08/03/2023

INTRODUCTION TO MACHINE LEARNING ITC2252

Chapter 5 – Support Vector Machine (SVM)

Dr. Chamara Liyanage
2023/03/07

Support Vector Machine (SVM)

another machine learning model → after linear regression

↓

not a ML model

- A popular classification technique
- In SVMs, the optimization objective is to maximize the margin
- The margin is defined as the distance between the separating hyperplane and the training samples that are closest to this hyperplane (support vectors)
- Intuitively, the large the margin, the lower generalization error
- Models with small margin prone to overfitting
- Supervised machine learning technique
- It is possible to use for regression problem also

line created in linear regression

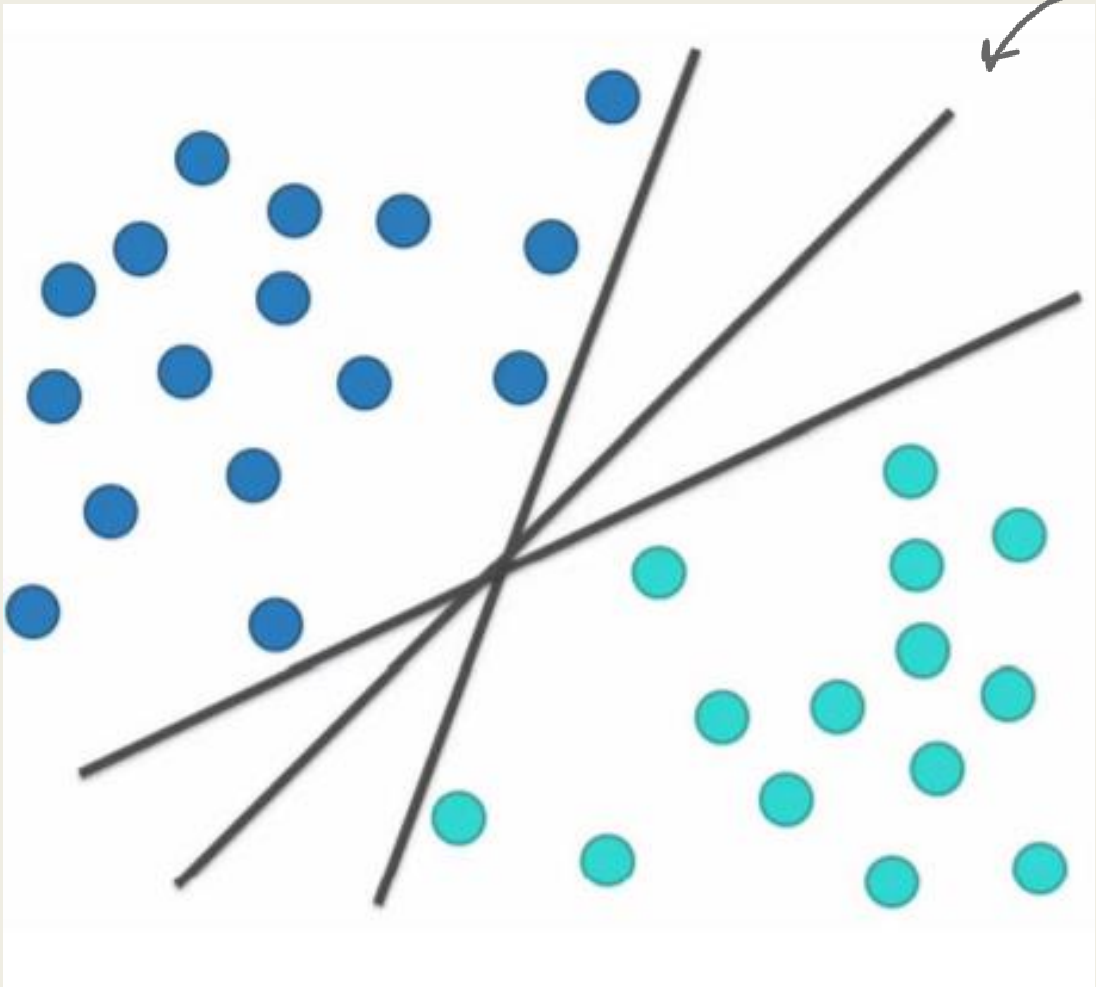
provide data and outputs both

What should you know ?

- Support vectors
- Hyperplanes regression line
- Marginal Distance
- Linear separable points Data can be seperated using line into classes
- Nonlinear Separable

can not seperate using a line

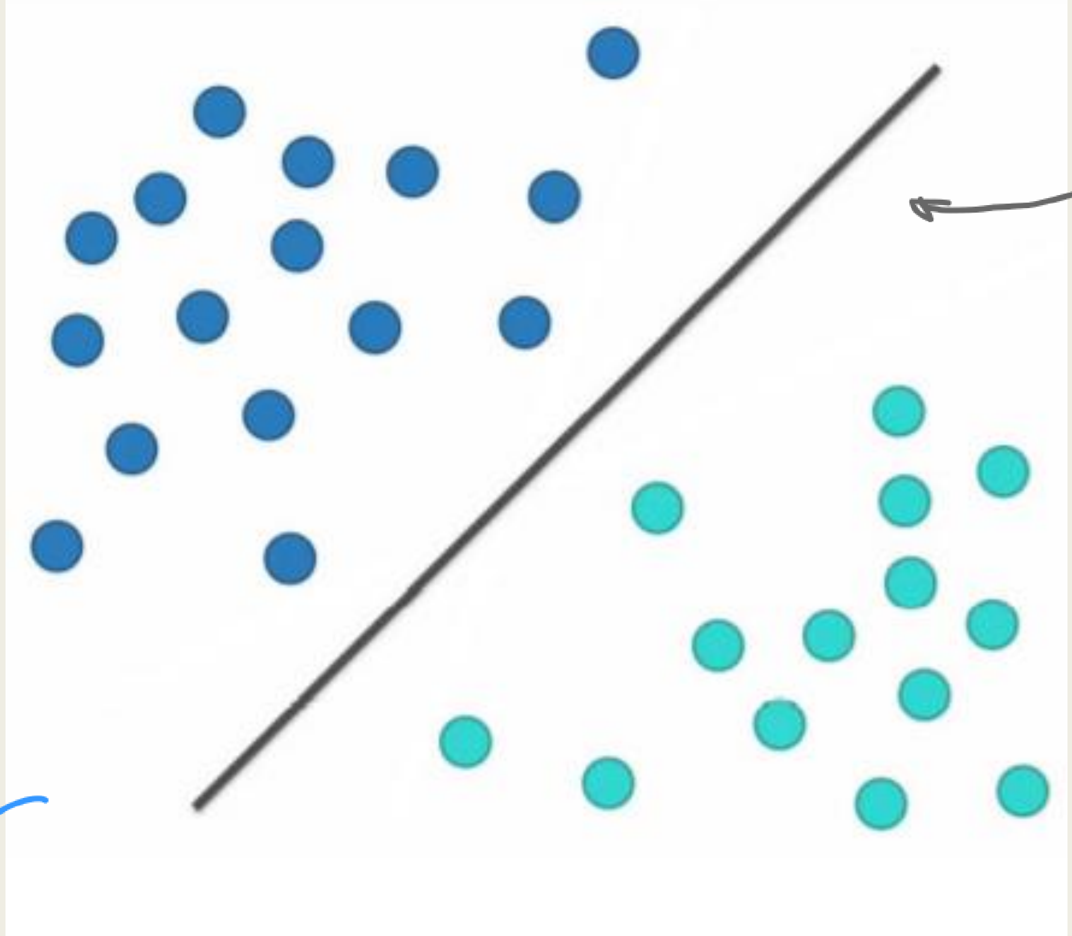
Splitting the data in the best possible way



