MACHINE LEARNING

Shoaib Farooq

Department of Computer Science

Instructor Information

Instructor	Shoaib Farooq	E-mail	m.shoaib1050@gmail.com
GitHub	https://github.com/SHOAIB1050		
YouTube	https://www.youtube.com/c/pakithub		
LinkedIn	https://www.linkedin.com/in/shoaib-farooq-b5b190105/		

Let’s Start …..

Lecture #9

GOALS

This Lecture Will Cover:

- **Support Vector Machine**
- **Support Vector Regression**
- **Support Vector Classifier**

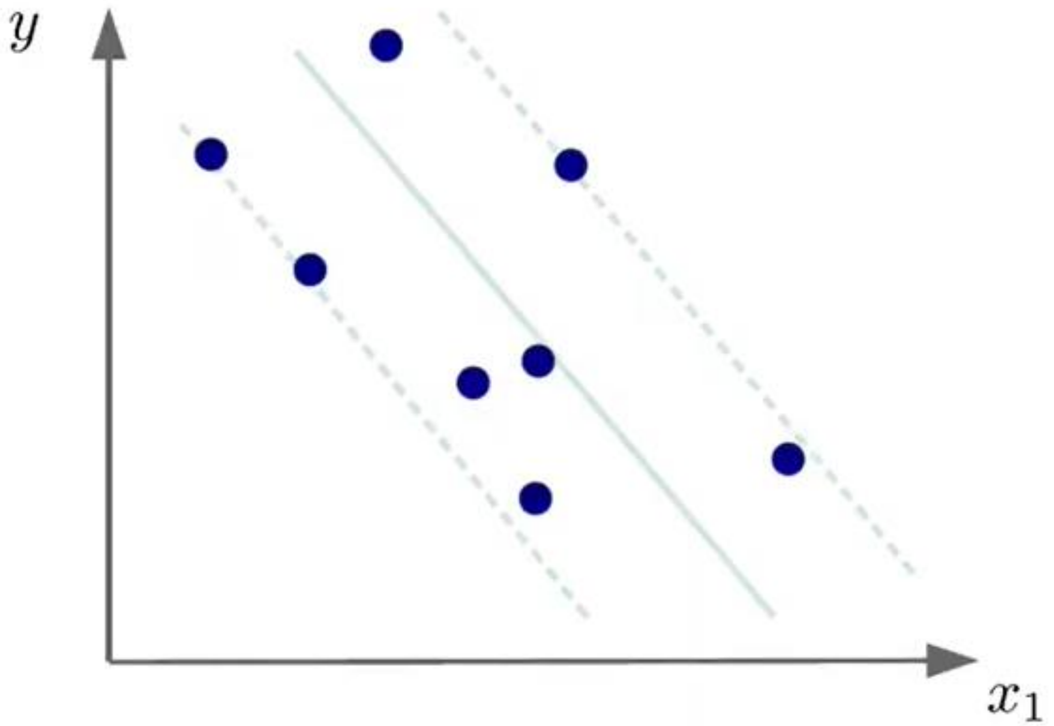
HISTORY OF SVM

- SVM is a statistical machine learning method introduced in 1992
- SVM becomes popular because of its success in handwritten digit recognition
  - 1.1% test error rate for SVM. This is the same as the error rates of a carefully constructed neural network, LeNet
  - SVM is now regarded as an important example of “kernel methods”, one of the key area in machine learning

Regression

Support Vector Regression

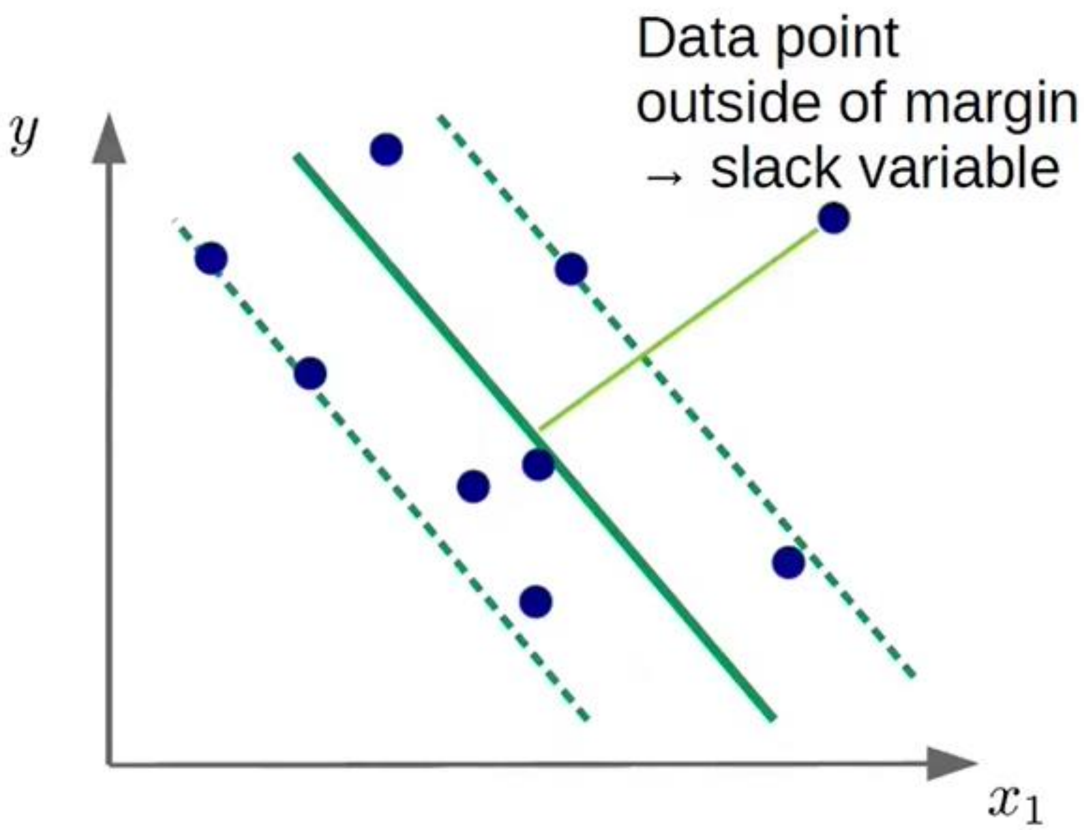
- Using the support vector machine to solve regression problems → support vector regression
- Margin lines are chosen so that they cover all data (hard margin) or allow for some violation (soft margin)
- Rest analogous to the classification



Regression

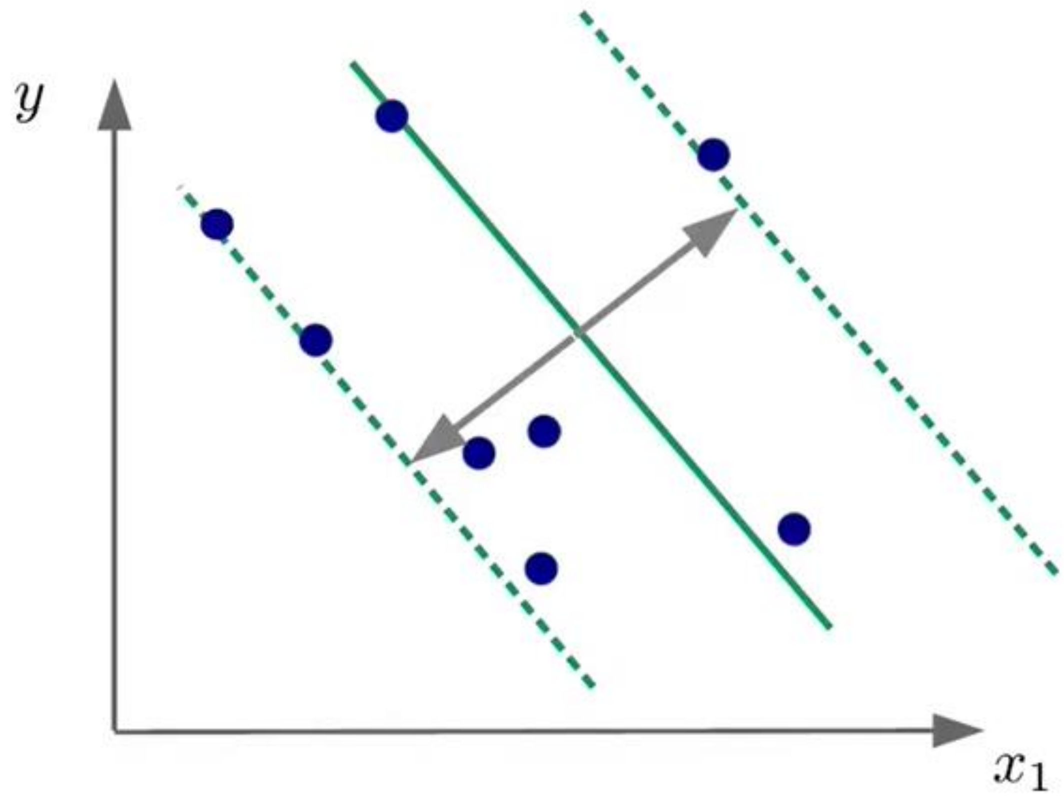
Support Vector Regression

- Using the support vector machine to solve regression problems → support vector regression
- Margin lines are chosen so that they cover all data (hard margin) or allow for some violation (soft margin)
- Rest analogous to the classification

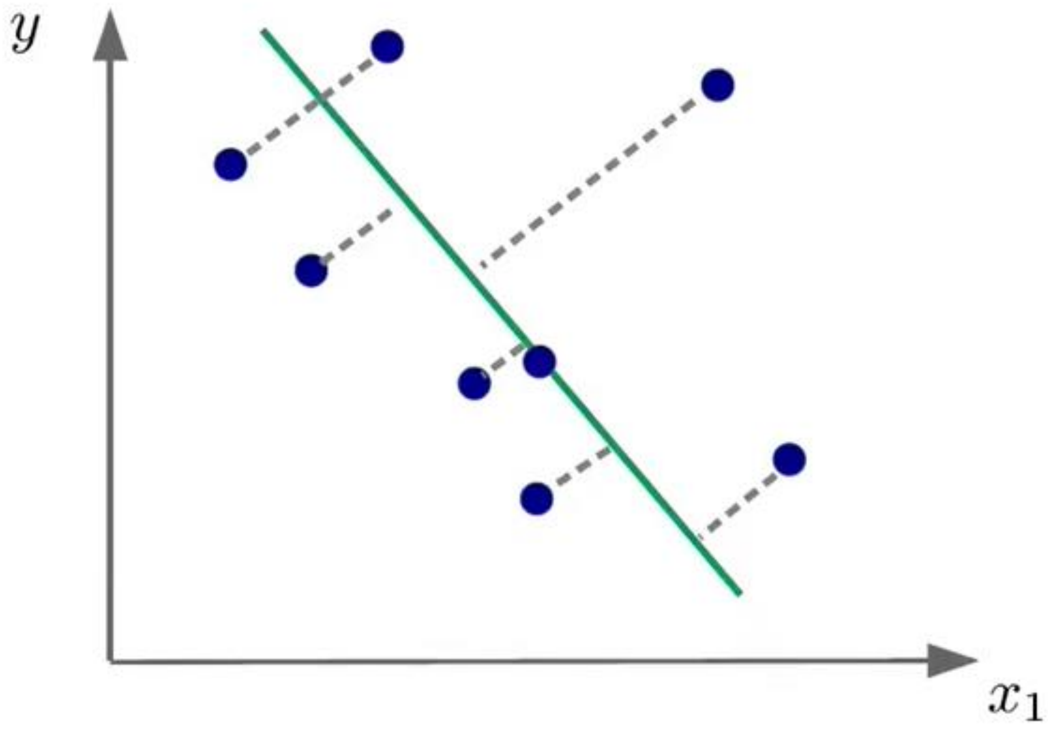


Regression

Support Vector Regression (SVR)

