

Machine Learning

Prepared by Dr. Yasser Abdelhamid

DS 620
Machine Learning

Lecture 9





OUTLINE

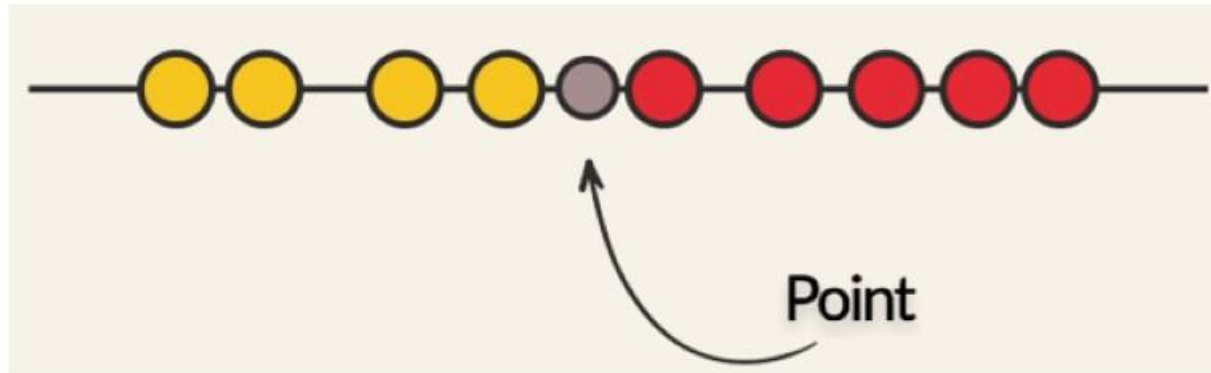
- Support vector machines (SVM)

SUPPORT VECTOR MACHINES(SVM)

- ❖ Support Vector Machines are widely used and powerful machine learning algorithms
- ❖ It is commonly used for
 - Classification and
 - Regression tasks.

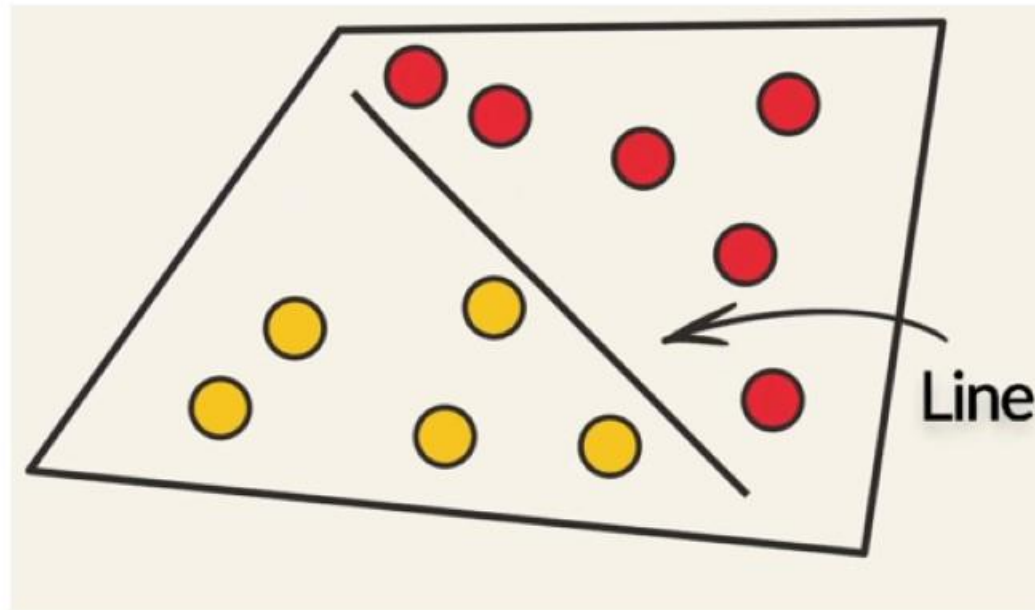
HYPERPLANE IN ONE DIMENSIONAL SPACE

❖ In one dimensional space a hyperplane is a **point**.



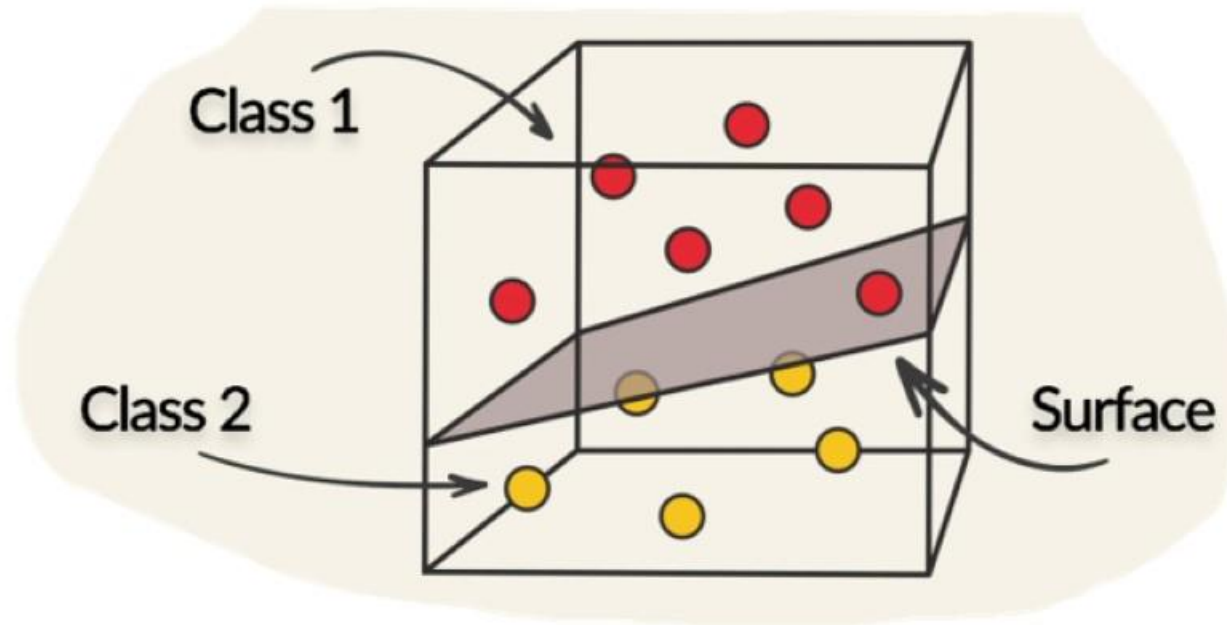
HYPERPLANE IN TWO DIMENSIONAL SPACE

❖ In two dimensional space a hyperplane is a **line**.