- Which hyperplane is best?

Example

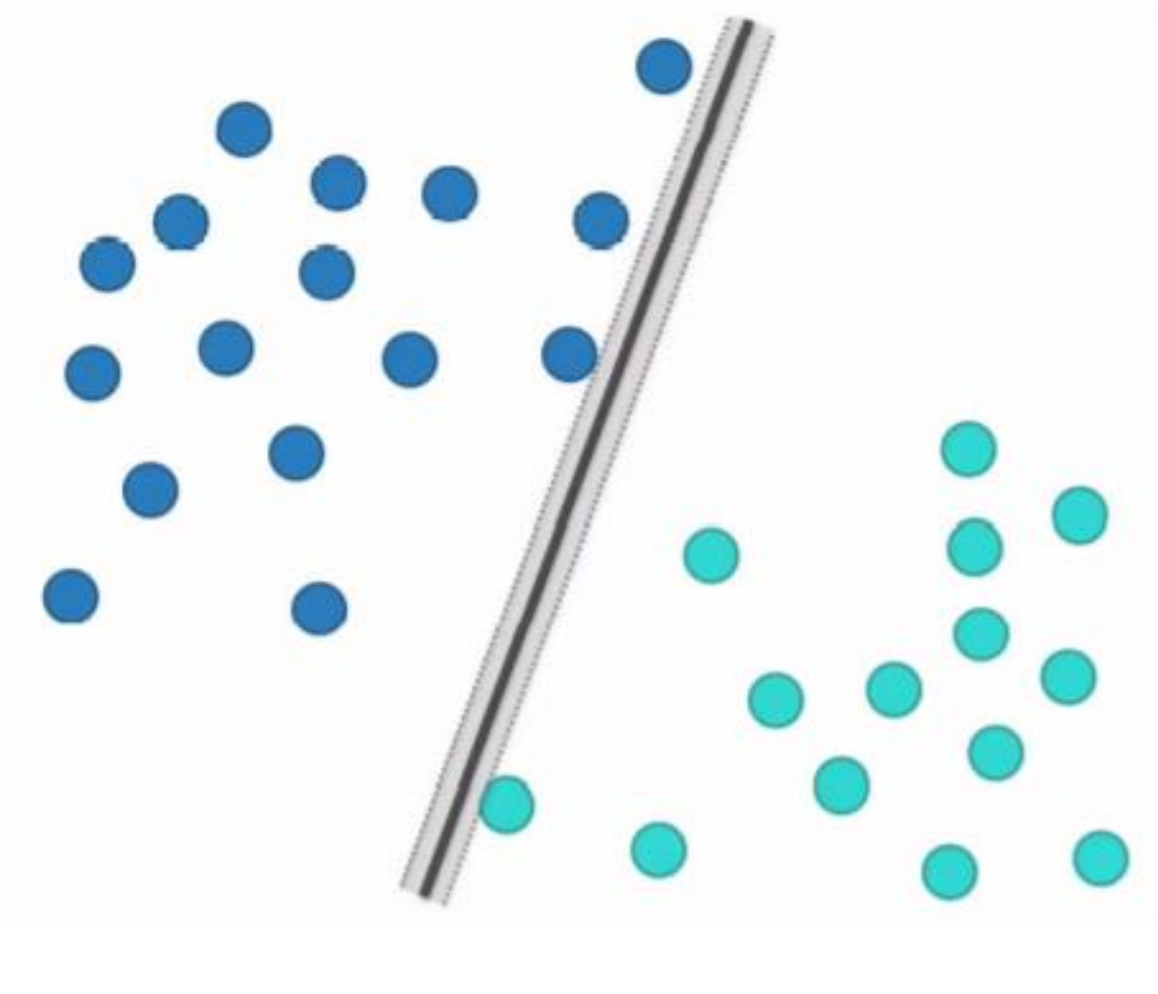
continue



- Split the data in the best possible way continue

adds a margine to the regression line in sum

Find the maximum margin



- Adjust the road with maximum width

optimizing sum

Basic idea of SVM

Find

- Optimal hyperplane for linearly separable patterns
- Extend to patterns that are not linearly separable, by transformations of original data to map into new space – Kernel function
- Unique features