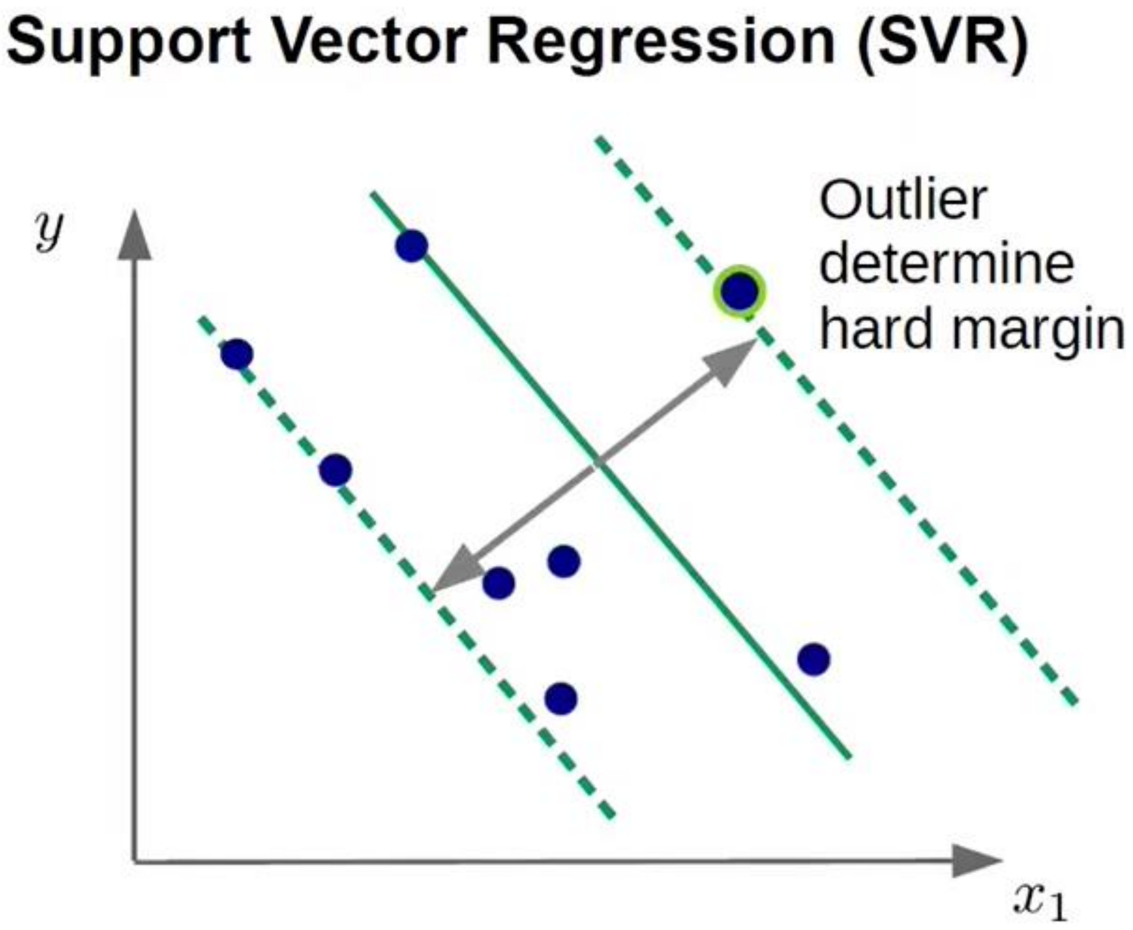


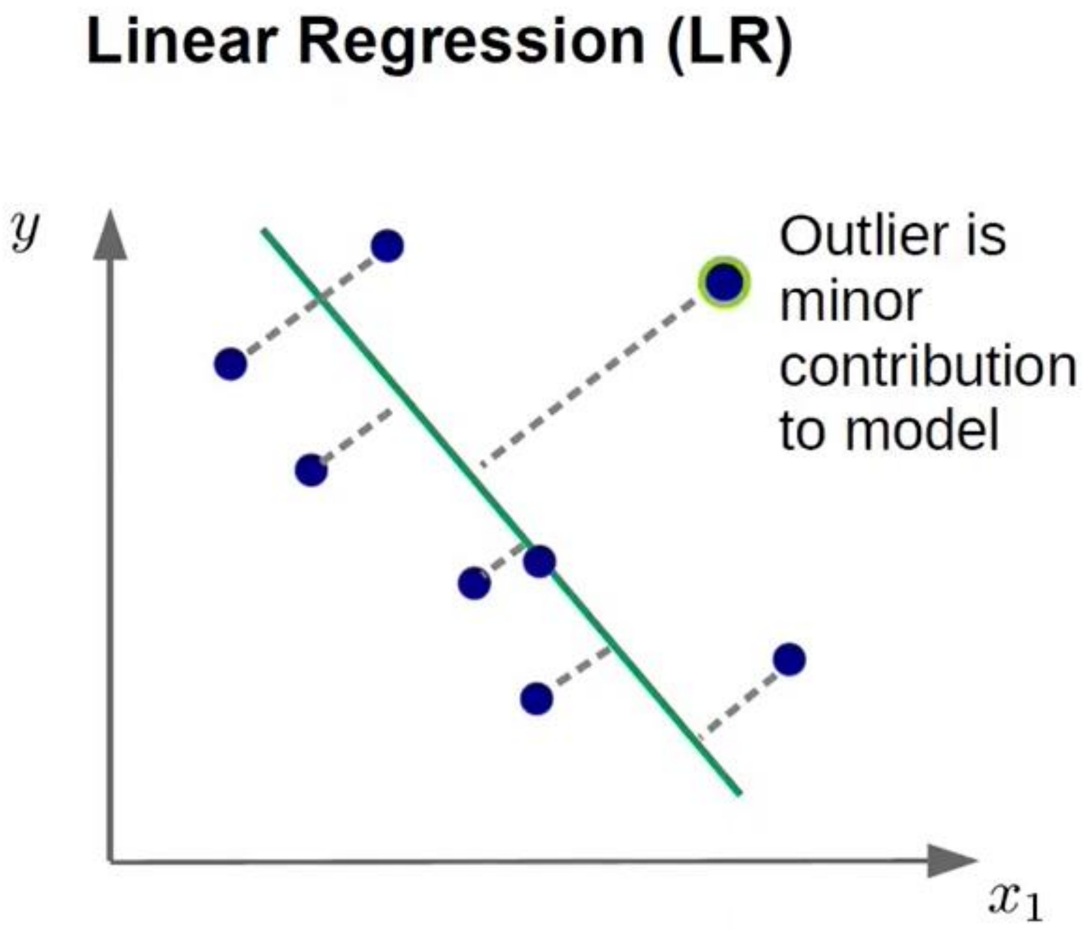
Linear Regression (LR)



- SVR without slack-variables more sensitive to outliers than LR
- SVR more insensitive to inner data points such as groups but efficient way to calculate

Regression

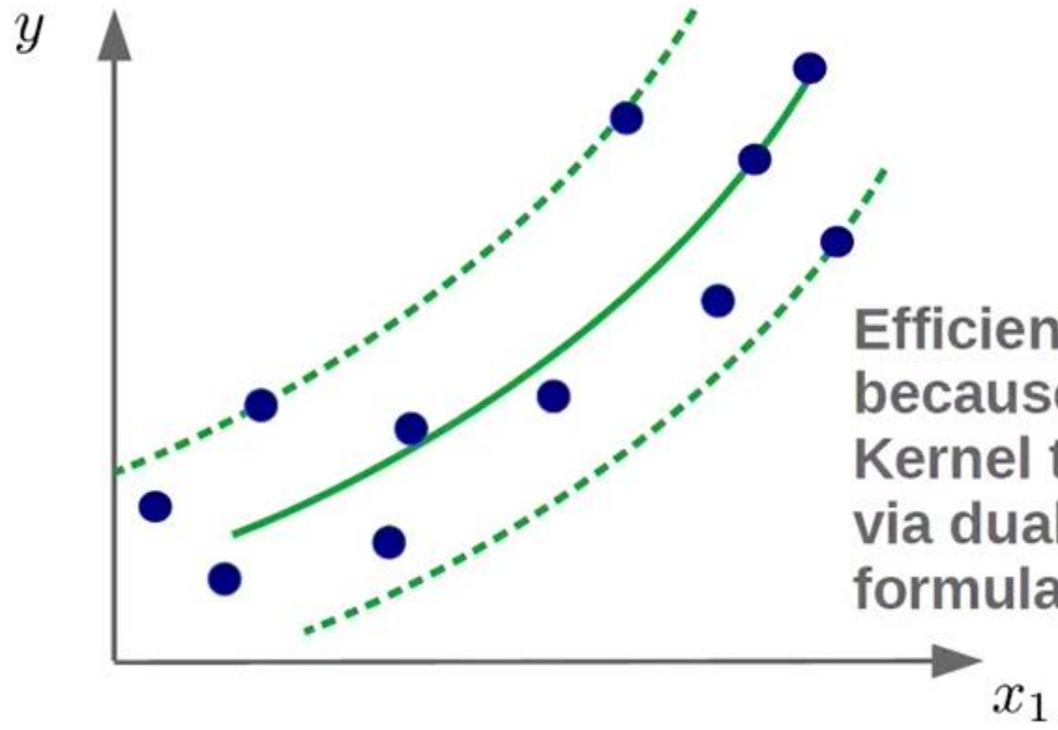




- SVR without slack-variables more sensitive to outliers than LR
- SVR more insensitive to inner data points such as groups but efficient way to calculate

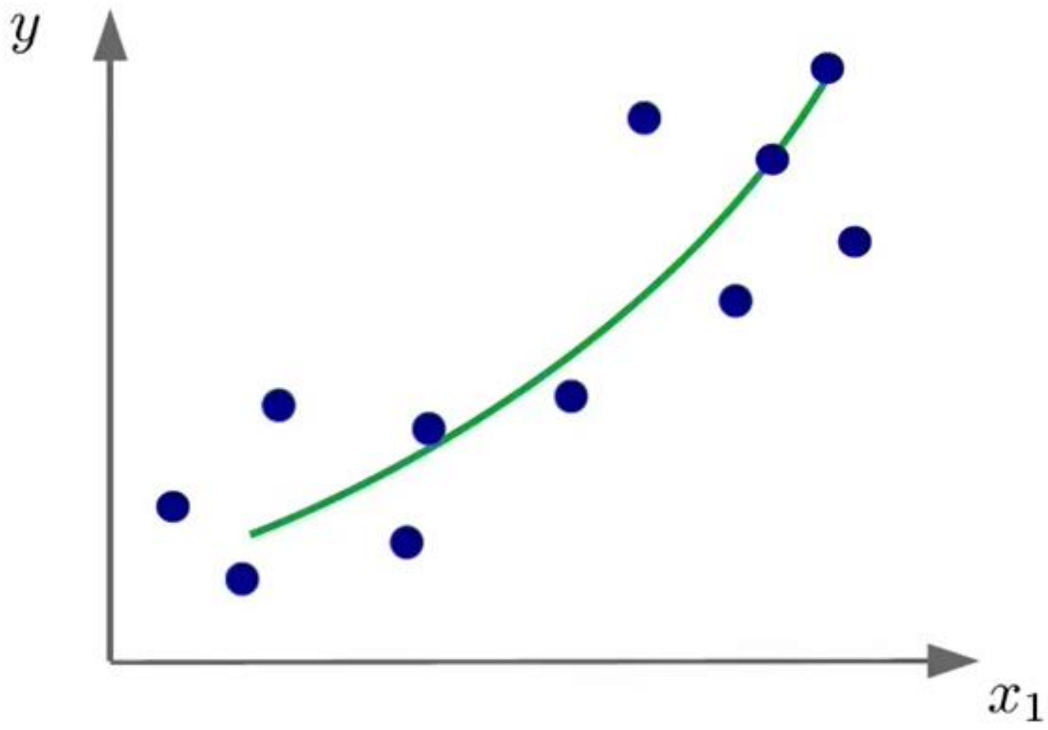
Regression

Support Vector Regression (SVR)



Efficient because of Kernel trick via dual formulation

Linear Regression (LR)