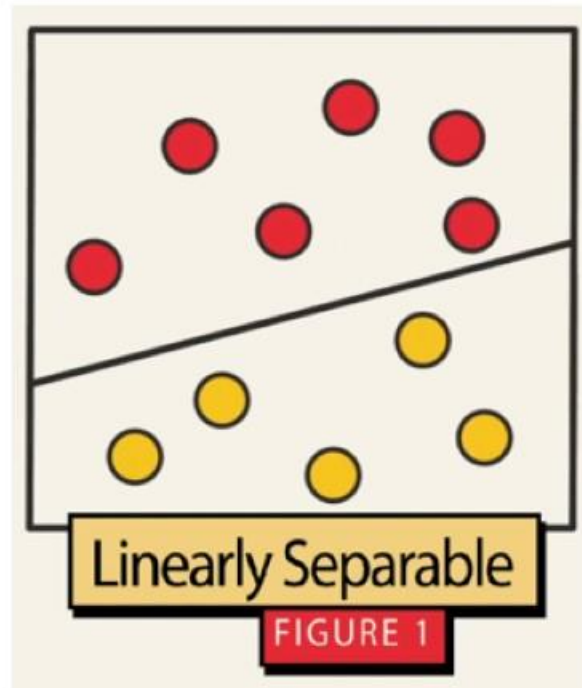
HYPERPLANE IN THREE DIMENSIONAL SPACE

- ❖ In three dimensional space the hyperplane is a **Plane/surface**.



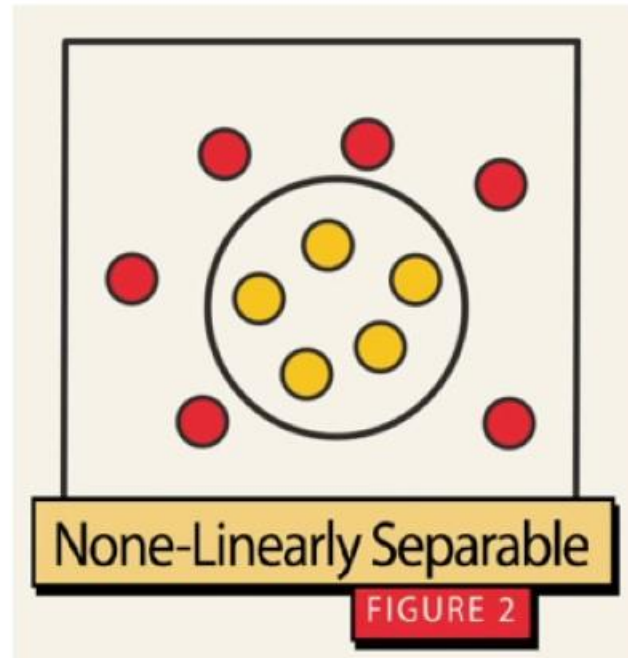
LINEARLY SEPARABLE DATA

❖ We can find a separating line between classes.



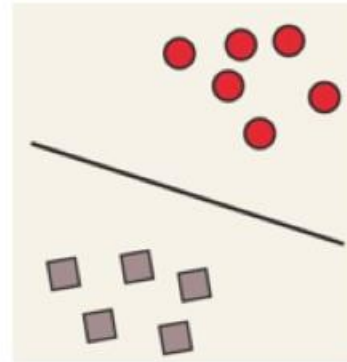
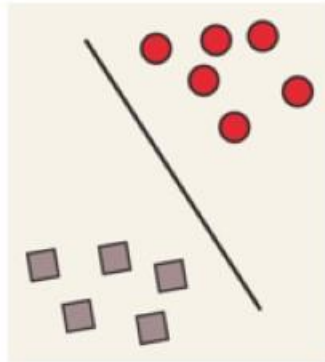
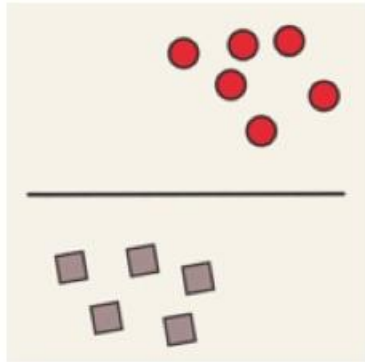
NON-LINEARLY SEPARABLE DATA

❖ We can not find a separating line between classes.



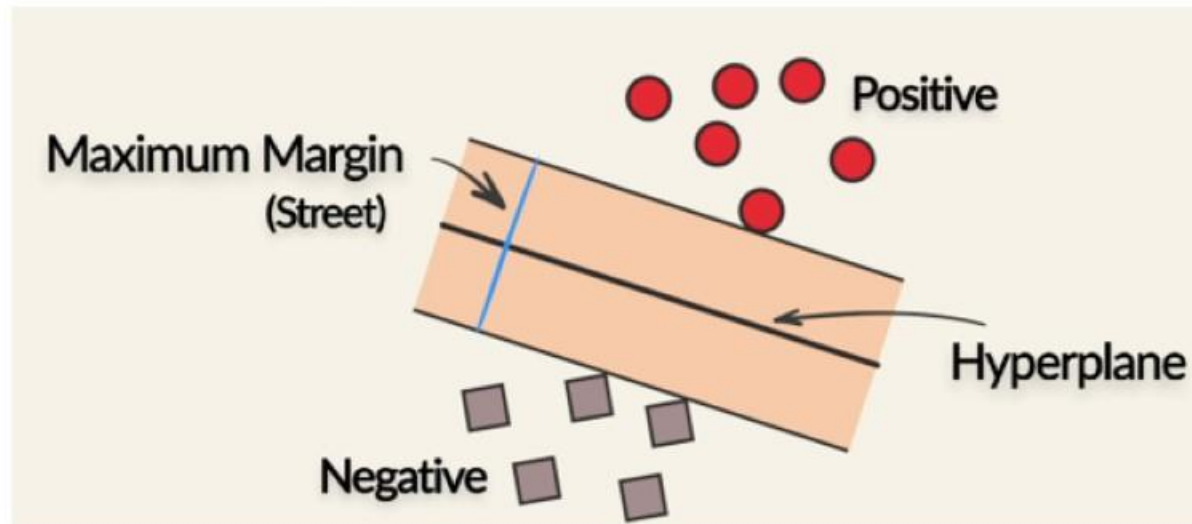
FINDING THE BEST LINE SEPARATOR

❖ Which of the following line separators gives the best classification results?