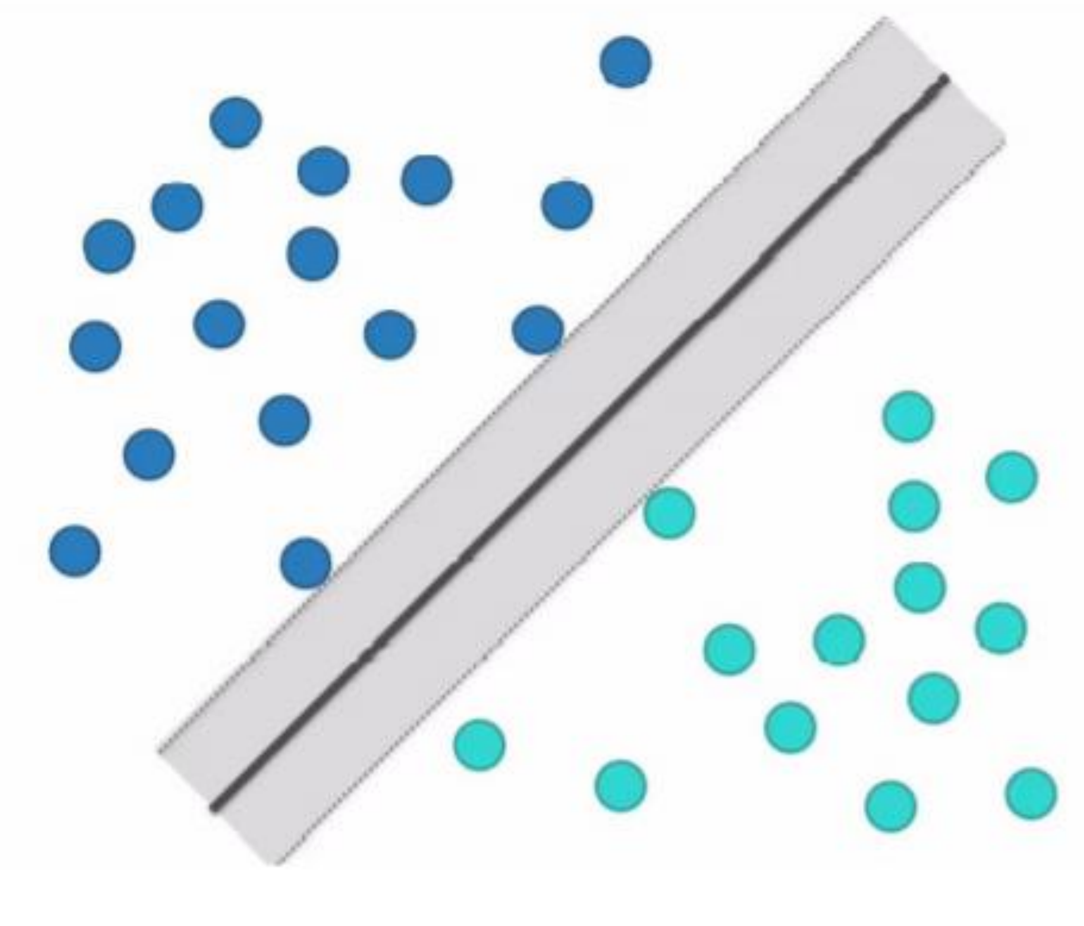
  It is explicitly based on a theoretical model of learning

  It is not affected by local minima

  no need to find local minima to find best fit like in gradient decent

  Do not suffer from the curse of dimensionality

  no limitation of number of features in dataset



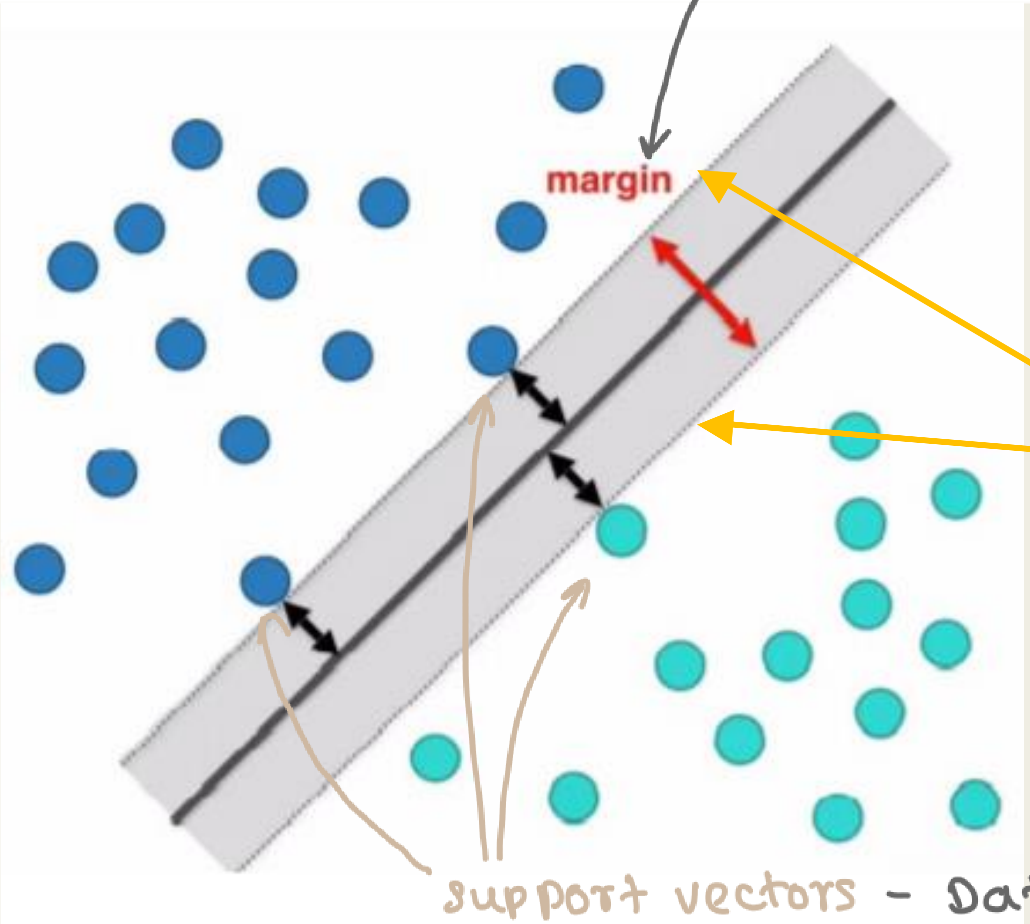
- Why is this the best split ?
- Adjust the road with maximum width

Important concepts in SVM

- **Support Vectors** – Datapoints that are closest to the hyperplane is called support vectors. Separating line will be defined with the help of these data points.
- **Hyperplane** – As we can see in the above diagram, it is a decision plane or space which is divided between a set of objects having different classes.
- **Margin** – It may be defined as the gap between two lines on the closet data points of different classes. It can be calculated as the perpendicular distance from the line to the support vectors. Large margin is considered as a good margin and small margin is considered as a bad margin.

[Support Vector Machine (SVM) (tutorialspoint.com)]

Distance from one edge to the other



- Adjust the road with maximum width
- The distance between the points and the line are as far as possible
- Why is this the best split?

