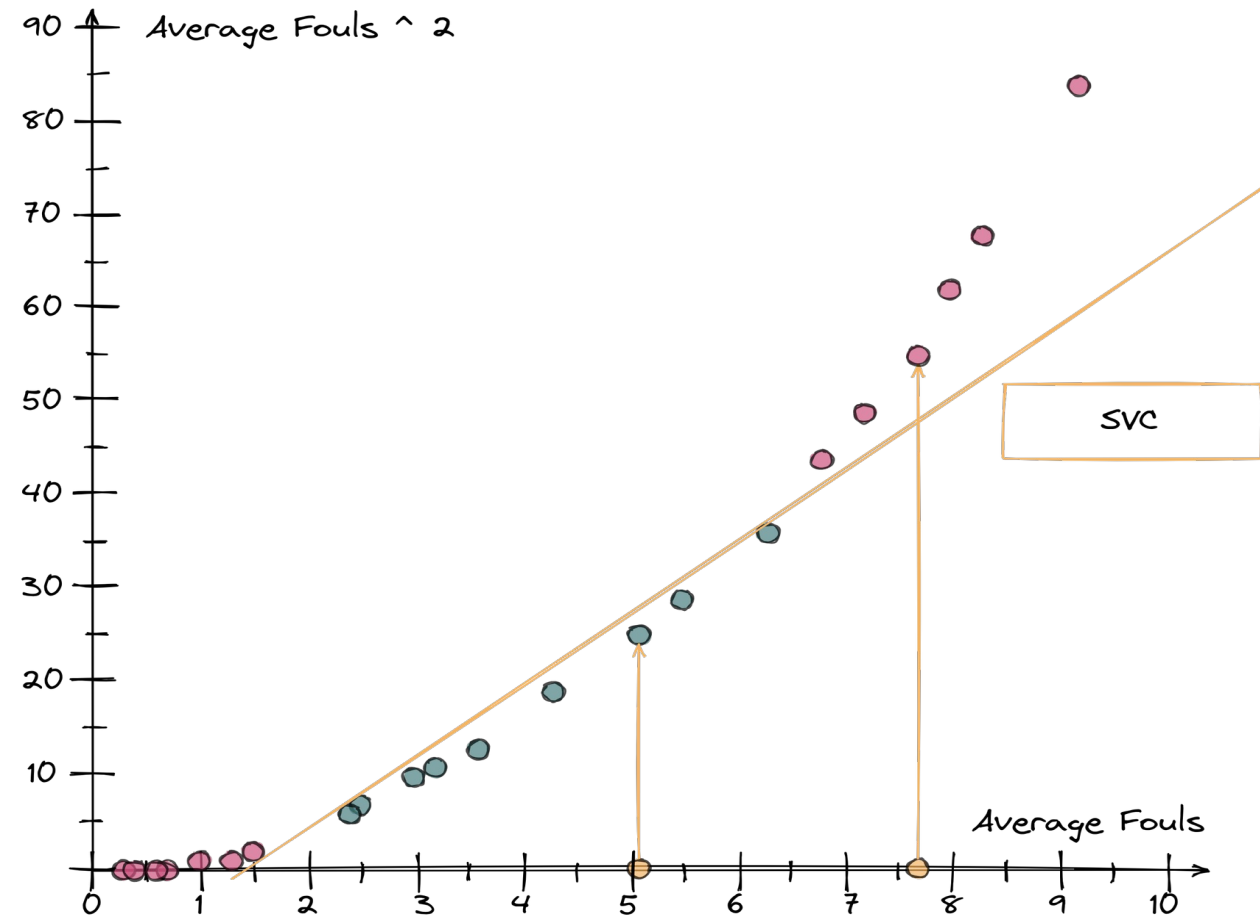
$$X = x$$

$$Y = x^2$$

Now we can look for a linear support vector classifier

The algorithm of transforming data and finding SVC is called
Support Vector Machines

SVM: Polynomial Kernel



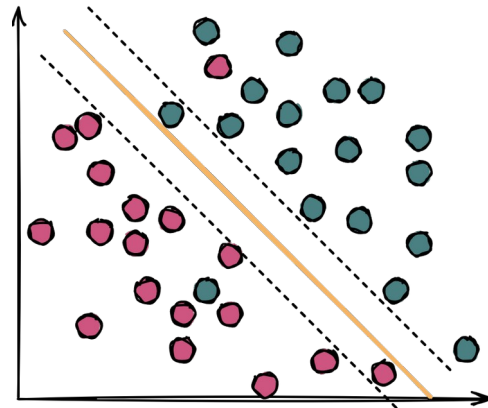
In order to classify a new data point, we'll need to classify it's x - x^2 position.