MAXIMUM MARGIN CLASSIFIER

- ❖ SVM can help us to choose the best hyperplane that maximizes the margin between classes. (Maximum Margin Classifier)



SUPPORT VECTOR MACHINES(SVM)

- ❖ Support Vector Machines are a type of **supervised learning algorithm** that works by finding an optimal hyperplane in a high-dimensional feature space to separate different classes of data points.

SUPPORT VECTOR MACHINES(SVM)

- ❖ The hyperplane is chosen in such a way that it **maximizes the margin**, i.e., the distance between the hyperplane and the nearest data points of each class.
- ❖ The data points that are closest to the hyperplane are called support vectors.

SUPPORT VECTOR MACHINES(SVM)

- ❖ By maximizing the margin,
 - We can achieve better generalization on unseen data, and
 - Improve the algorithm's ability to handle noise and outliers.

LINEAR SVM

- ❖ In linear SVM, the goal is to find the best hyperplane that separates the data points of different classes.
- ❖ The hyperplane can be represented by the equation:
$$w \cdot x + b = 0,$$
- ❖ where w represents the weights vector, x is the input feature vector, and b is the bias term.

LINEAR SVM

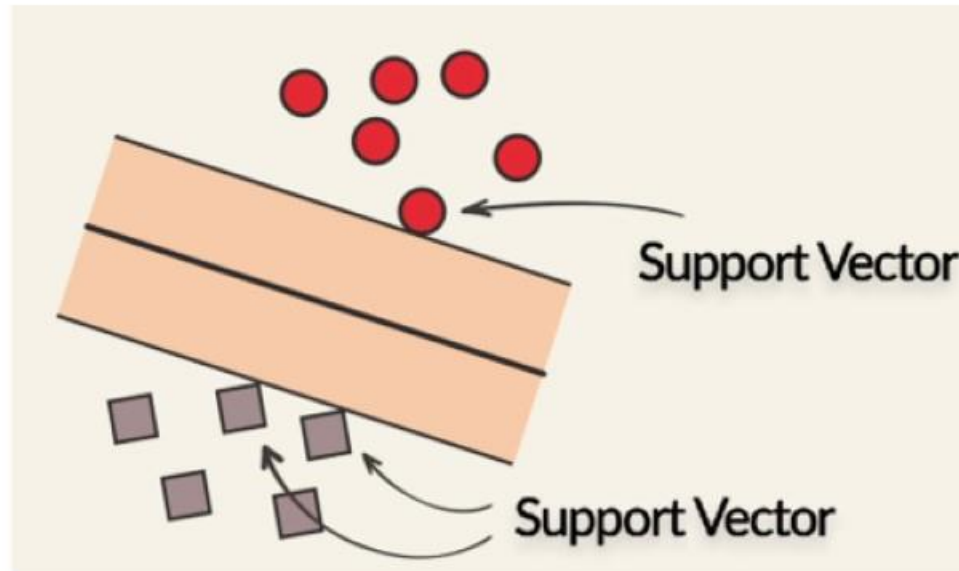
- ❖ The objective of SVM is to find the optimal $\mathbf{w}$ and $\mathbf{b}$ values that maximize the margin while correctly classifying the training data.
- ❖ To achieve this, we formulate an optimization problem that minimizes the objective function subject to certain constraints.

LINEAR SVM

- ❖ To handle non-linearly separable data, we can introduce a slack variable ξ that allows for some misclassification of training samples.
- ❖ This leads to the formulation of the soft-margin SVM, where the objective function includes a penalty term for misclassifications.

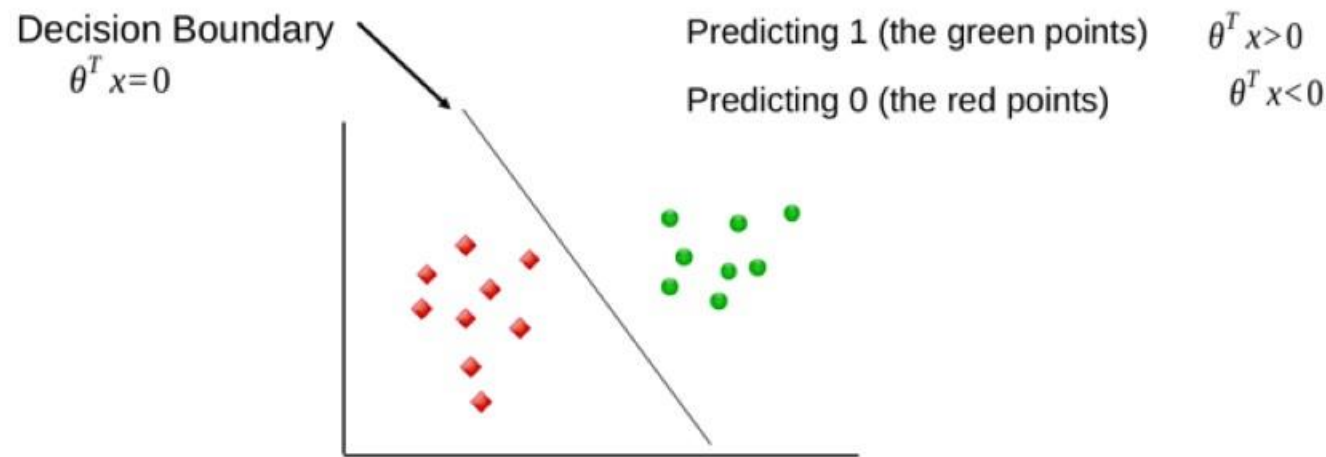
SUPPORT VECTORS

- ❖ Support vectors are the data points that lie closest to the hyperplane. (The most difficult data points to classify)