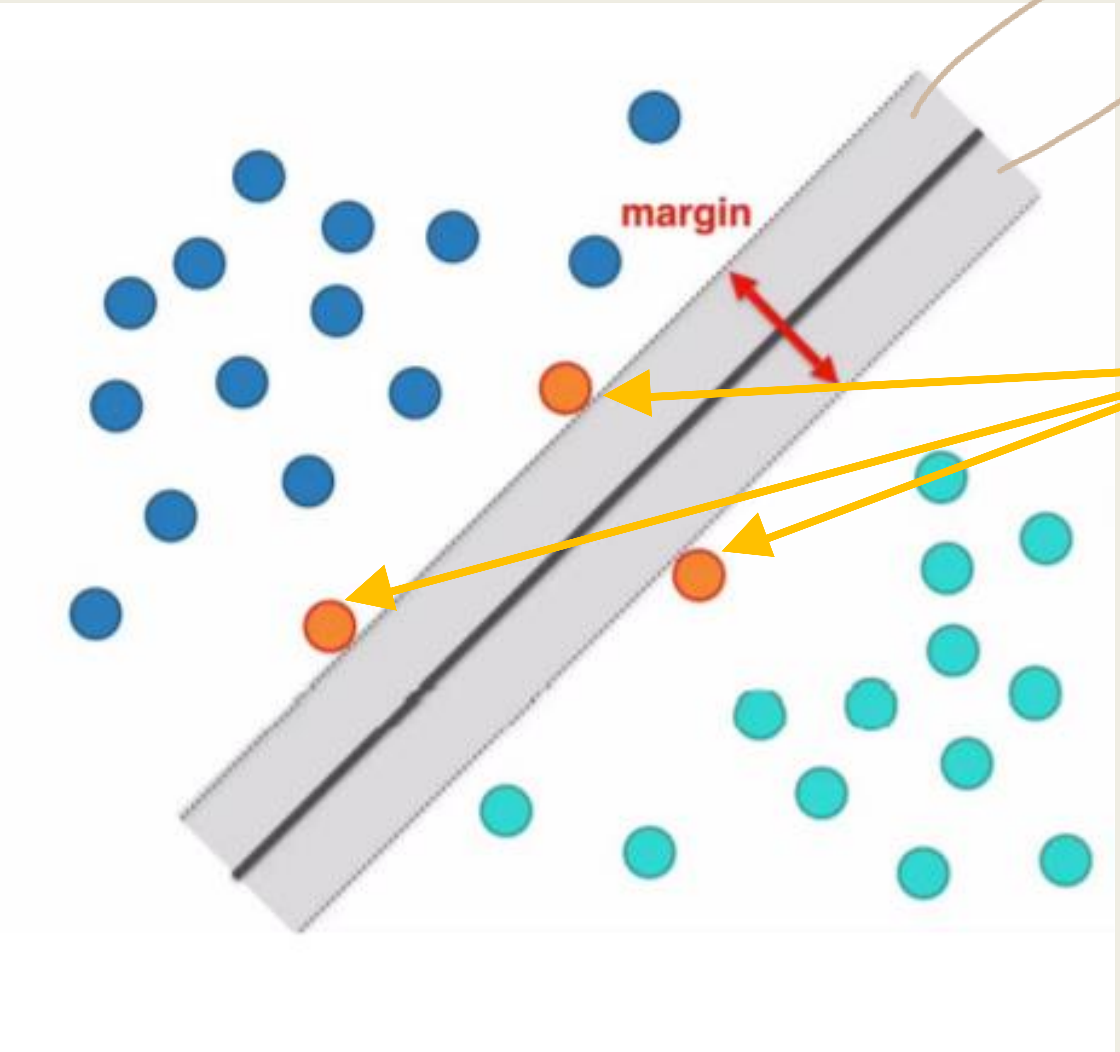
margin

The decision boundaries with large margins is tend to have a lower generalization error or overcome the overfitting when have small margins

support vectors - Data points which touch the margin

Support Vectors

same size



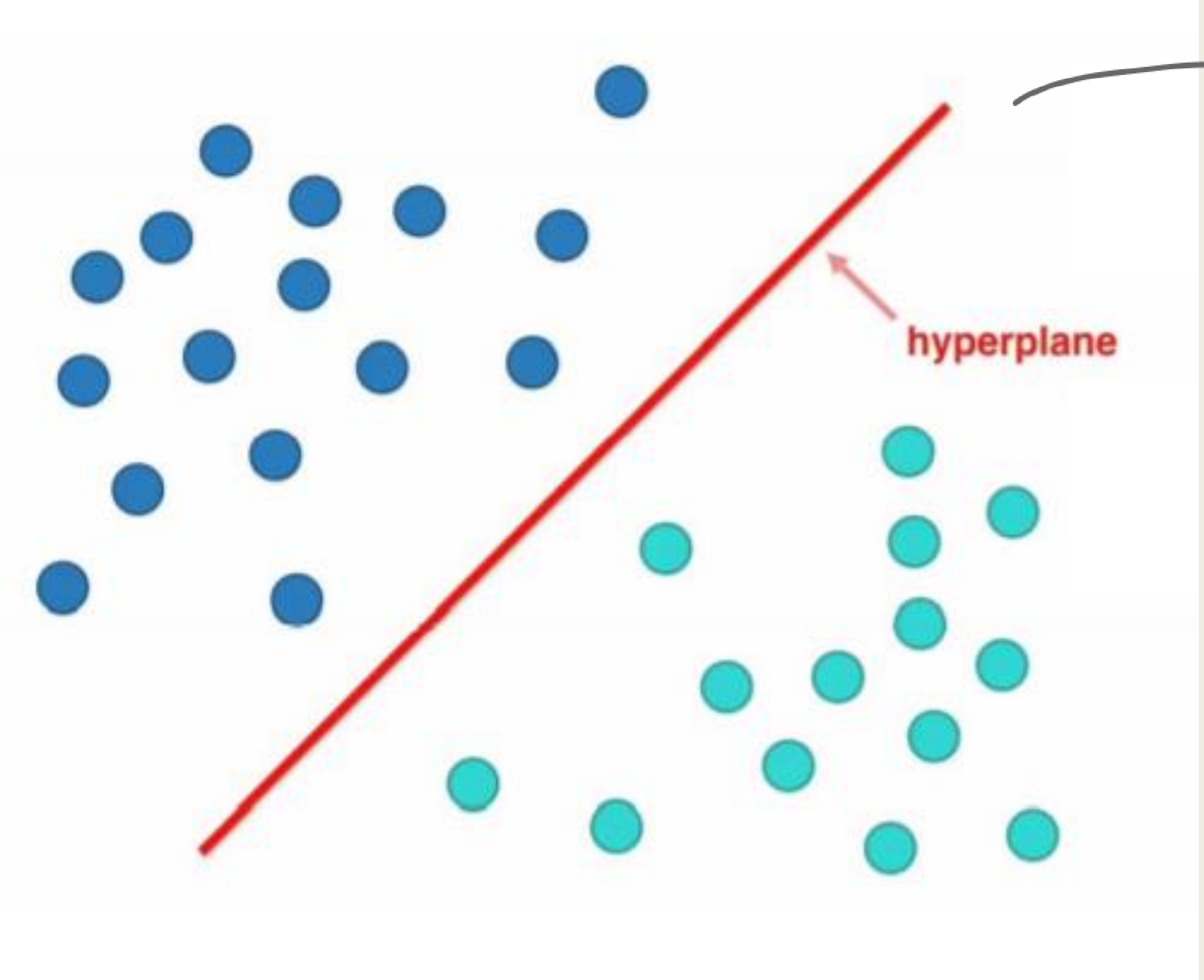
- Why is this the best split?

means again

The distance between the **support vectors** and the line are as for as possible

should be far as possible

Hyperplane



- Split the data in the best possible way
- This hyperplane best splits the data

In general ,lots of possible solutions i.e. lines for best splits the data.