We could've used x^3 , x^4 or other formula. These are called **kernel functions**

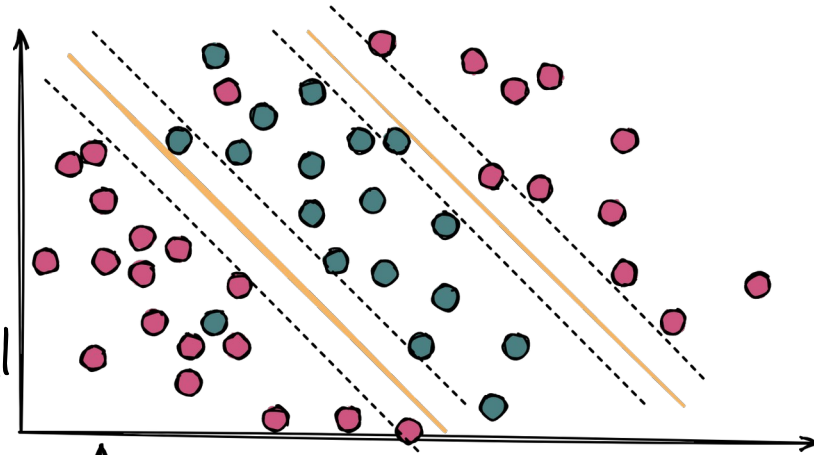
Also, rather than transforming data we can calculate pair relationships between data points.