CLASSIFICATION

- ❖ Binary Classification can be view as the task of separating classes in feature spaces.
- ❖ For a hypothesis $h_{\theta}(x) = \theta^T x$



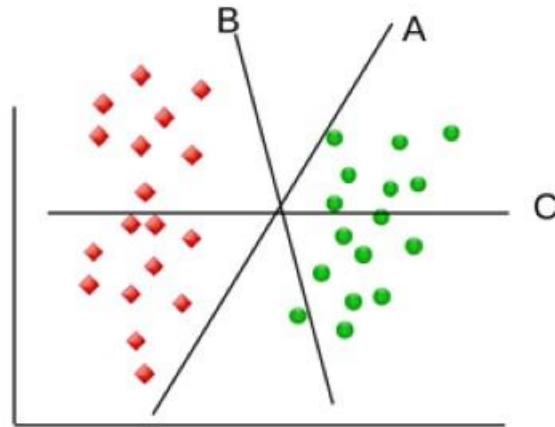
- ❖ The goal of Classifier is to train a model that assigns a new **unseen object** into a specific category.
- ❖ It places the object **above or below the separator hyperplane**

CLASSIFICATION

- ❖ Classification is the process of categorizing the two classes with a hyperplane.
- ❖ How can we Identify the (best) hyperplane

Scenario 1

Identify the right
hyper-plane

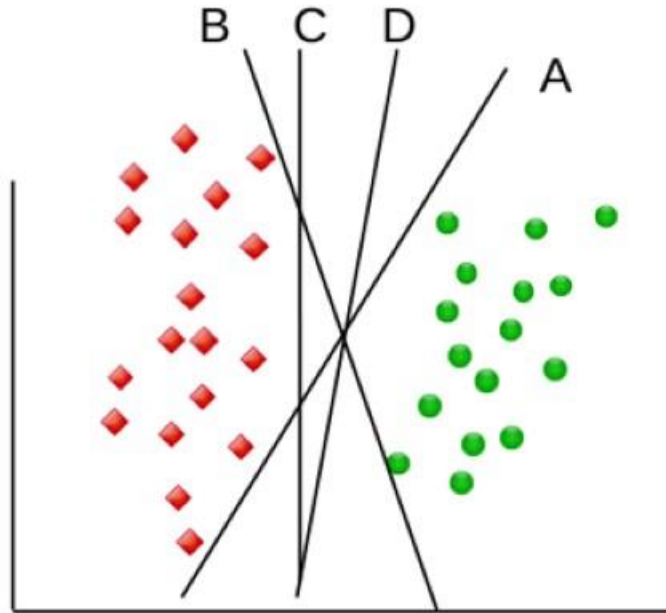


MULTIPLE CORRECT HYPERPLANES

❖ What if we have more than one hyperplane?

Scenario 2

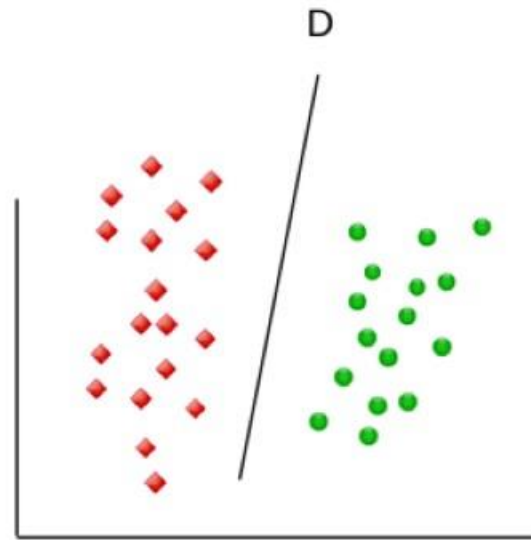
Identify the right hyper-plane



We have 4 hyper-planes A, B,C, and D. All are separating the classes well

MAXIMUM MARGIN CLASSIFIER (MMC)

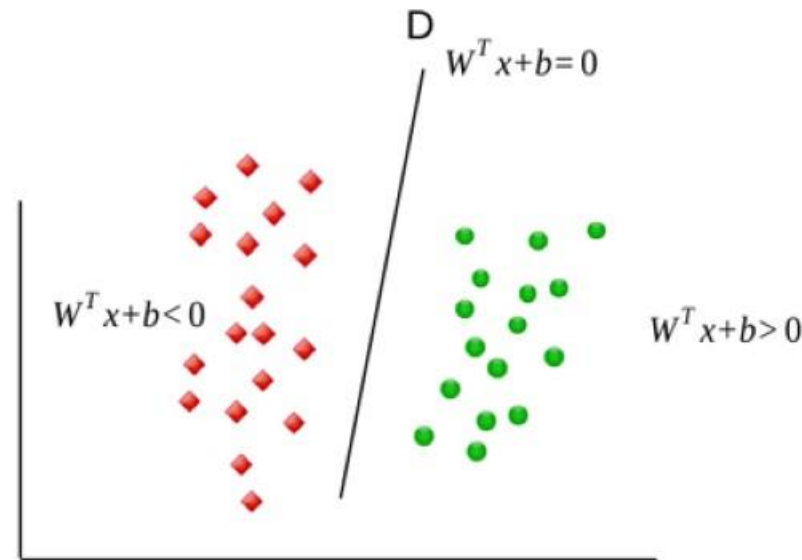
- ❖ D stays as far as possible from both classes.
- ❖ By being **right in the middle of the two classes**, it is less “risky” to misclassify the unseen data, thus generalizes well on test data



MAXIMUM MARGIN CLASSIFIER (MMC)

❖ MMC tries to find the best separator hyperplane. Here is a simple version of What MMC (as a kind of SVMs) do:

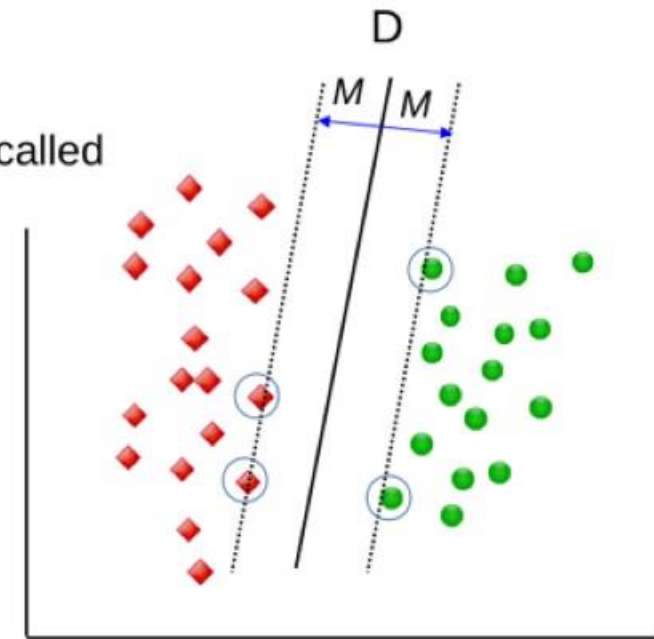
1. Find hyperplanes that correctly classify the training data.
2. Among all such hyperplanes, pick the one that has the greatest distance to the closest points.



MAXIMUM MARGIN CLASSIFIER (MMC)

- ❖ The closest points that identify this hyper-plane are known as **support vectors**.
- ❖ The region around the line is known as the Margin.
 - Only support vectors matter, other training examples are ignored.
 - Data that can be separated by a hyper-plane is known as **Linearly separable** data
 - The hyperplane that classifies the linear separable data acts as a linear classifier.

- $2M$ is the margin
- Circled points are called support vectors

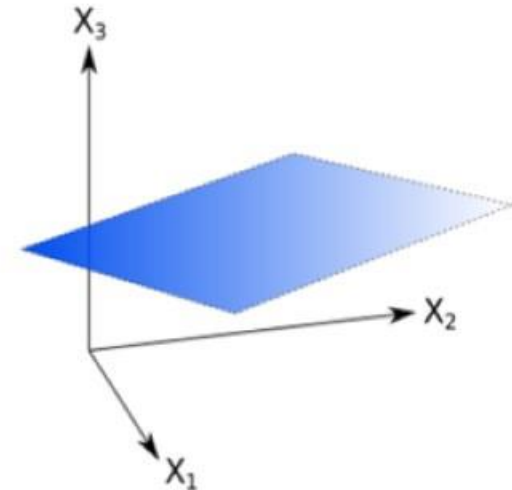
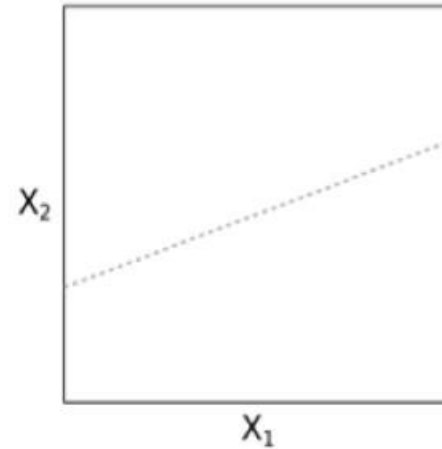


SUPPORT VECTOR CLASSIFIER

- ❖ SVC is a classifier formally defined by a separating hyperplane $W^T X + b = 0$
- ❖ A hyperplane is a subspace of one dimension less than its ambient space.
- ❖ This means a hyperplane of two dimension space is one dimension separator (line).
- ❖ A hyperplane of three dimension space is two dimension separator (plane).

SUPPORT VECTOR CLASSIFIER

- ❖ Elements above the hyperplane satisfy $w^T x + b > 0$
- ❖ Elements below the hyperplane satisfy $w^T x + b < 0$
- ❖ The weight vector W represents the orientation of the hyperplane and b represent the bias.

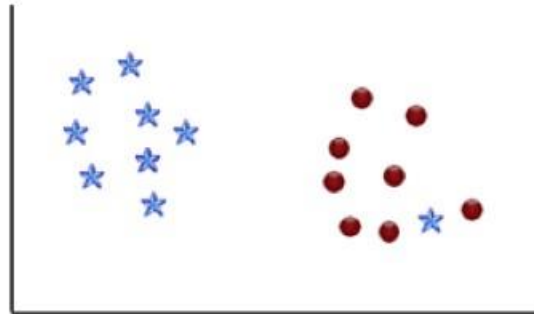


ALLOWING FOR ERRORS

- ❖ Look at the case indicated in scenario 3,
- ❖ How can we define the class separator?

Scenario 3

How can you find the right hyperplane ?

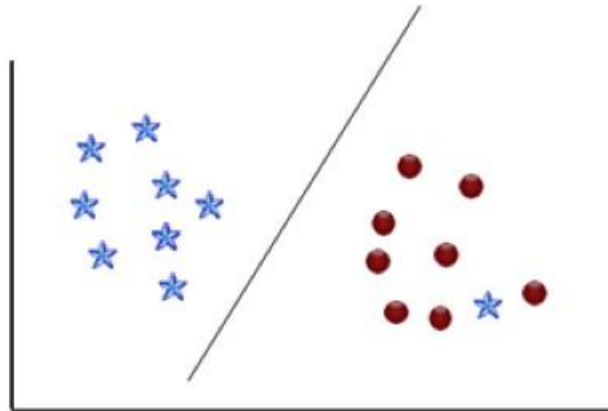


ALLOWING FOR ERRORS

- ❖ SVC has a feature to ignore outliers and find the hyperplane that has maximum margin.
- ❖ i.e SVC (hence SVM) is robust against outliers

Scenario 3

How can you find the right hyperplane ?