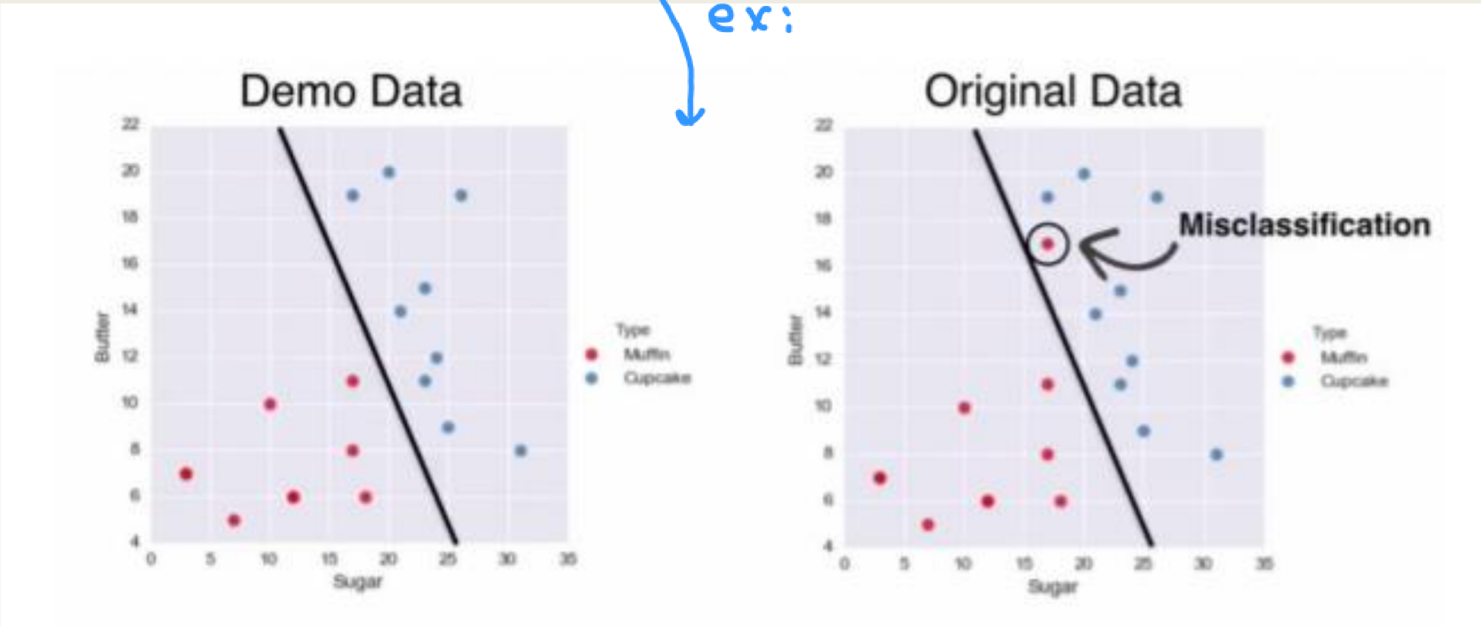
Two things in SVM

also two parameters

- kernels → transform data into nigher pane
- C parameter → adujus degree of overfitting

- Setting a larger margin
- Lowering misclassification rate (how much a model mis qualifies a data)

ex:



- Misclassification can be avoided with *C* parameter. The degree of overfitting will be decided by *C* parameter. Just check the next slide

Regularization

- The "C" parameter in Python's SkLearn Library
- Optimizes SVM classifier to avoid misclassifying the data
- C → large, Margin of hyperplane → small or soft margin
- C→ small, margin of hyperplane → large (possibility of misclassification is high)

not allow misclassification

C → large, chance of overfit

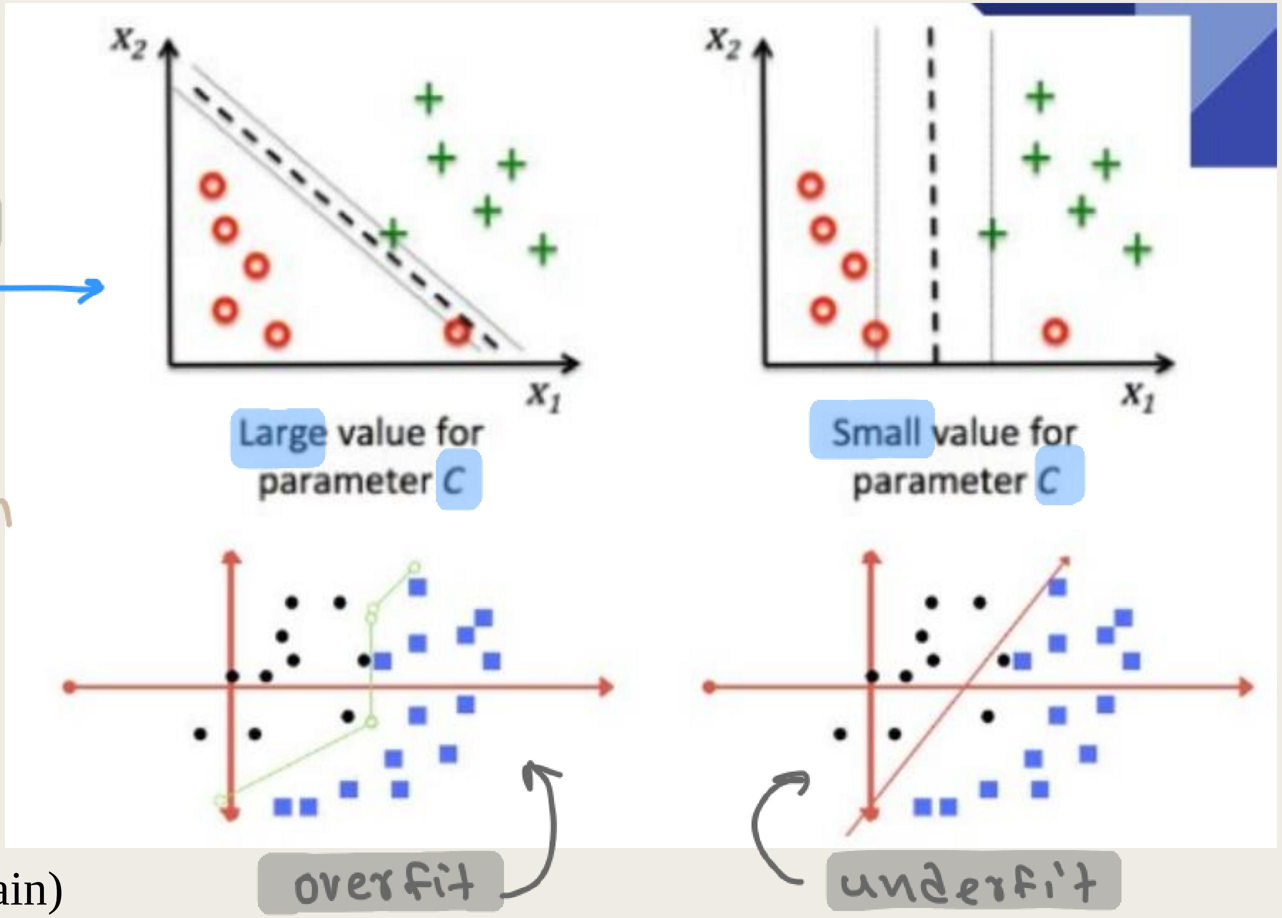
C → small, chance of underfitting

allow misclassification

```
clf = svm.SVC(kernel='linear', C=1).fit(X_train, y_train)
```