This is known as **kernel trick**

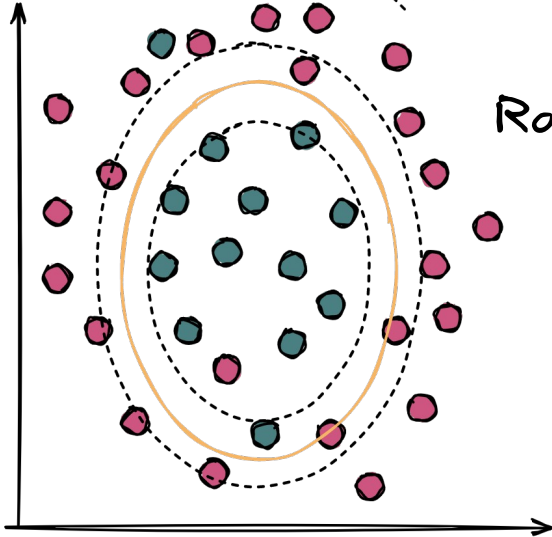
SVM: Types of Kernels



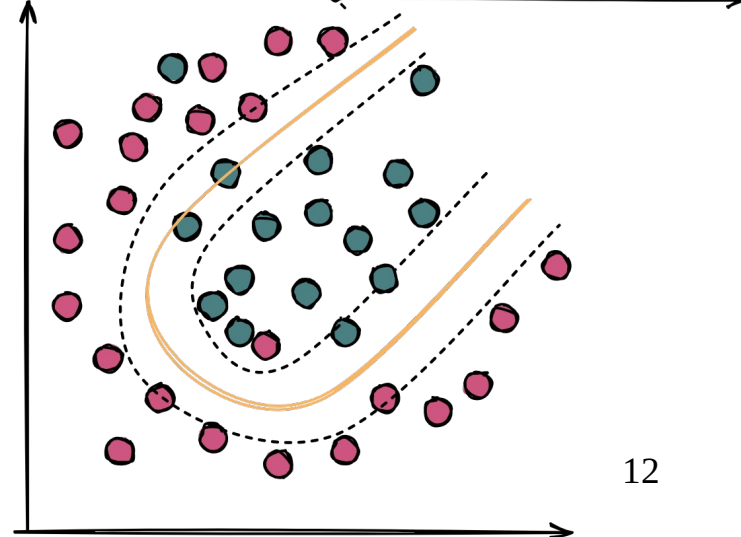
Linear kernel



Sigmoid kernel

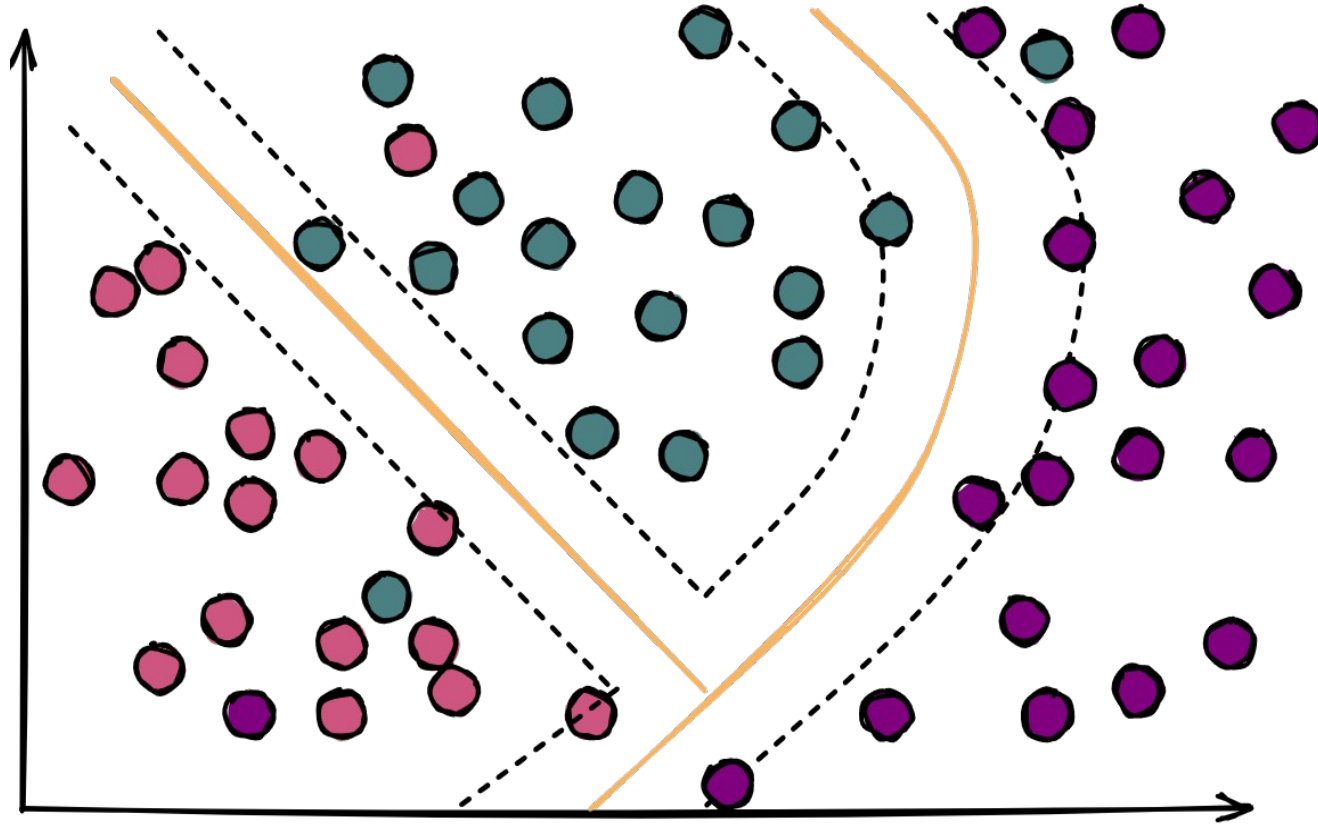


Radial kernel



Polynomial kernel

SVM: Multi-Class



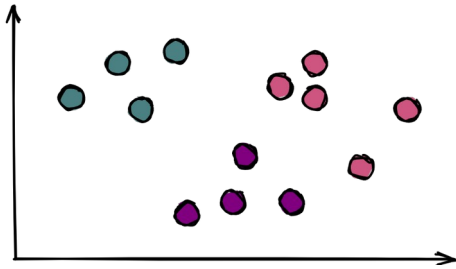
Besides binary classification (2 classes) SVM can handle tasks classifying more than 2 classes (multi-class).

Strategies involved:

- 1 vs all
- 1 vs 1

SVM: Multi-Class: 1 vs all

Feature1	Feature2	Target1
1	1	Red
2	10	Green
4	100	Purple
...