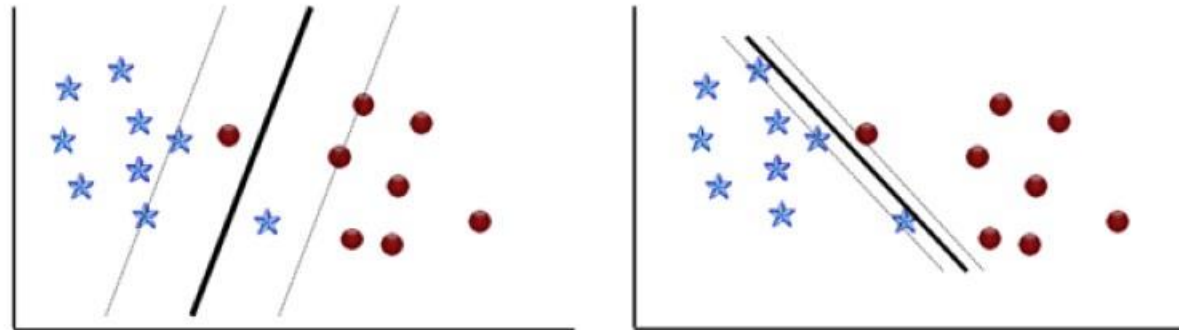


SOFT MARGIN CLASSIFIER

- ❖ For scenario 4, which separator is better.
- ❖ Will you maximize the margin and allow misclassified instance?
- ❖ Or
- ❖ Will you choose to correctly classify with narrow margin ?

Scenario 4

Which one is the best ?



Soft margin Classifier

ALLOWING FOR ERRORS

- ❖ How SVC let you handle this situation ?
 - It allows you to specify how many errors you are willing to accept.
- ❖ Providing a hyper parameter called ' C ' to your SVM.
- ❖ This allows to control the trade-off between:
 - A wide Margin
 - Correctly classify the training data

ALLOWING FOR ERRORS

- ❖ C is a non-negative “Tuning” parameter. If $C=0$, implies that no violation of the margin is possible
- ❖ (in this case, we have MMC situation)
- ❖ In addition to C , The SVC introduces a parameter ϵ_i called slack variable to each data point x_i
- ❖ It allows the data points to be on the wrong side of the margin or hyperplane.

SLACK VARIABLE

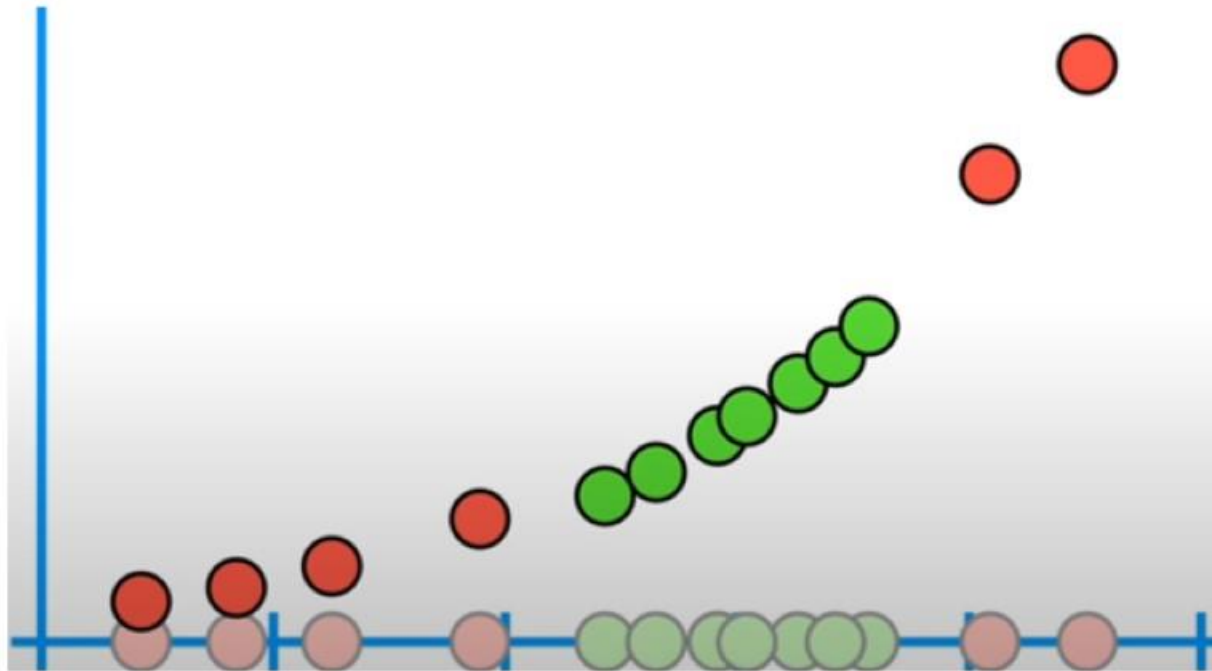
- ❖ If $\varepsilon_i = 0$, it states that the training point x_i is on the correct side of the margin,
- ❖ For $\varepsilon_i > 0$ that means x_i is on the wrong side of the margin.
- ❖ For $\varepsilon_i > 1$ means x_i is on the wrong side of the hyperplane