Please follow Chapter -05 code

Traing accurancy may higher but lower when predict

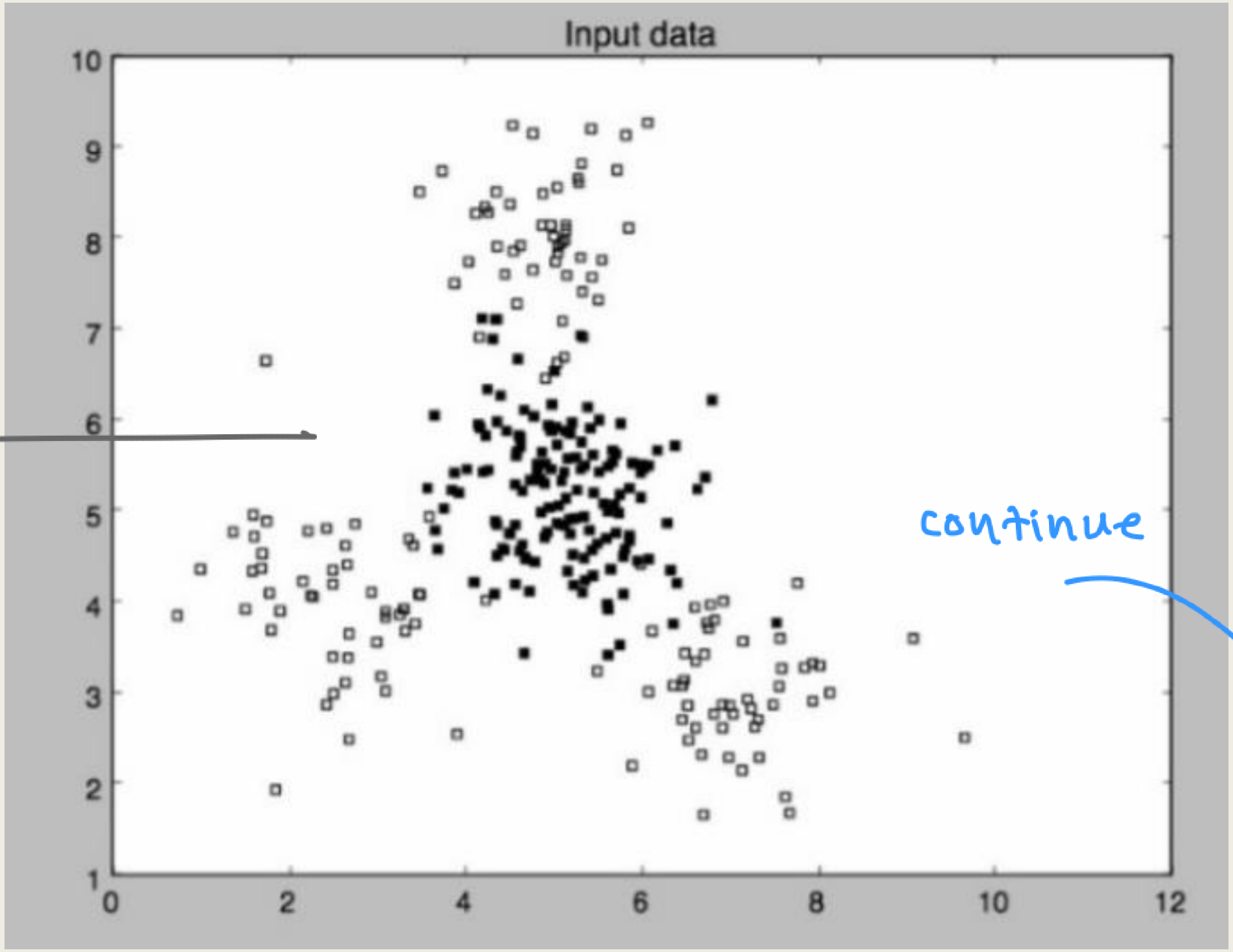


Source: https://pt.slideshare.net/JunaidKhan281/support-vector-machines-svm-118550627/15

Example

difficult to classify

use kernels to tranform into classifiable data



continue

- The preceding figure consists of two types of points – **solid squares** and empty squares.

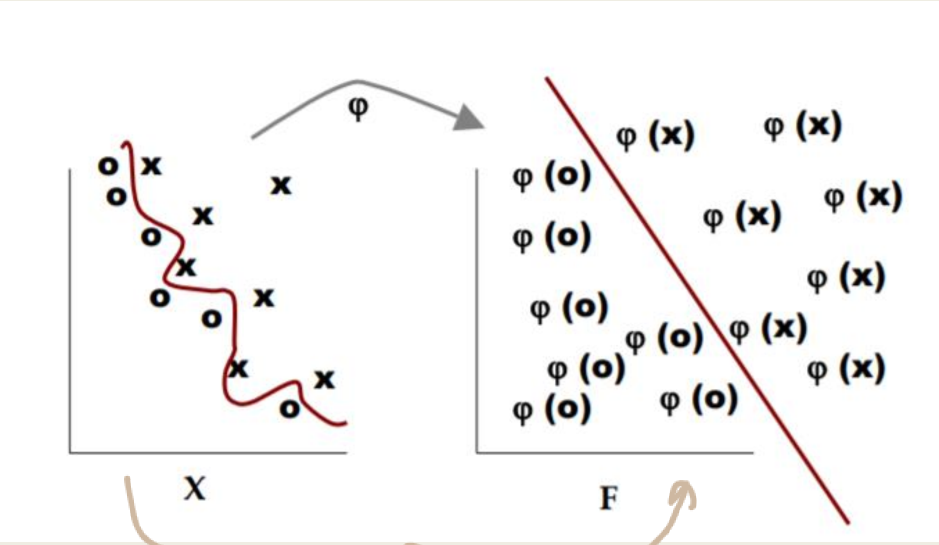
Kernels

tranform into higher dimention

- Mathematical functions for transforming data
- Using some linear algebra
- Different SVM algorithms use different types of kernel functions

In practice, SVM algorithm is implemented with kernel that transforms an input data space into the required form. SVM uses a technique called the kernel trick in which kernel takes a low dimensional input space and transforms it into a higher dimensional space. In simple words, kernel converts non-separable problems into separable problems by adding more dimensions to it.