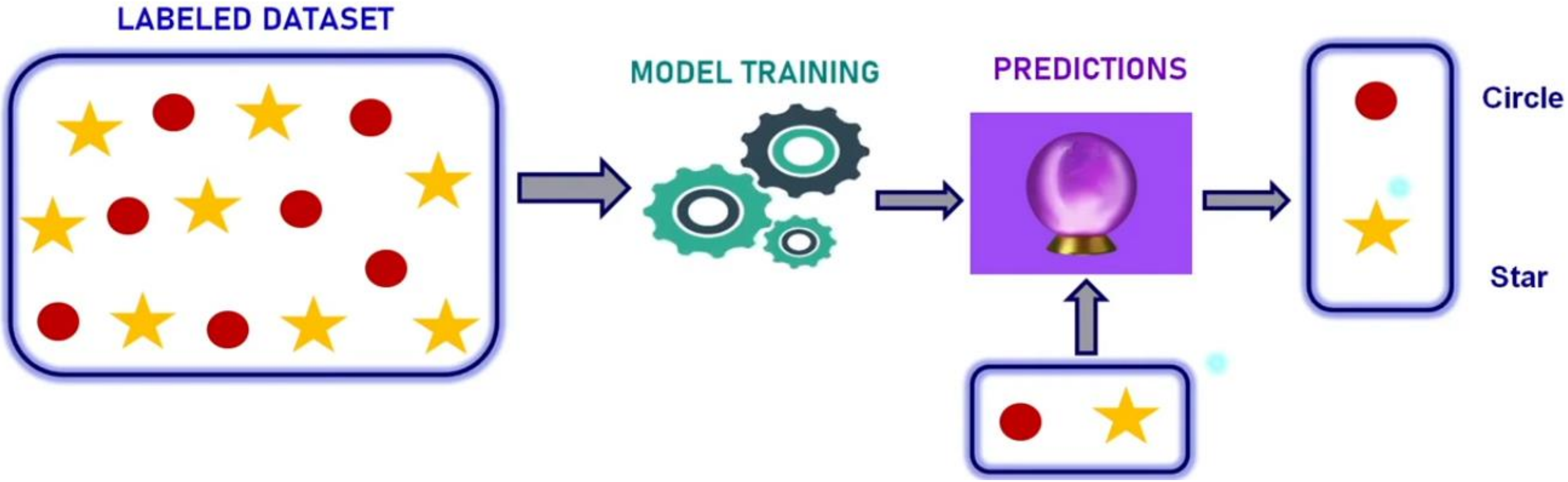
- SVR without slack-variables more sensitive to outliers than LR
- SVR more insensitive to inner data points such as groups but efficient way to calculate

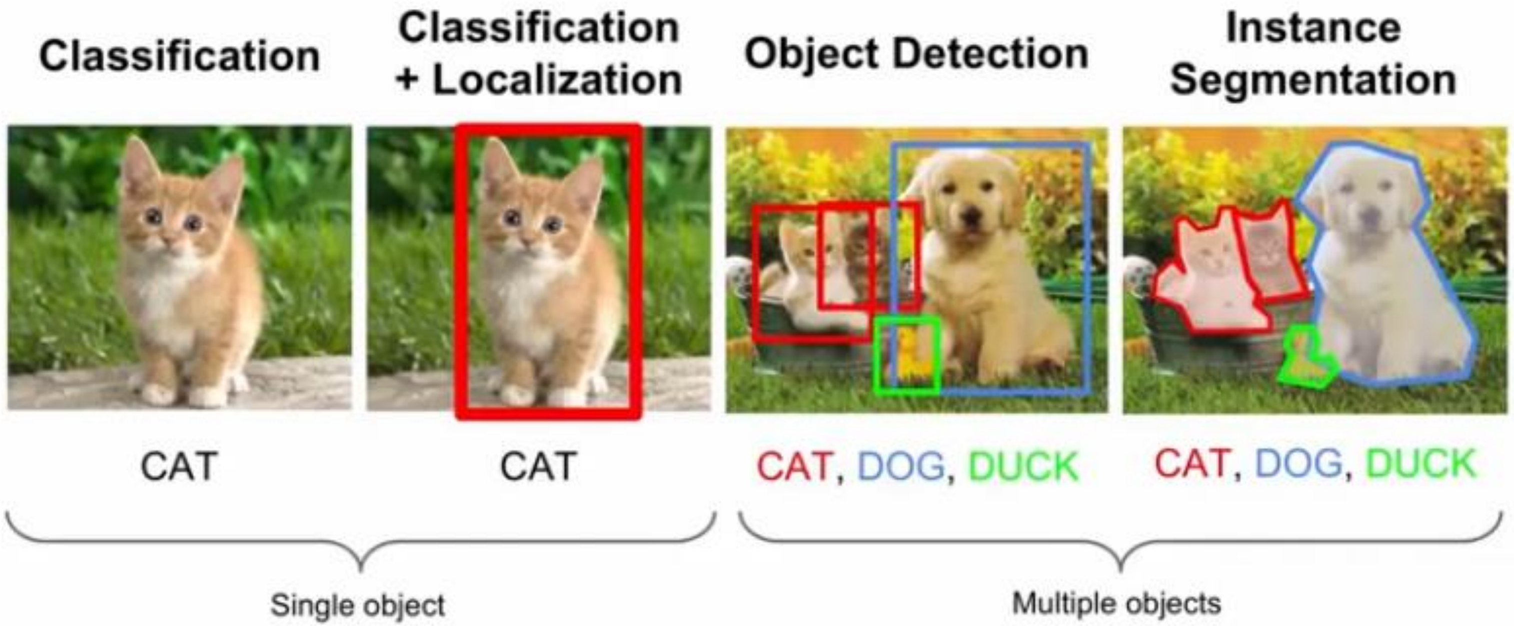
SVM CLASSIFIER

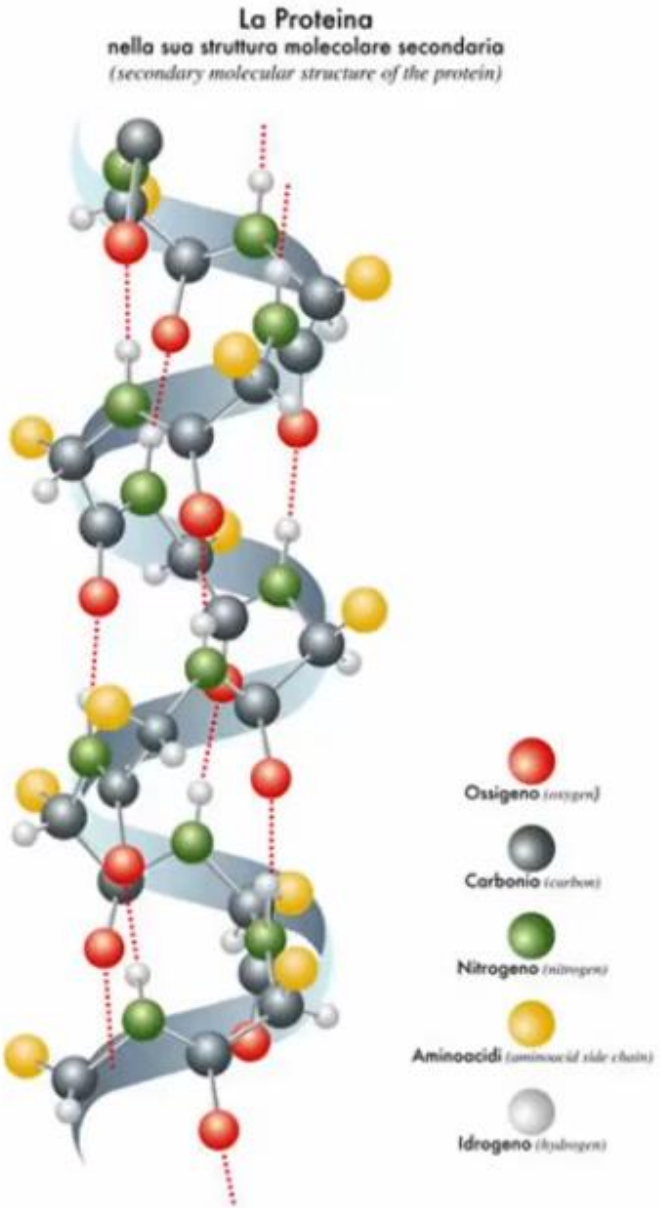
SVM is a type of machine learning algorithm that finds the optimal decision boundary (hyperplane) between two classes of data while maximizing the margin between the classes. Though they are used for classification of both linear and non-linear data.

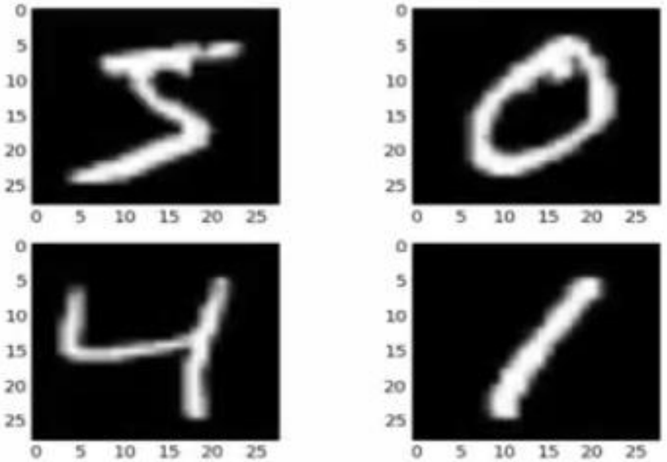
SUPPORT VECTOR MACHINE

Support Vector Machine is a machine learning algorithm based on supervised learning, that can be used for both regression and classification problems.

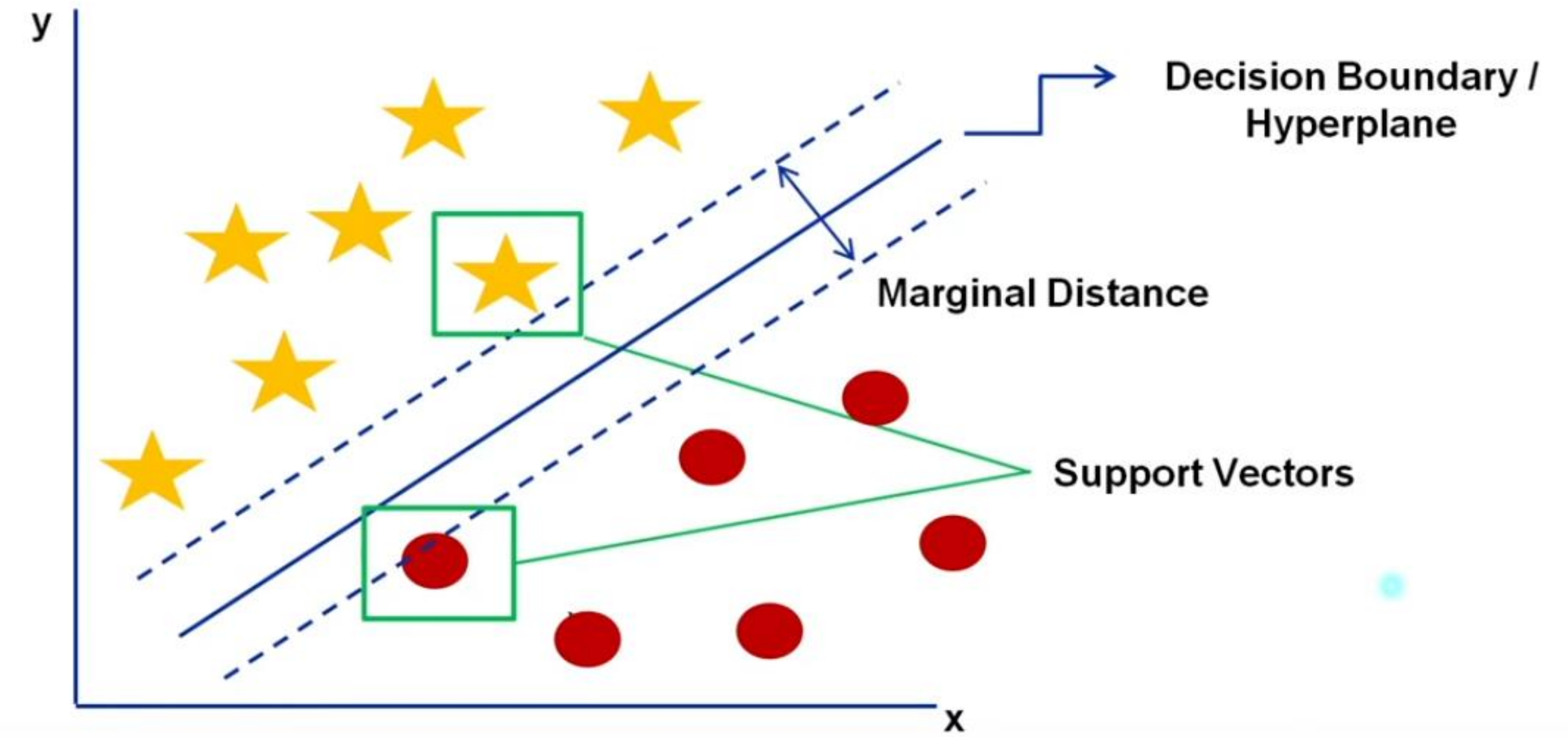



Classification
Classification + Localization
Object Detection
Instance Segmentation
CAT
CAT
CAT, DOG, DUCK
CAT, DOG, DUCK
Single object
Multiple objects