NON-LINEARLY SEPARABLE DATA

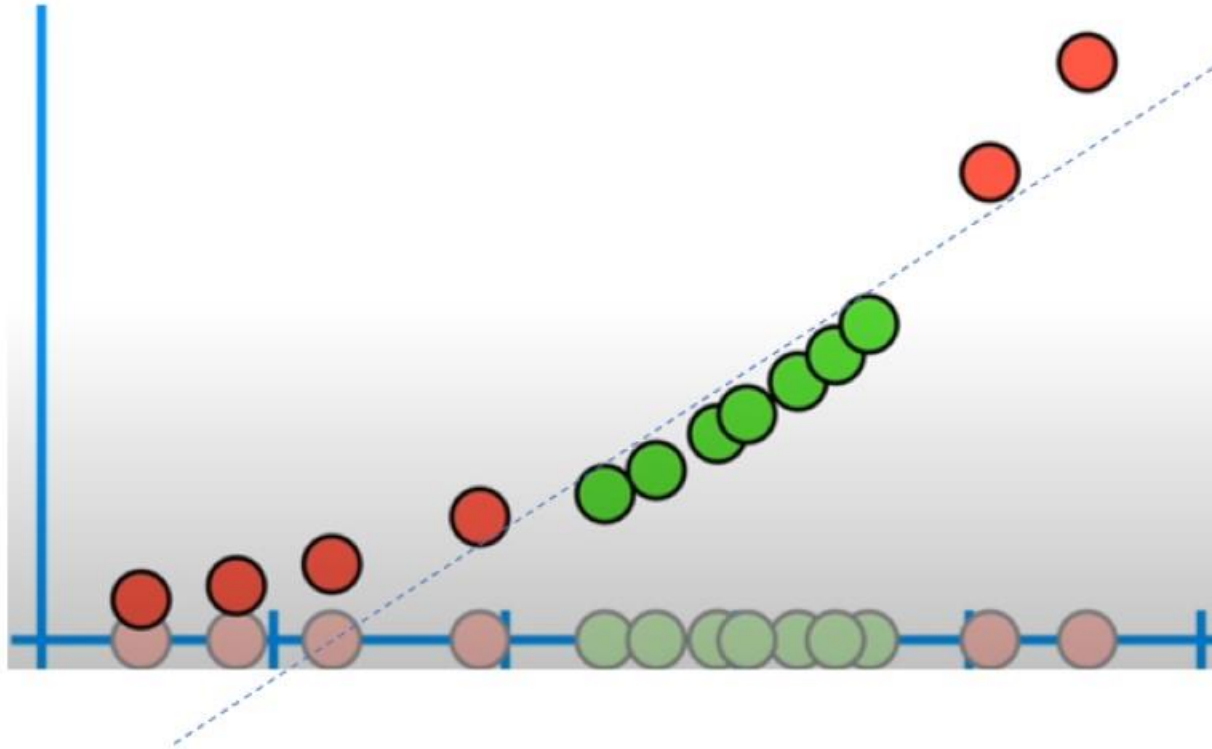
❖ How can we deal with data set like this?