- Various kernels available
  - *linear kernel*
  - *Non-linear kernel*
  - *Radial basis function (RBF)*
  - *Sigmoid*
  - *Polynomial*
  - *Exponential*



trasform

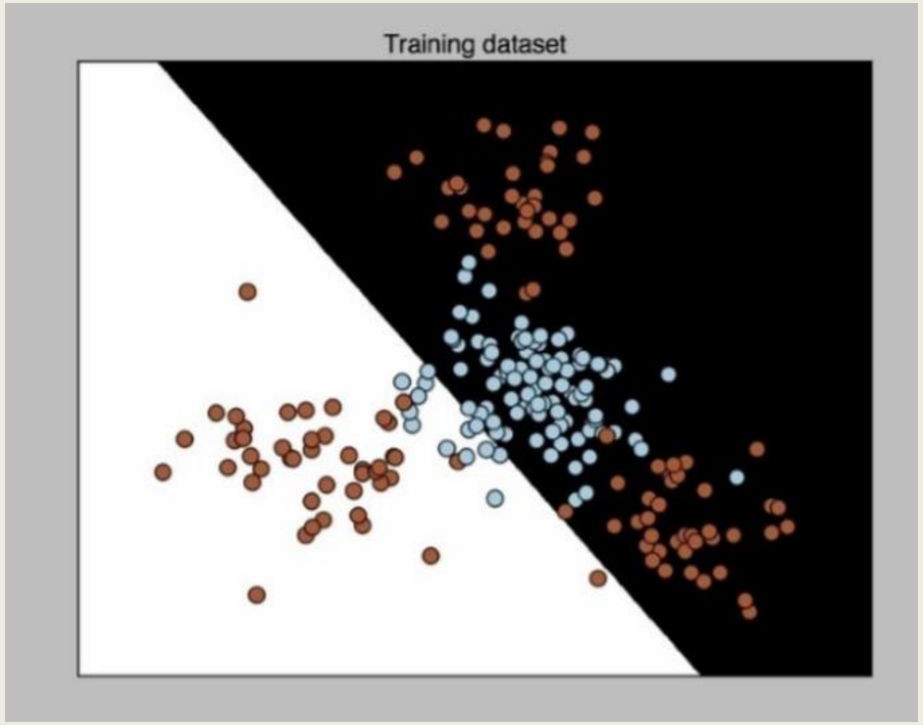
Codes ; kernel Trick 1

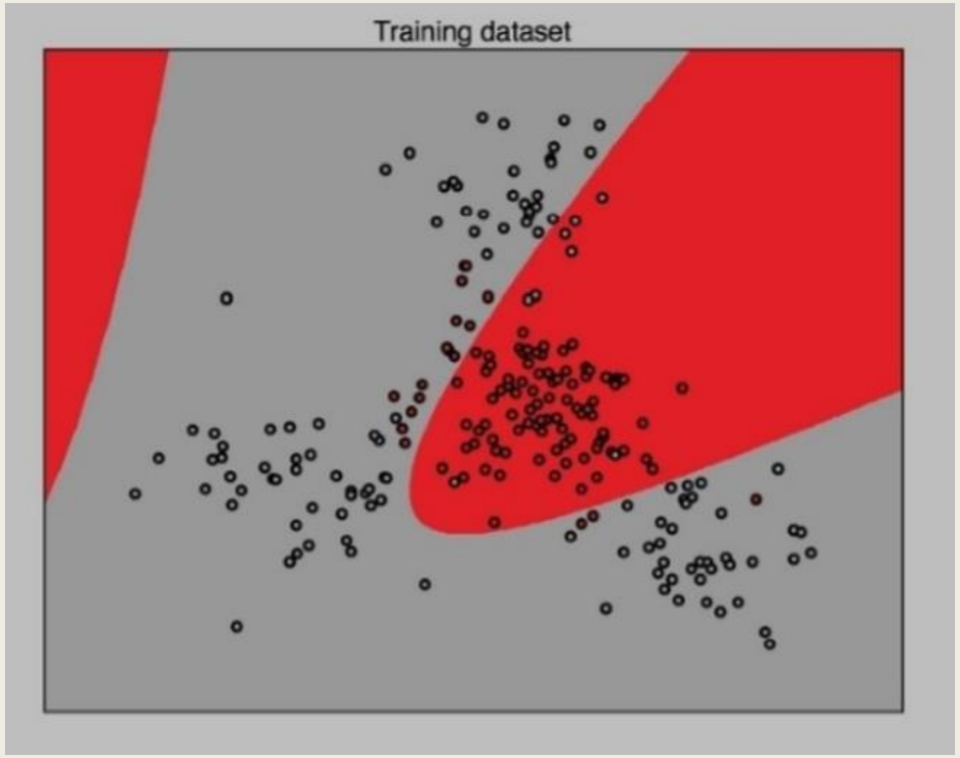
polynomial with degree

- params = {'kernel': 'linear'}
- classifier = SVC(**params)

Defaut

- params = {'kernel': '**poly**', 'degree': 3}
  classifier = SVC(**params)