La Proteina
nella sua struttura molecolare secondaria
(secondary molecular structure of the protein)
Ossigeno
Carbonio
Nitrogeno
Aminoacidi
Idrogeno


0 5 10 15 20 25

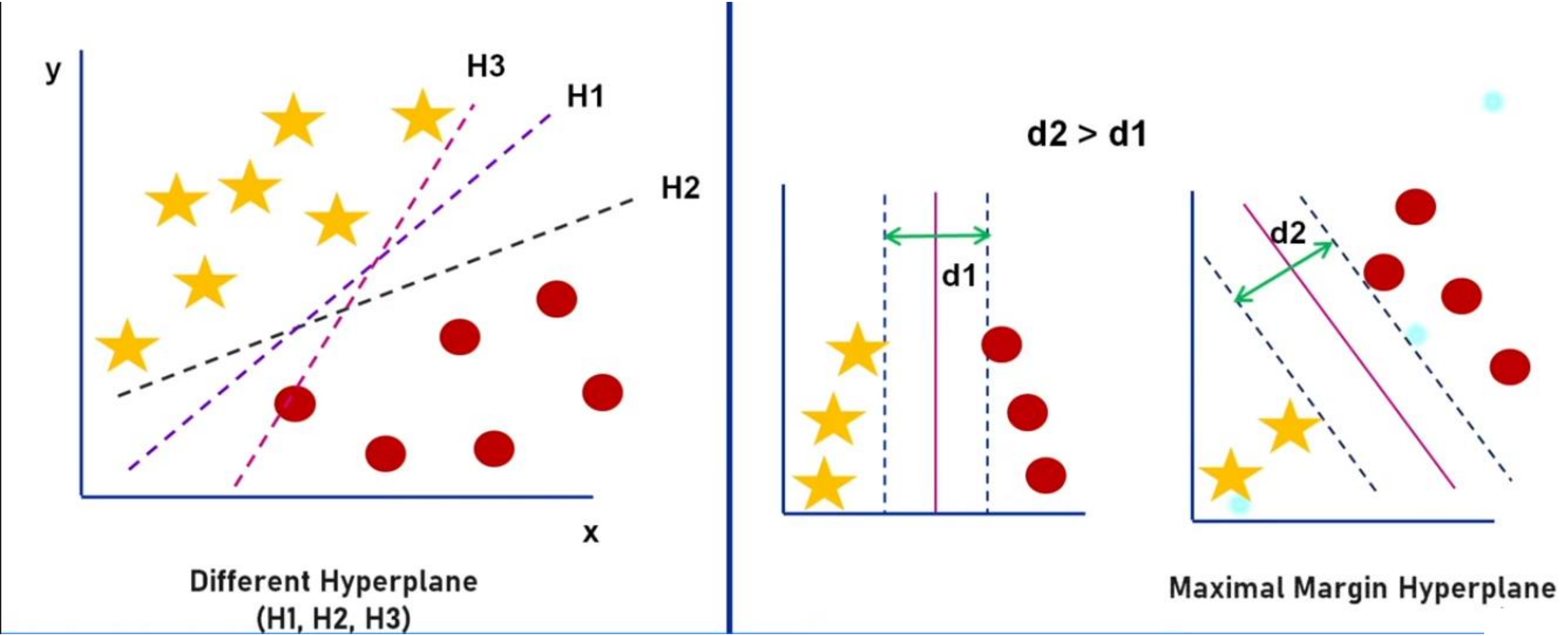
SUPPORT VECTOR MACHINE



SUPPORT VECTOR MACHINE

- **Decision Boundary** also known as the hyperplane, this is the line that separates the data points into different classes.
- **Margin Distance** The distance between the decision boundary and the closest data points to the boundary, also known as the support vectors.
- **Support vectors** The data points that are closest to the decision boundary and have the most influence on its position.

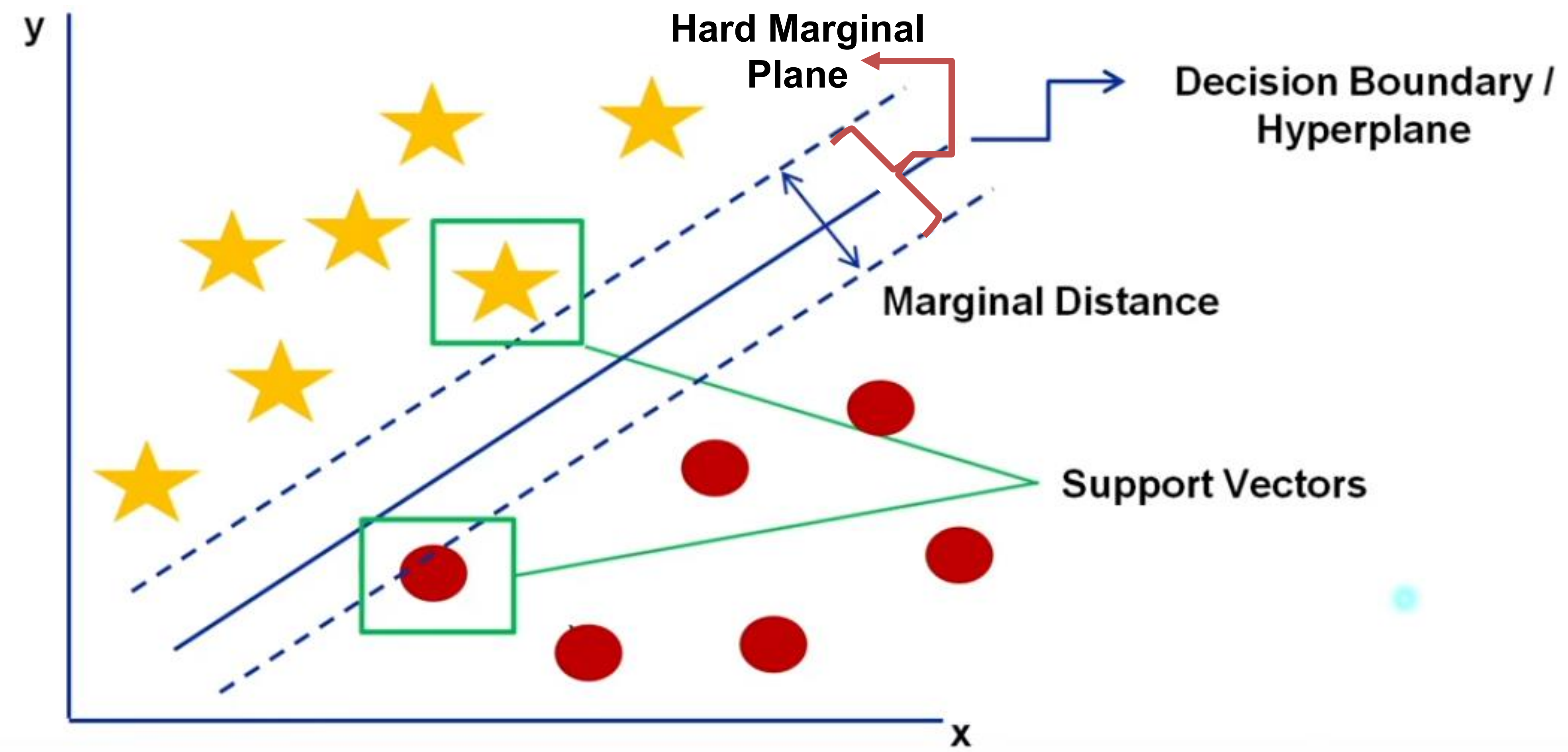
WHICH HYPERPLANE TO SELECT?



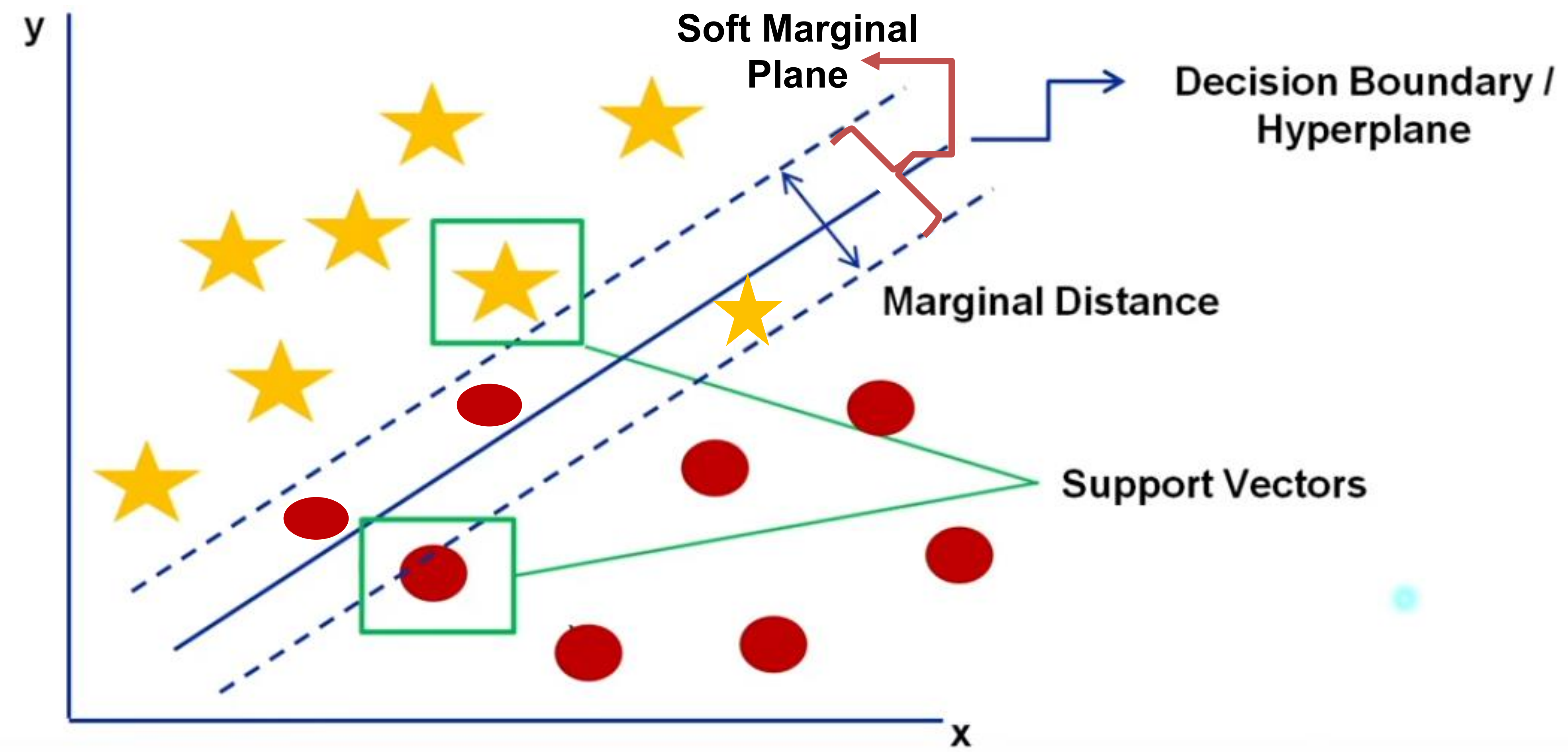
LARGE-MARGIN DECISION BOUNDARY

- Large Margin refers to the wide separation or “margin” between the decision boundary (hyperplane) and the nearest data points of each class.
- In a large margin classification, the SVM tries to find a decision boundary that not only correctly classifies the data points but also maximizes the margin between the classes.
- It provides effective classification for newly added instances and fewer outliers in classification.