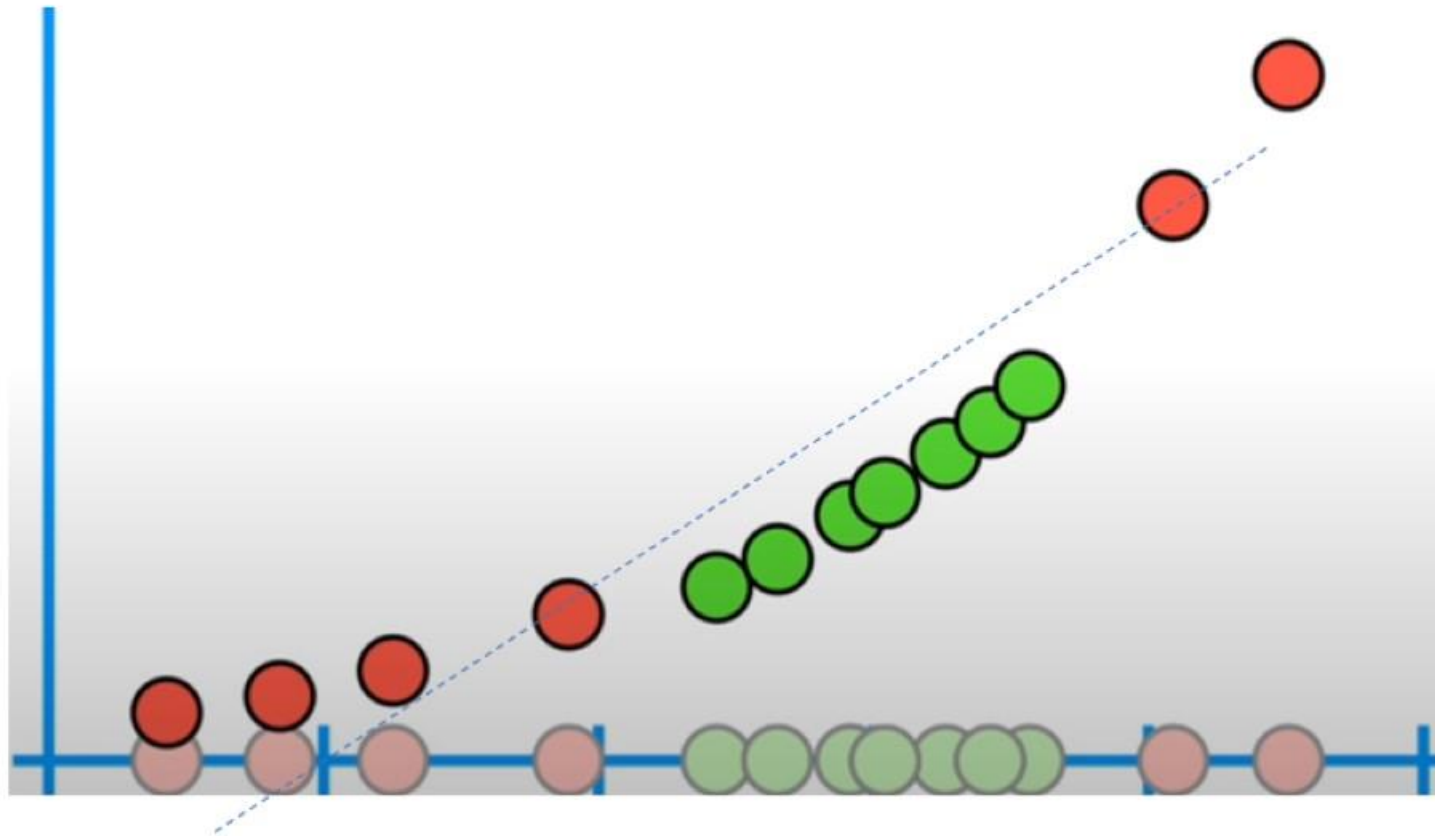
MOVING DATA TO HIGHER DIMENSION



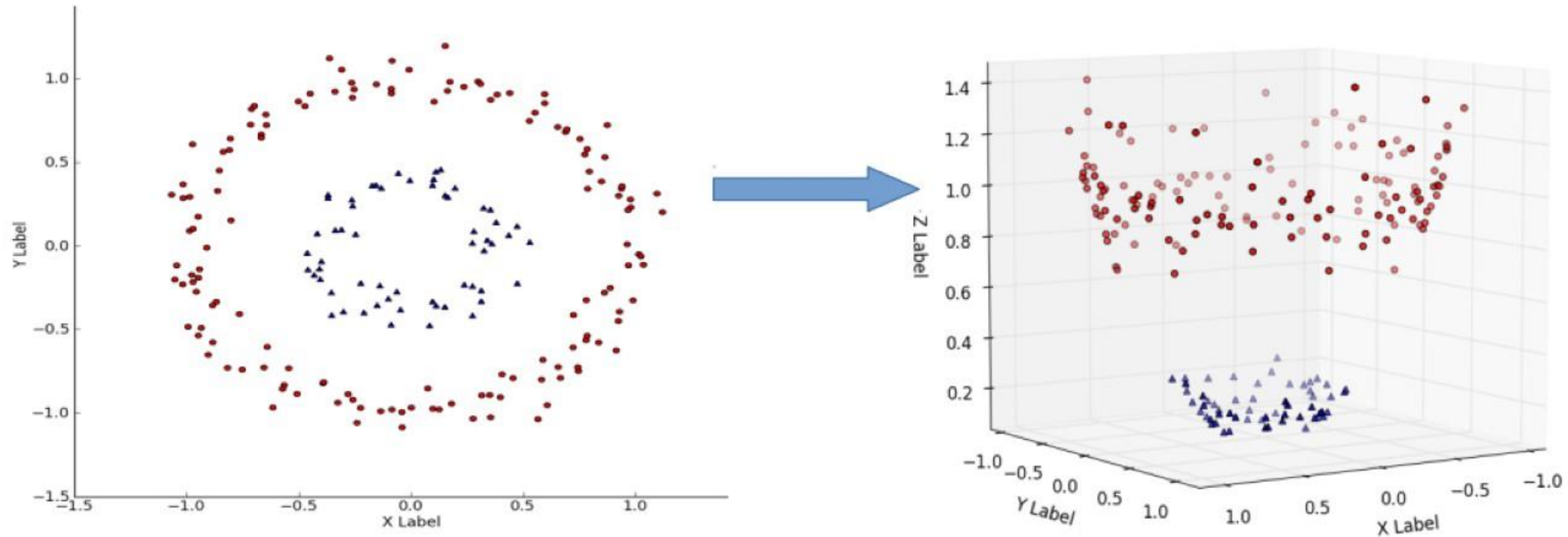
FIND A SUPPORT VECTOR CLASSIFIER
FOR THE DATA IN THE NEW DIMENSION



USE THE CLASSIFIER TO CLASSIFY NEW UNSEEN EXAMPLES

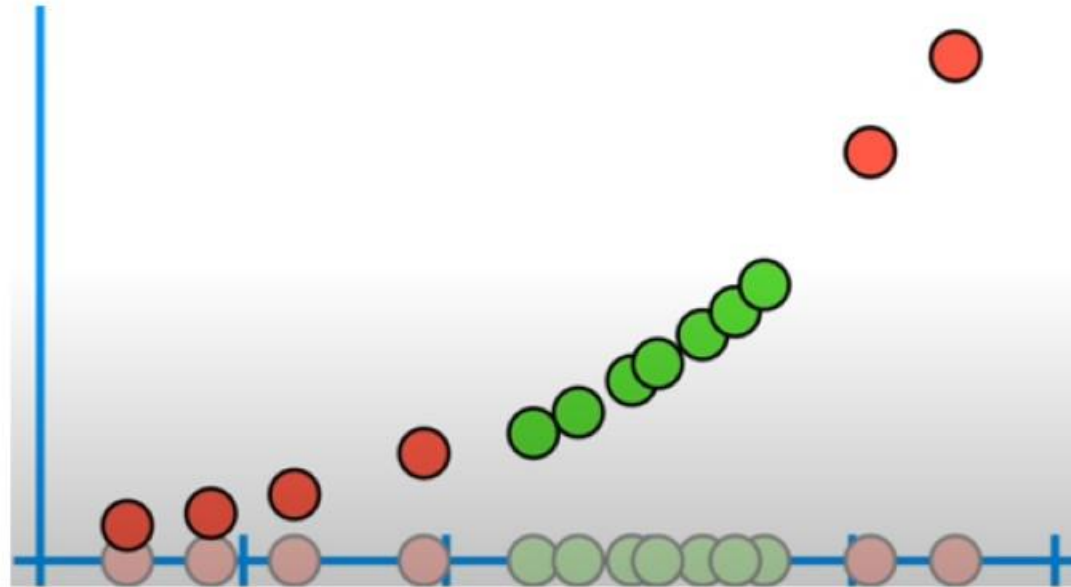


TRANSFORMATION FROM 2D TO 3D



HOW CAN WE TRANSFORM THE DATA

- ❖ Support vector machines use Kernel functions to systematically find support vector classifiers in higher dimensions



KERNEL FUNCTIONS

Some Popular Kernels

 **Polynomial Kernel** $K(x,y) = (x^T y + 1)^d$

 **Radial Base Function Kernel** $K(x,y) = \exp(-\gamma \|x - y\|^2)$

 **Linear Kernel** $K(x,y) = x^T y$

WHEN TO USE SVM?

❖ Binary Classification

- SVMs excel in binary classification problems, where the task is to classify data into two distinct classes.
- SVMs aim to find an optimal hyperplane that separates the classes by maximizing the margin between the closest data points (support vectors) from each class.
- They can handle both linearly separable and non-linearly separable data by using different kernel functions.

DATA WITH CLEAR MARGIN

- ❖ SVMs work well when there is a clear margin of separation between the classes.
- ❖ If the classes are overlapping or inseparable, SVMs may struggle to find an appropriate decision boundary.
- ❖ In such cases, kernel tricks can be used to transform the data into a higher-dimensional space where it becomes separable.

HIGH-DIMENSIONAL FEATURE SPACES

- ❖ SVMs are effective when dealing with high-dimensional feature spaces, where the number of features is large compared to the number of samples.
- ❖ SVMs can handle such cases without suffering from the "curse of dimensionality" as they rely only on a subset of the training points (the support vectors) to define the decision boundary.

OUTLIER DETECTION

- ❖ SVMs are robust to outliers due to their reliance on support vectors.
- ❖ The decision boundary is determined by these critical points, making it less sensitive to outliers that lie far from the decision boundary.



SUMMARY

- Support vector machines