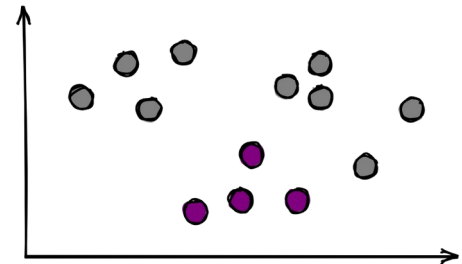
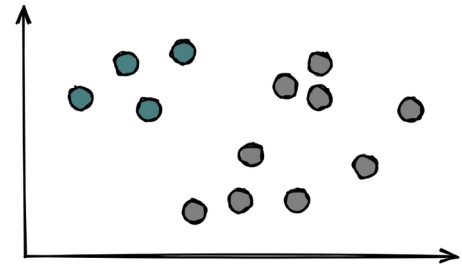
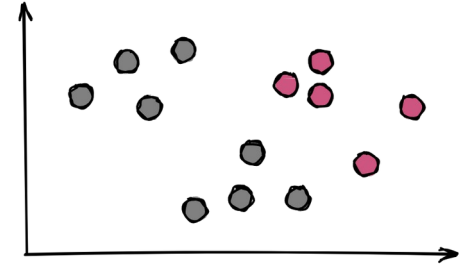
With the updated data we now calculate prob of each class. The class with highest score wins.

Creating training sets

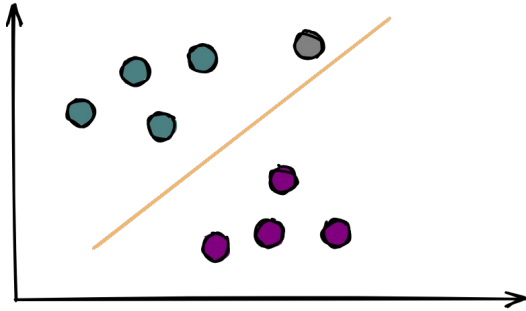
Feature1	Feature2	Target1
1	1	1
2	10	-1
4	100	-1
...

Feature1	Feature2	Target1
1	1	-1
2	10	1
4	100	-1
...

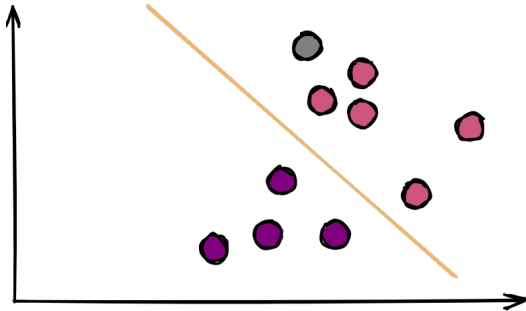
Feature1	Feature2	Target1
1	1	-1
2	10	-1
4	100	1
...



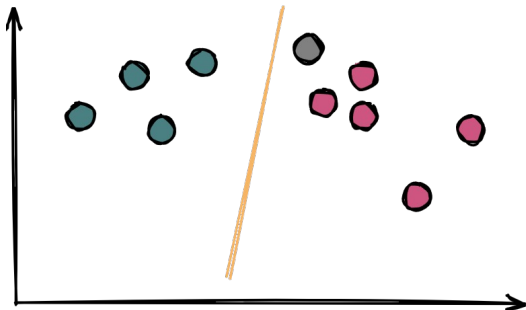
SVM: Multi-Class: 1 vs 1



Green vs Purple
Winner: Green

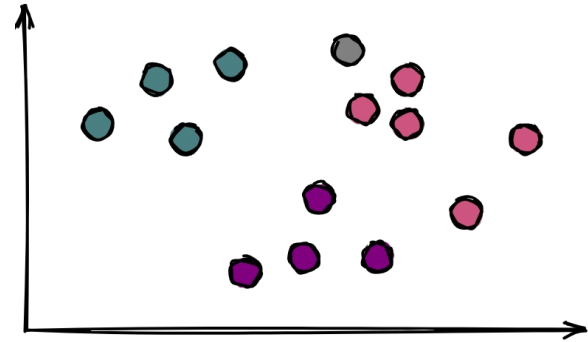


Red vs Purple:
Winner: Red



Green vs Red:
Winner: Red

Result: Red



Compared to 1 vs all, 1 vs 1 compares whether the data point should be classified as either class. The data point is assigned to class with most "wins".

Support Vector Machines



- Works well on datasets with many features
- Provides a clear separation margin
- Effective for datasets where the number of features are greater than the number of data points
- Possible to specify different kernel functions to make a proper decision boundary



- Require high training time, so not recommended for large datasets.
- Very sensitive to outliers.