SUPPORT VECTOR MACHINE



SUPPORT VECTOR MACHINE

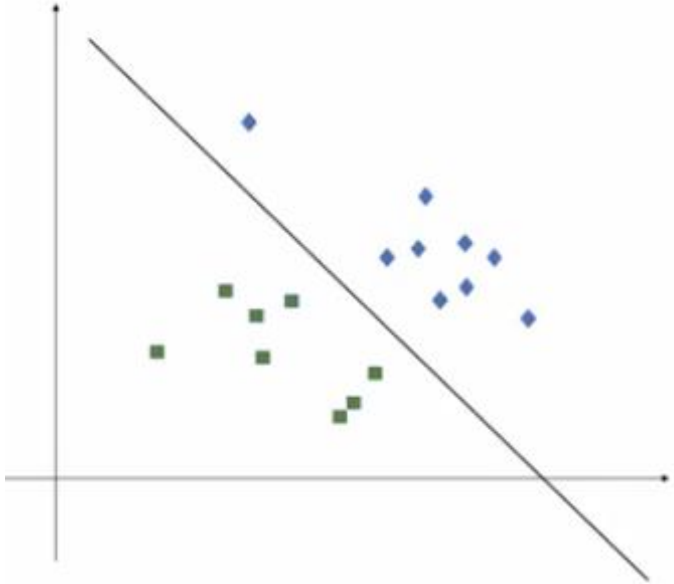


HARD MARGINAL PLANE

- A **Hard Margin SVM** is used when the data is perfectly linearly separable, meaning there is a clear linear boundary that can separate the data points into distinct classes with no errors or overlap.
- **Soft Margin SVM** Soft margin SVM allows for some margin violations, meaning that it permits certain data points to fall within the margin or even on the wrong side of the decision boundary

SVM CLASSIFIER

2D



3D

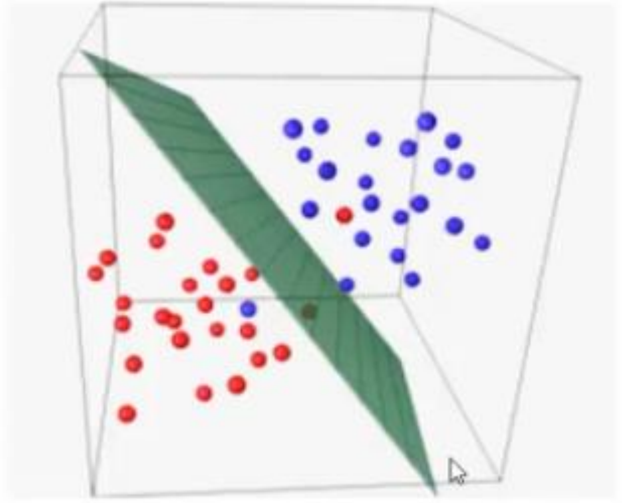
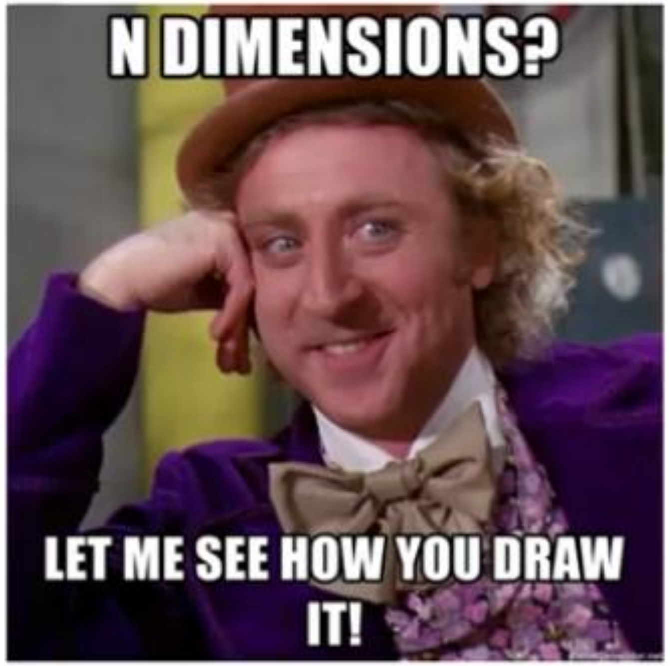
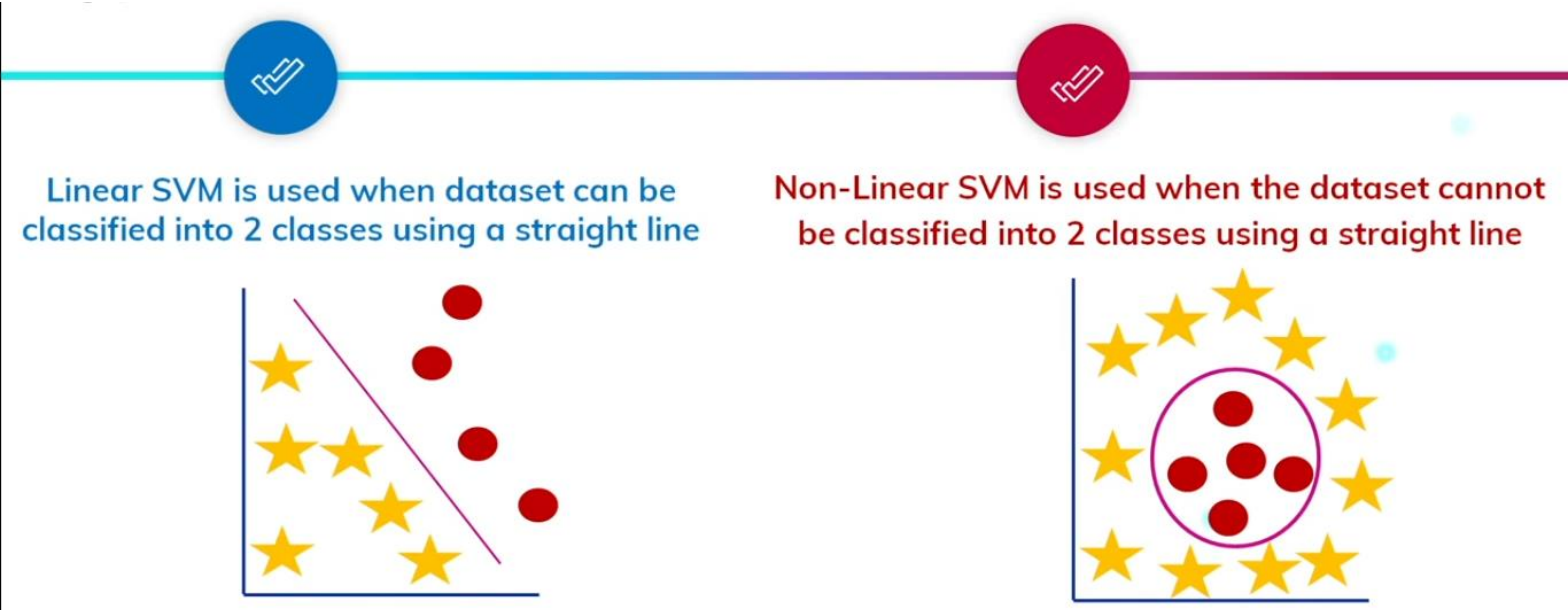


Image Credit: https://appliedmachinelearning.blog

nD



TYPES OF SVM

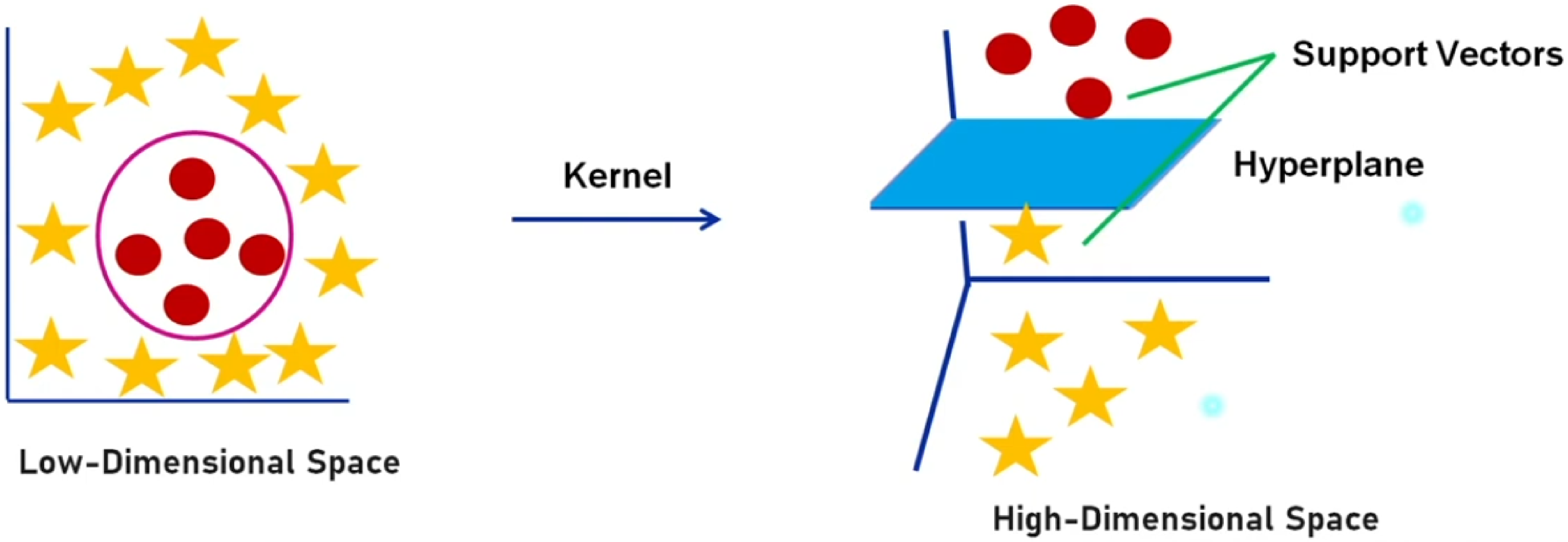


Liner SVM

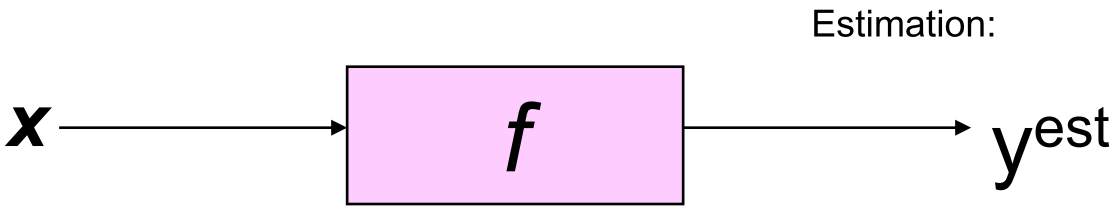
Non Liner SVM

KERNEL FUNCTION

Kernel functions takes low dimensional input space and transform it into a higher-dimensional space, i.e., it converts not separable problem to separable problem.