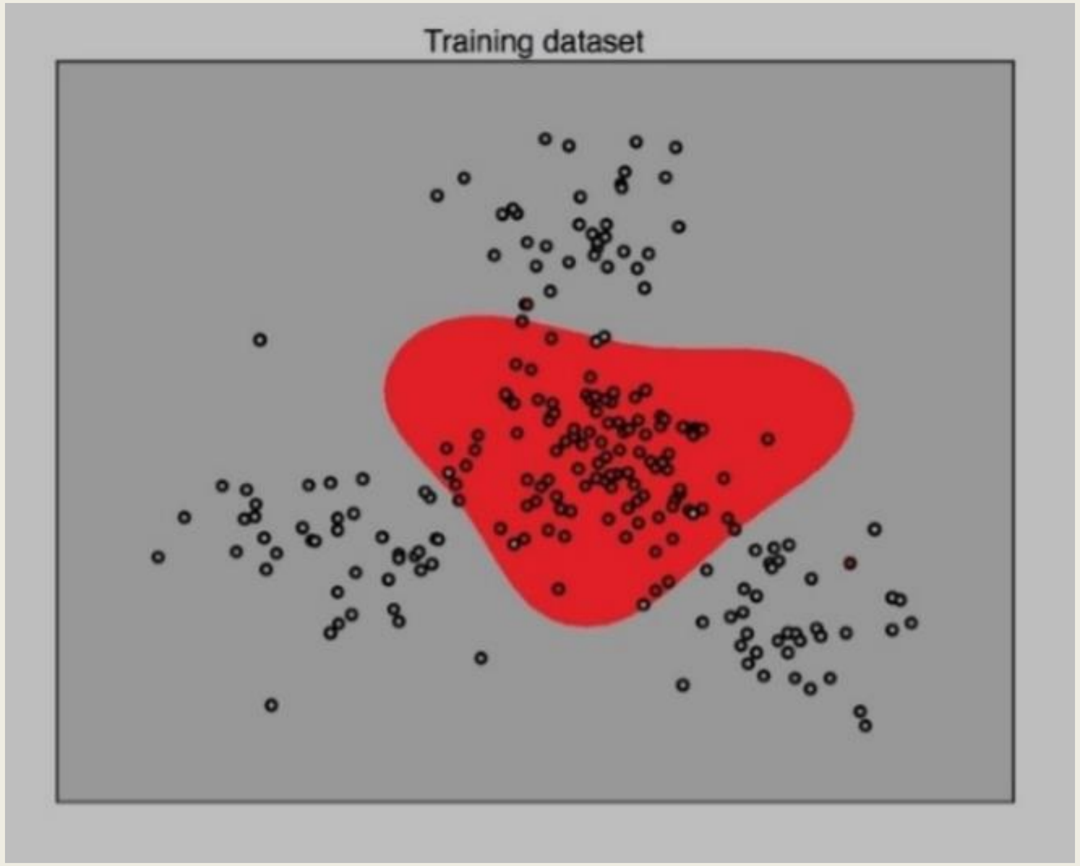
The SVM based classier is called the SVC (Support Vector Classifier) and used in classification problems.

Codes ; kernel Trick 2

- Radial basis function (RBF)

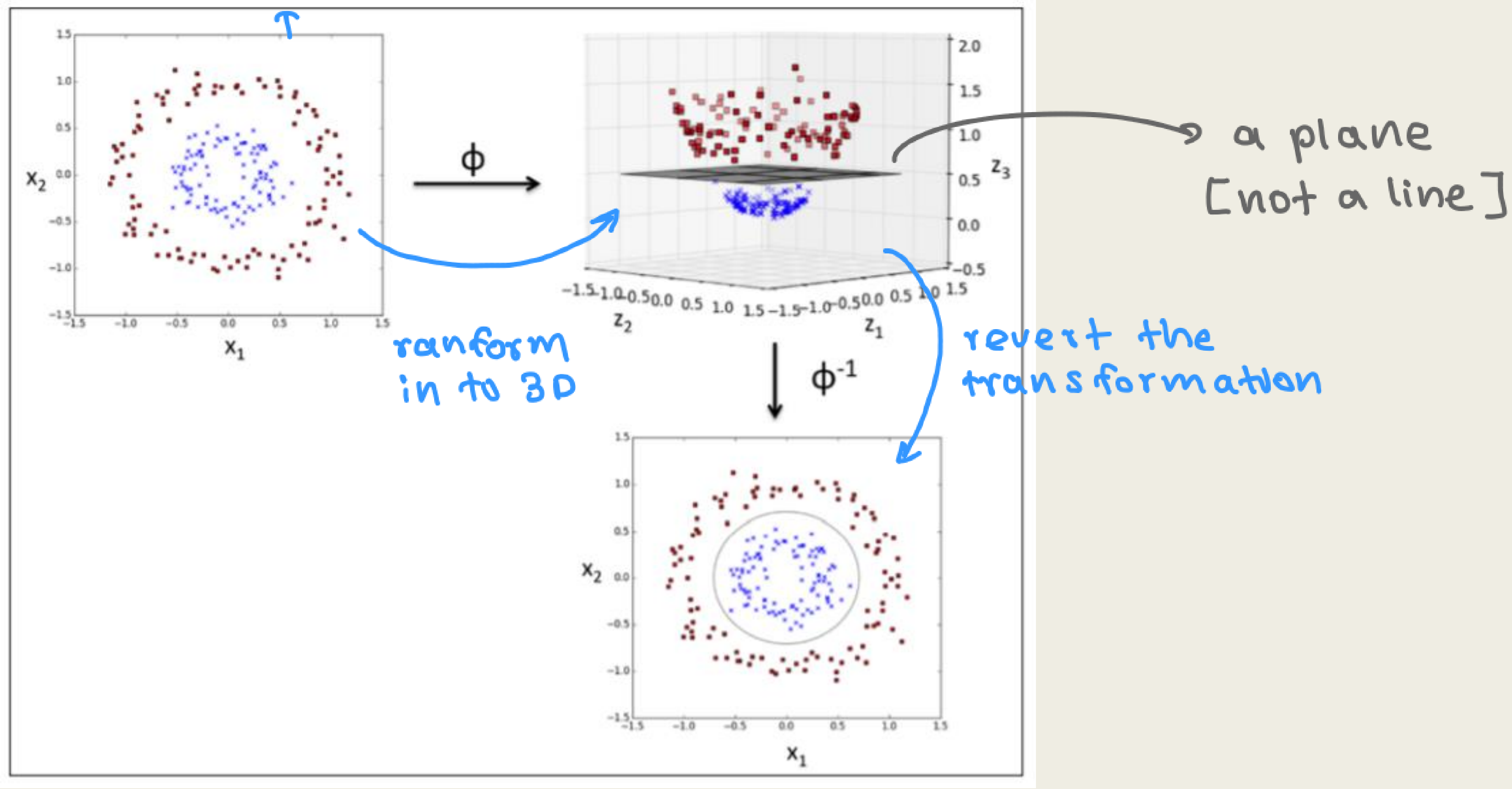
for splitting in the middle

```
params = {'kernel': 'rbf', 'probability': True}
classifier = SVC(**params)
```



Codes ; kernel Trick 3

- This allows us to separate the two classes shown in the plot via a linear hyperplane that becomes a nonlinear decision boundary if we project it back onto the original feature space:



Pros

- It works really well with clear margin of separation
- It is effective in high dimensional spaces
- It is effective in cases where number of dimensions is greater than the number of samples

  features
- It uses a subset of training points in the decision function (called support vectors), so it also memory efficient

Cons:

- It doesn't perform well, when we have large data set because the required training time is higher
- It also doesn't perform very well, when the data set has more noise i.e. target classes are overlapping

Application

- Face detection
- Text and hypertext categorization
- Classification of images
- Bioinformatics
- Handwriting recognition
- Protein fold and remote homology detection
- Generalized predictive control (GPC)

Predictive modeling

extract useful
info from large dataset

- Concerning data mining to forecast future trends. used for
- Predictive modeling is an analysis technique that is used to predict the future behavior of a system