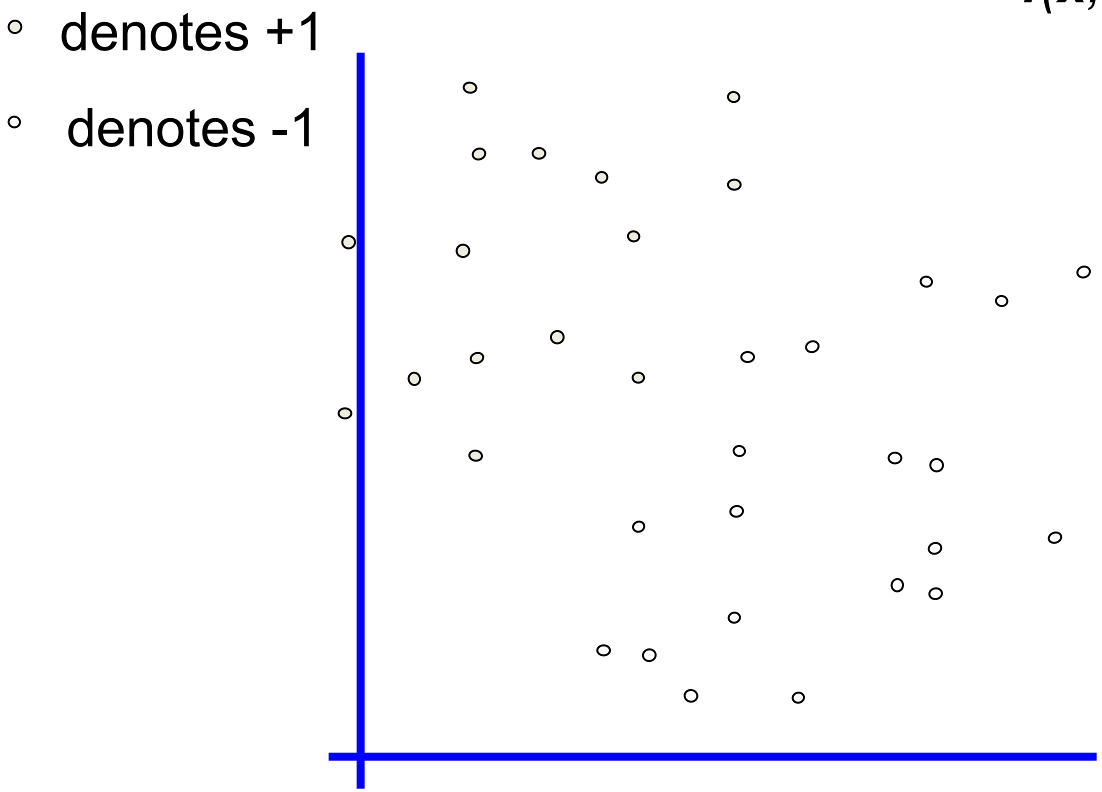


LINEAR CLASSIFIERS



$$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} + b)$$

w: weight vector
x: data vector



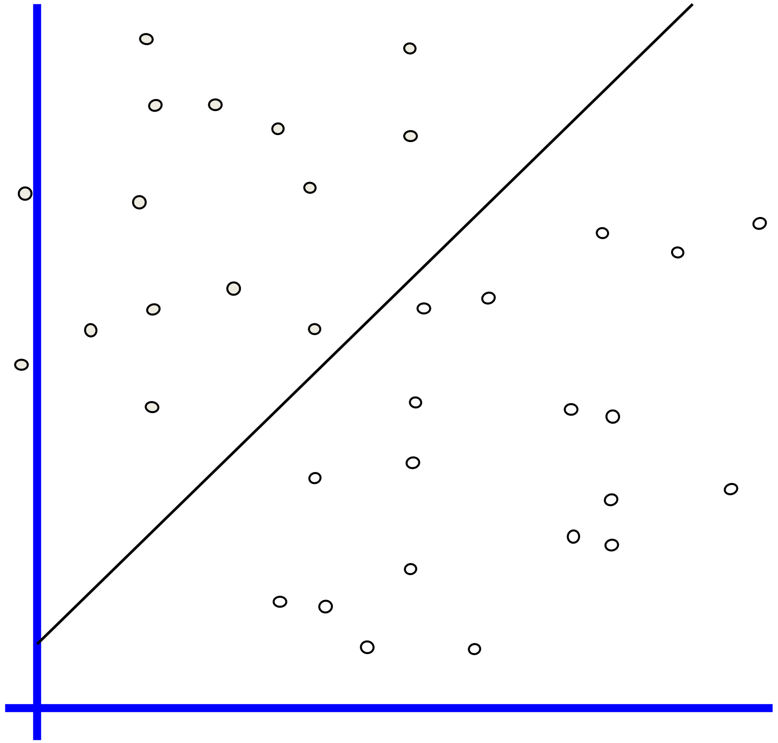
How would you classify this data?

LINEAR CLASSIFIERS

$$\boldsymbol{x} \longrightarrow \boxed{f} \longrightarrow y^{est}$$

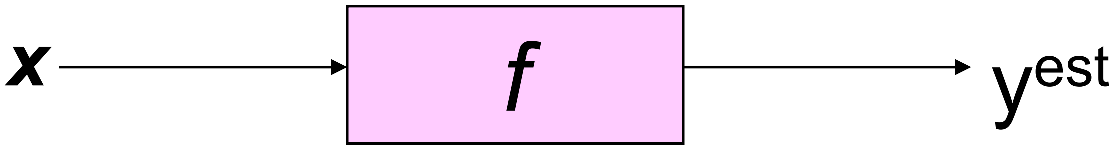
$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w}, b) = sign(\boldsymbol{w}.\ \boldsymbol{x} + b)$

- denotes +1
- denotes -1



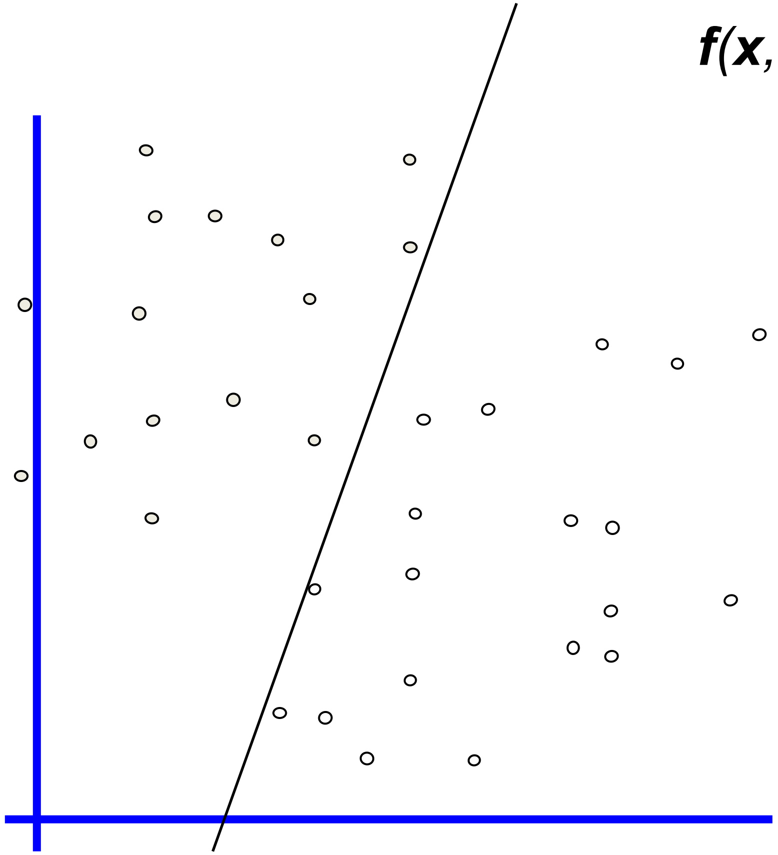
How would you classify this data?

LINEAR CLASSIFIERS



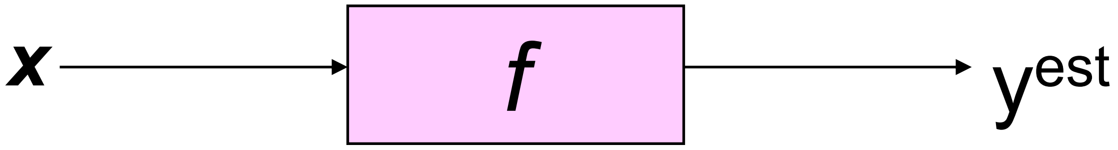
$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w} \cdot \mathbf{x} + b)$

- denotes +1
- denotes -1

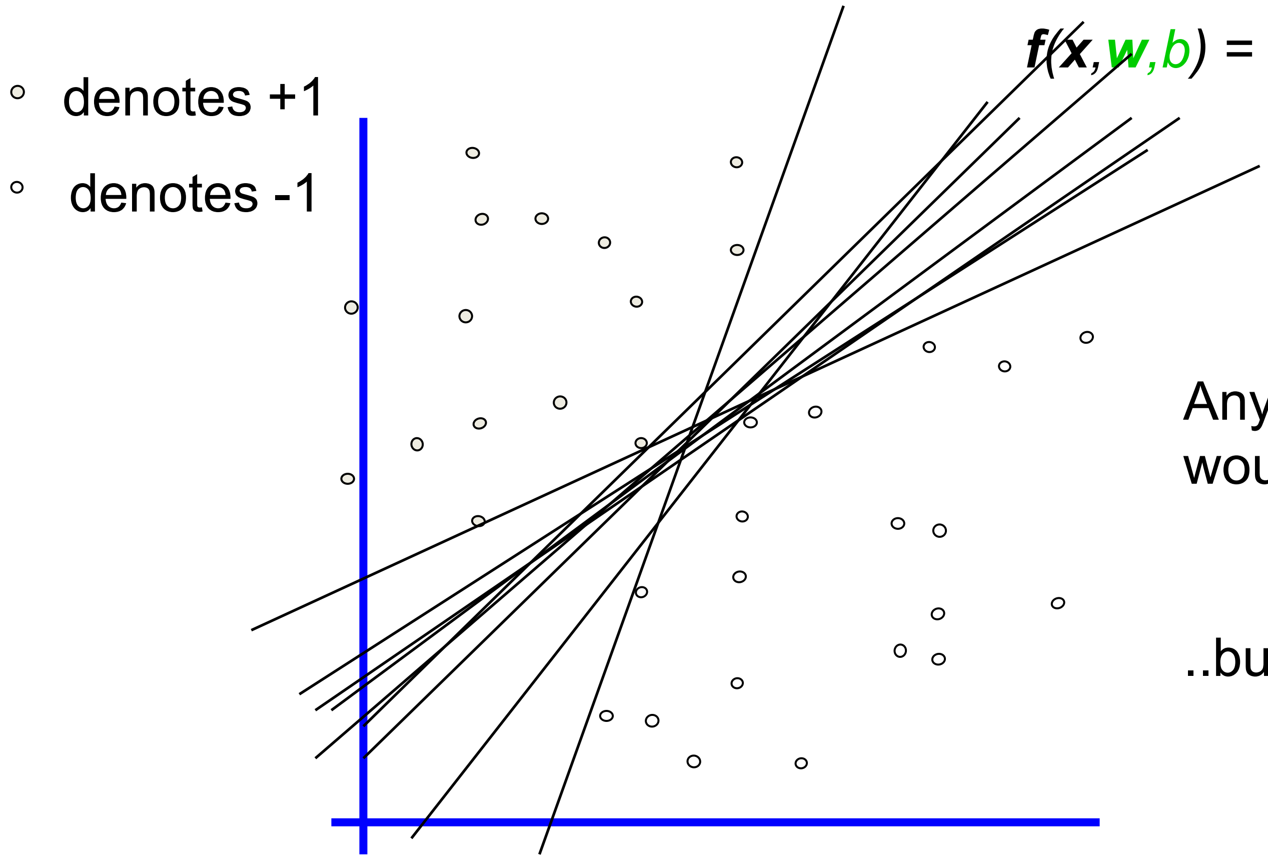


How would you classify this data?

LINEAR CLASSIFIERS



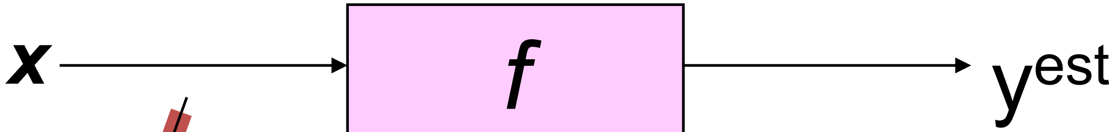
$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} + b)$



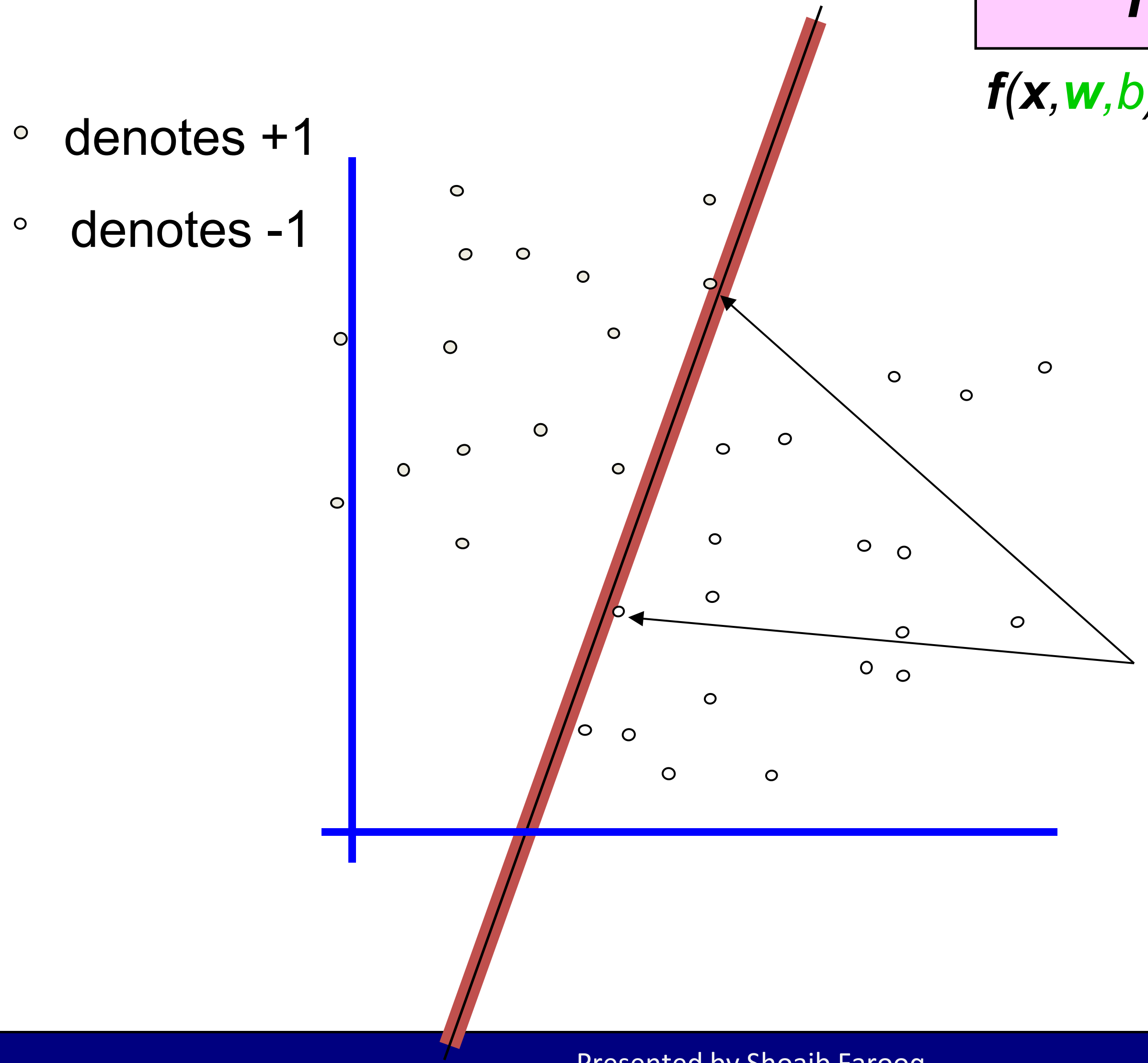
Any of these would be fine..

..but which is best?

CLASSIFIER MARGIN

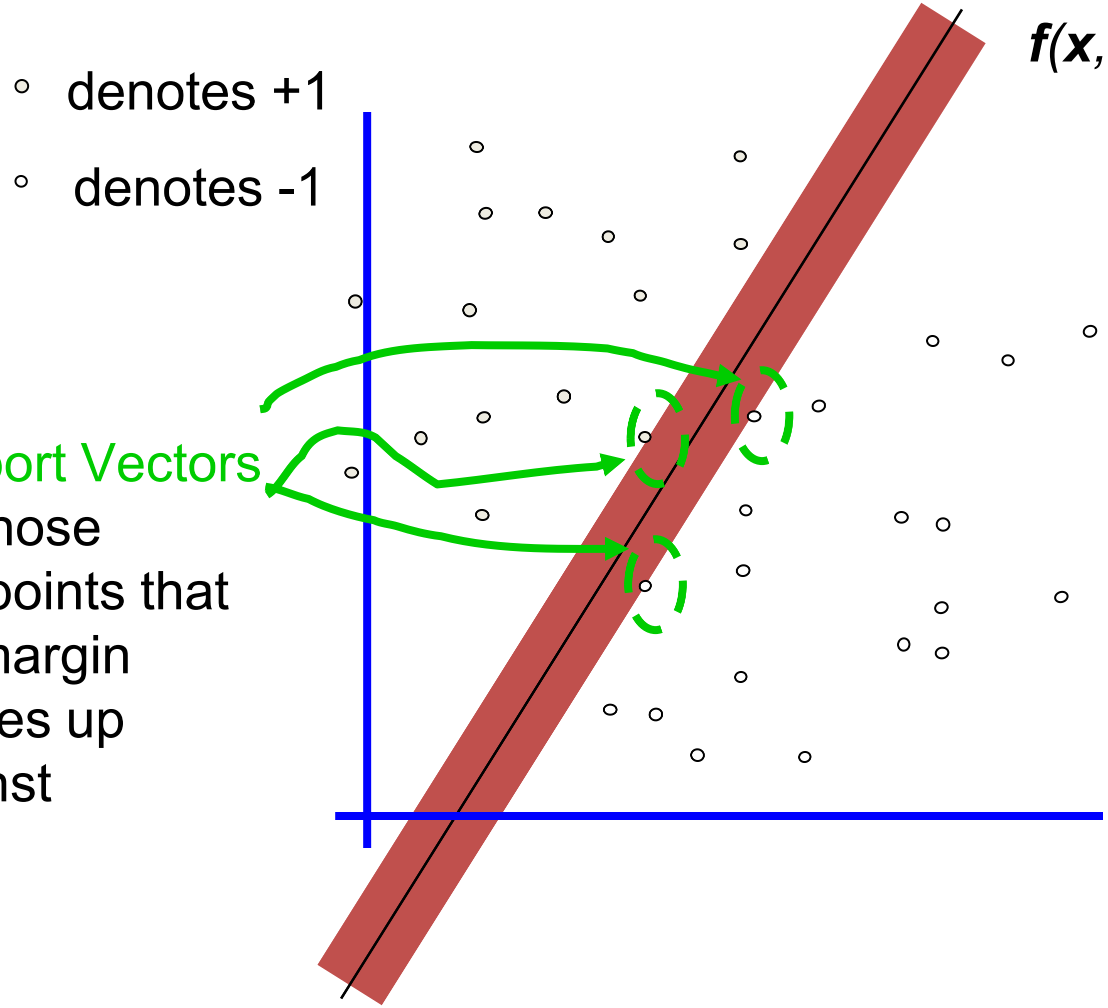


$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} + b)$



Define the margin of a linear classifier as the width that the boundary could be increased by before hitting a datapoint.

MAXIMUM MARGIN



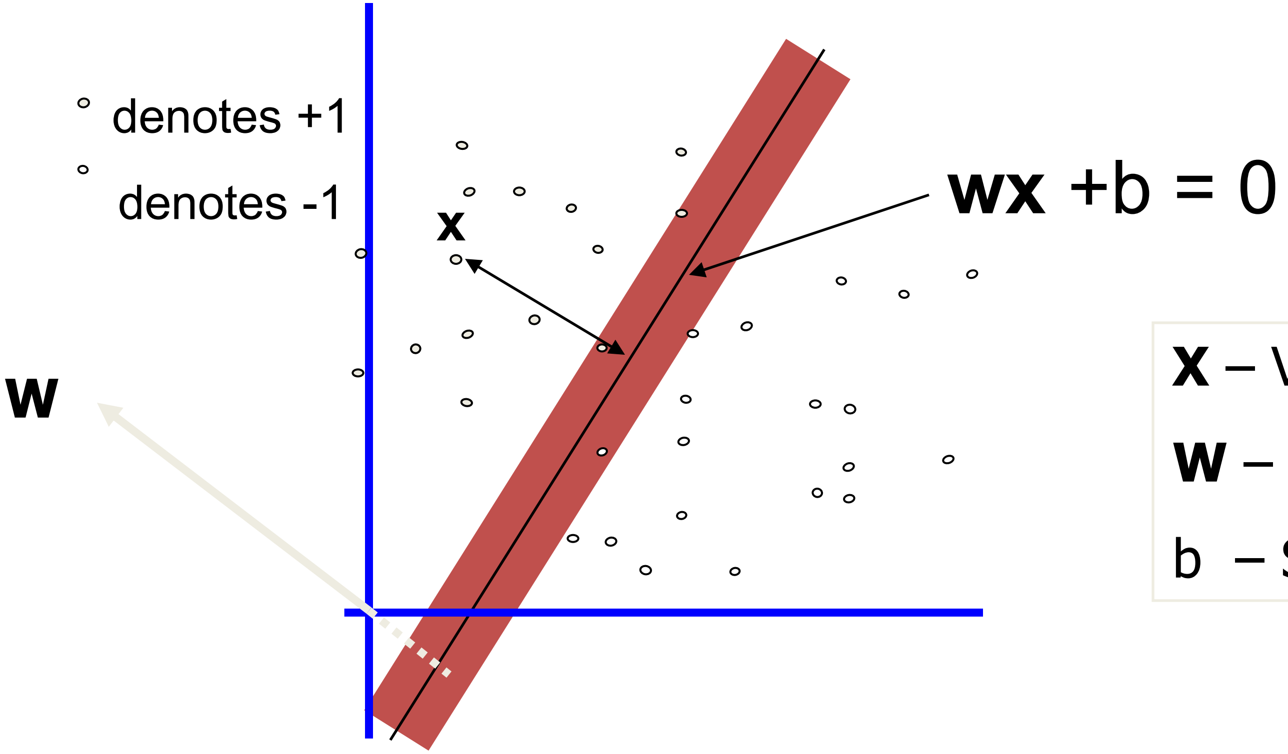
Support Vectors are those datapoints that the margin pushes up against

$$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} + b)$$

The maximum margin linear classifier is the linear classifier with the, maximum margin for better classification.

This is the simplest kind of SVM (Called an LSVM)

MAXIMUM MARGIN



X – Vector

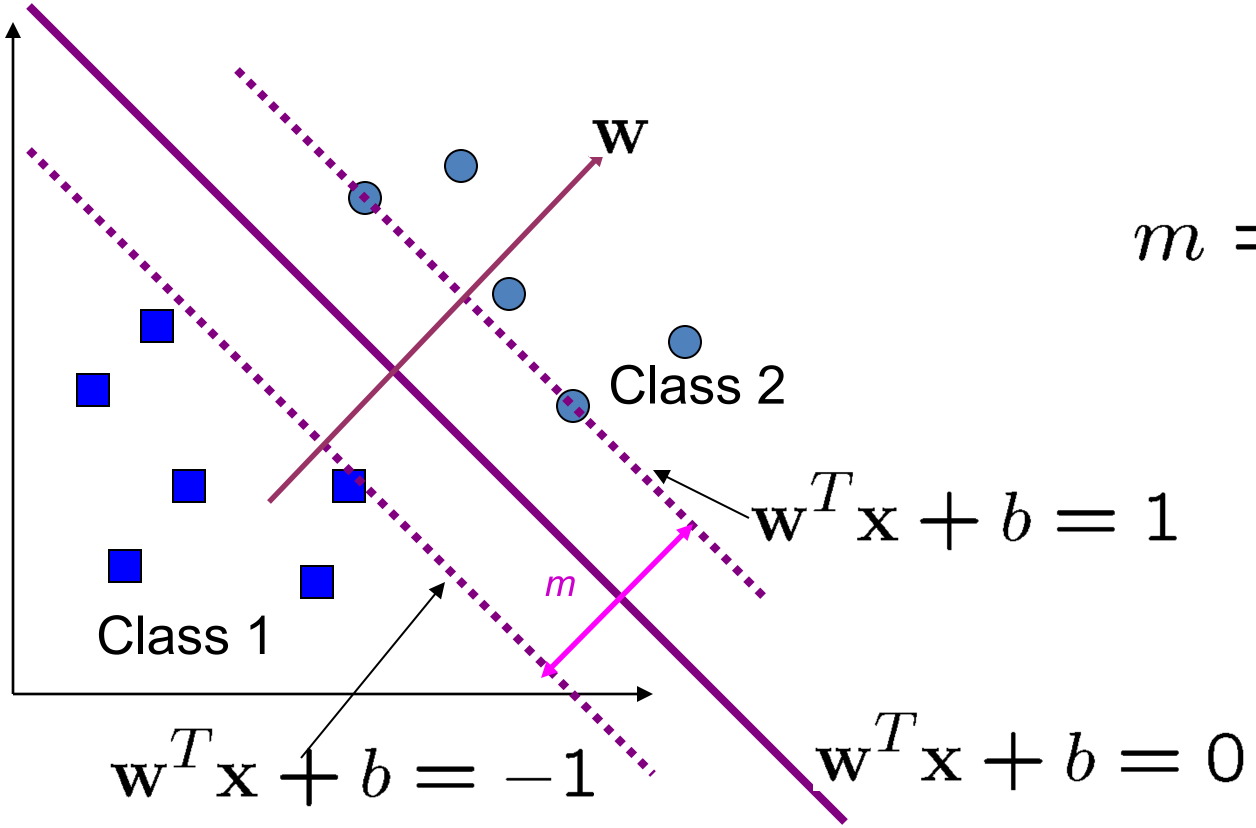
W – Normal Vector

b – Scale Value

- In our case, $w_1*x_1+w_2*x_2+b=0$,
- thus, $\mathbf{w}=(w_1,w_2)$, $\mathbf{x}=(x_1,x_2)$

LARGE-MARGIN DECISION BOUNDARY

- The decision boundary should be as far away from the data of both classes as possible
  - We should maximize the margin, *m*



$$m = \frac{2}{||\mathbf{w}||}$$

FINDING THE DECISION BOUNDARY

- Using the SVM algorithm, find the hyperplane with maximum margin for the following data.

X1	X2	Class
2	1	+1
4	3	-1

FINDING THE DECISION BOUNDARY

- $N = 2$
- $\vec{x}_1 = (2, 1)$
- $\vec{x}_2 = (4, 3)$
- $y_1 = +1$
- $y_2 = -1$
- $\vec{\alpha} = (\alpha_1, \alpha_2)$

X1	X2	Class
2	1	+1
4	3	-1

Decision Rule

$$f(\vec{x}) = \vec{w} \cdot \vec{x} - b$$

FINDING THE DECISION BOUNDARY

$$\phi(\vec{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{N} \alpha_i \alpha_j y_i y_j (\vec{x}_i \cdot \vec{x}_j)$$

$$= (\alpha_1 + \alpha_2) - \frac{1}{2}[\alpha_1 \alpha_1 y_1 y_1 (\vec{x}_1 \cdot \vec{x}_1) + \alpha_1 \alpha_2 y_1 y_2 (\vec{x}_1 \cdot \vec{x}_2) +$$

$$\alpha_2 \alpha_1 y_2 y_1 (\vec{x}_2 \cdot \vec{x}_1) + \alpha_2 \alpha_2 y_2 y_2 (\vec{x}_2 \cdot \vec{x}_2)]$$

$$= (\alpha_1 + \alpha_2) -$$

$$\frac{1}{2}[\alpha_1^2 (+1)(+1)(2 \times 2 + 1 \times 1) + \alpha_1 \alpha_2 (+1)(-1)(2 \times 4 + 1 \times 3) +$$

$$\alpha_2 \alpha_1 (-1)(+1)(4 \times 2 + 3 \times 1) + \alpha_2^2 (-1)(-1)(4 \times 4 + 3 \times 3)]$$

$$= (\alpha_1 + \alpha_2) - \frac{1}{2}\left[5\alpha_1^2 - 22\alpha_1 \alpha_2 + 25\alpha_2^2\right]$$

X1	X2	Class
2	1	+1
4	3	-1

$$N = 2$$
$$\vec{x}_1 = (2, 1)$$
$$\vec{x}_2 = (4, 3)$$
$$y_1 = +1$$
$$y_2 = -1$$
$$\vec{\alpha} = (\alpha_1, \alpha_2)$$
$$\alpha_1 = \alpha_2$$
$$\alpha_1, \alpha_2 > 0$$

FINDING THE DECISION BOUNDARY

- Find values of α_1 and α_2 which maximizes

$$\phi(\vec{\alpha}) = (\alpha_1 + \alpha_2) - \frac{1}{2}\left[5\alpha_1^2 - 22\alpha_1\alpha_2 + 25\alpha_2^2\right]$$

$$\phi(\vec{\alpha}) = 2\alpha_1 - 4\alpha_1^2$$

- For Ø to be maximum we must have

$$\frac{d\phi}{d\alpha_1} = 2 - 8\alpha_1 = 0$$

$$\alpha_1 = \frac{1}{4}$$

$$\alpha_2 = \frac{1}{4}$$

X1	X2	Class
2	1	+1
4	3	-1

$$N = 2$$
$$\vec{x}_1 = (2, 1)$$
$$\vec{x}_2 = (4, 3)$$
$$y_1 = +1$$
$$y_2 = -1$$
$$\vec{\alpha} = (\alpha_1, \alpha_2)$$
$$\alpha_1 = \alpha_2$$

FINDING THE DECISION BOUNDARY

$$\vec{w} = \sum_{i=1}^{N} \alpha_i y_i \vec{x}_i$$

$$= \alpha_1 y_1 \vec{x}_1 + \alpha_2 y_2 \vec{x}_2$$

$$= \frac{1}{4}(+1)(2,1) + \frac{1}{4}(-1)(4,3)$$

$$= \frac{1}{4}(-2,-2)$$

$$= \left(-\tfrac{1}{2}, -\tfrac{1}{2}\right)$$

$$\alpha_1 = \frac{1}{4}$$

$$\alpha_2 = \frac{1}{4}$$

X1	X2	Class
2	1	+1
4	3	-1

$$N = 2$$
$$\vec{x}_1 = (2,1)$$
$$\vec{x}_2 = (4,3)$$
$$y_1 = +1$$
$$y_2 = -1$$
$$\vec{\alpha} = (\alpha_1, \alpha_2)$$
$$\alpha_1 = \alpha_2$$
$$\alpha_1, \alpha_2 > 0$$

FINDING THE DECISION BOUNDARY

$$b = \frac{1}{2}\left(\min_{i:y_i=+1}(\vec{w}\cdot\vec{x}_i) + \max_{i:y_i=-1}(\vec{w}\cdot\vec{x}_i)\right)$$

$$= \frac{1}{2}((\vec{w}\cdot\vec{x}_1) + (\vec{w}\cdot\vec{x}_2))$$

$$= \frac{1}{2}((-\tfrac{1}{2}\times 2 - \tfrac{1}{2}\times 1) + (-\tfrac{1}{2}\times 4 - \tfrac{1}{2}\times 3))$$

$$= \frac{1}{2}\left(-\frac{10}{2}\right) = -\frac{5}{2}$$

$$\alpha_1 = \frac{1}{4}$$

$$\alpha_2 = \frac{1}{4}$$

$$\vec{w} = (-\tfrac{1}{2}, -\tfrac{1}{2})$$

X1	X2	Class
2	1	+1
4	3	-1

$$N = 2$$
$$\vec{x}_1 = (2,1)$$
$$\vec{x}_2 = (4,3)$$
$$y_1 = +1$$
$$y_2 = -1$$
$$\vec{\alpha} = (\alpha_1, \alpha_2)$$
$$\alpha_1 = \alpha_2$$
$$\alpha_1, \alpha_2 > 0$$

FINDING THE DECISION BOUNDARY

- The SVM classifier function is given by

$$f(\vec{x}) = \vec{w} \cdot \vec{x} - b$$

$$= \left(-\tfrac{1}{2}, -\tfrac{1}{2}\right) \cdot (x_1, x_2) - \left(-\tfrac{5}{2}\right)$$

$$= -\frac{1}{2}x_1 - \frac{1}{2}x_2 + \frac{5}{2}$$

$$= -\frac{1}{2}(x_1 + x_2 - 5)$$

$$\alpha_1 = \frac{1}{4}$$

$$\alpha_2 = \frac{1}{4}$$

$$\vec{w} = \left(-\tfrac{1}{2}, -\tfrac{1}{2}\right)$$

$$b = -\frac{5}{2}$$

X1	X2	Class
2	1	+1
4	3	-1

$$N = 2$$
$$\vec{x}_1 = (2, 1)$$
$$\vec{x}_2 = (4, 3)$$
$$y_1 = +1$$
$$y_2 = -1$$
$$\vec{\alpha} = (\alpha_1, \alpha_2)$$
$$\alpha_1 = \alpha_2$$
$$\alpha_1, \alpha_2 > 0$$

FINDING THE DECISION BOUNDARY

- The equation of the maximal margin hyperplane is

$$f(\vec{x}) = 0$$

$$f(\vec{x}) = -\frac{1}{2}(x_1 + x_2 - 5)$$

$$-\frac{1}{2}(x_1 + x_2 - 5) = 0$$

$$x_1 + x_2 - 5 = 0$$

$$\alpha_1 = \frac{1}{4}$$

$$\alpha_2 = \frac{1}{4}$$

$$\vec{w} = (-\tfrac{1}{2}, -\tfrac{1}{2})$$

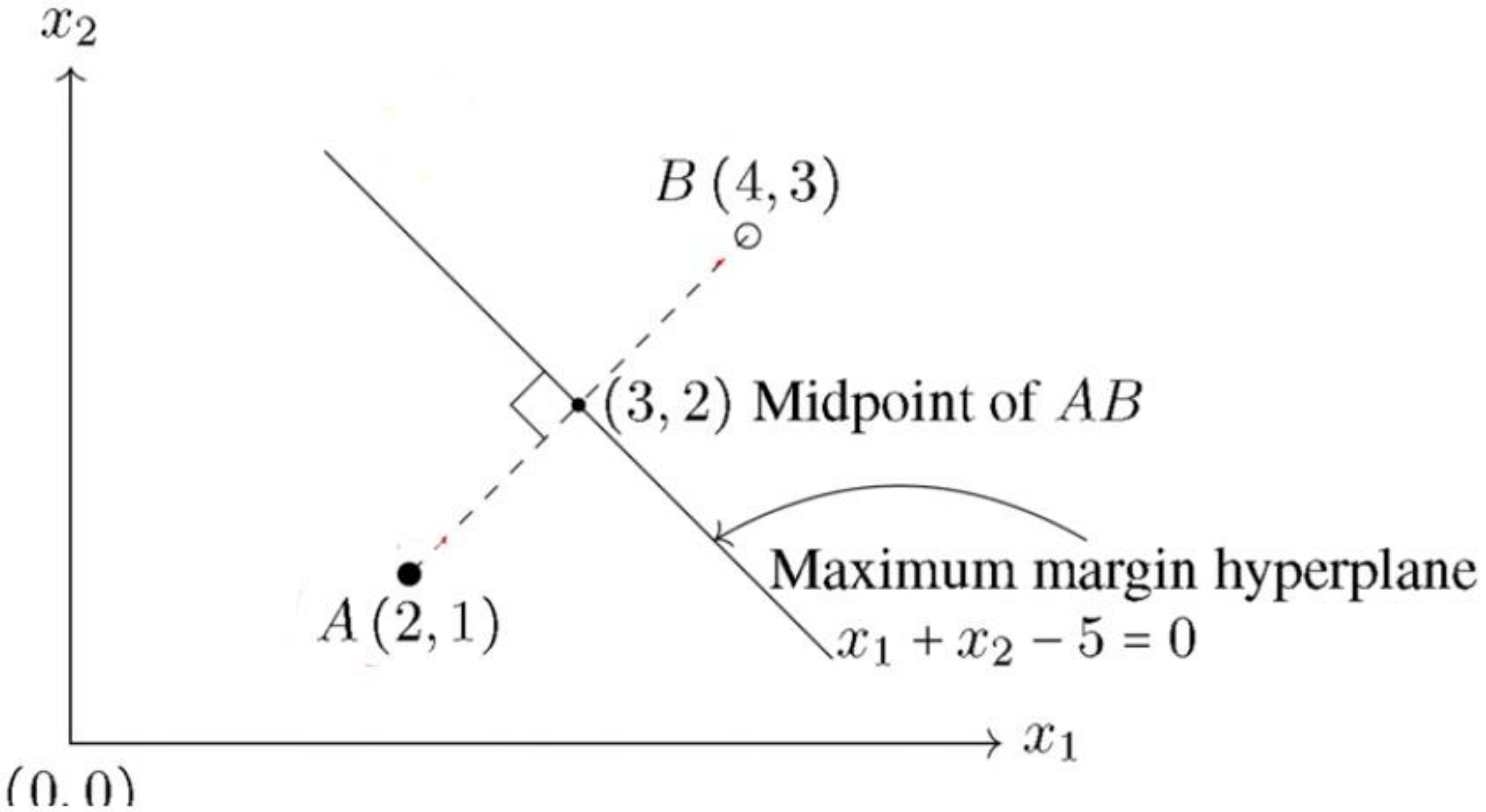
$$b = -\frac{5}{2}$$

X1	X2	Class
2	1	+1
4	3	-1

$$N = 2$$
$$\vec{x}_1 = (2, 1)$$
$$\vec{x}_2 = (4, 3)$$
$$y_1 = +1$$
$$y_2 = -1$$
$$\vec{\alpha} = (\alpha_1, \alpha_2)$$
$$\alpha_1 = \alpha_2$$
$$\alpha_1, \alpha_2 > 0$$

FINDING THE DECISION BOUNDARY

The equation $\boldsymbol{x_1 + x_2 - 5 = 0}$ is the perpendicular bisector of the line segment joining the points (2, 1) and (4, 3)



X1	X2	Class
2	1	+1
4	3	-1

$$N = 2$$
$$\vec{x}_1 = (2, 1)$$
$$\vec{x}_2 = (4, 3)$$
$$y_1 = +1$$
$$y_2 = -1$$
$$\vec{\alpha} = (\alpha_1, \alpha_2)$$
$$\alpha_1 = \alpha_2$$
$$\alpha_1, \alpha_2 > 0$$

Thank You ☺