Introduction to Machine Learning

Ass. Prof. Sara El-Metwally
Computer Science Department,
Faculty of Computers and Information,
Mansoura University, Egypt.

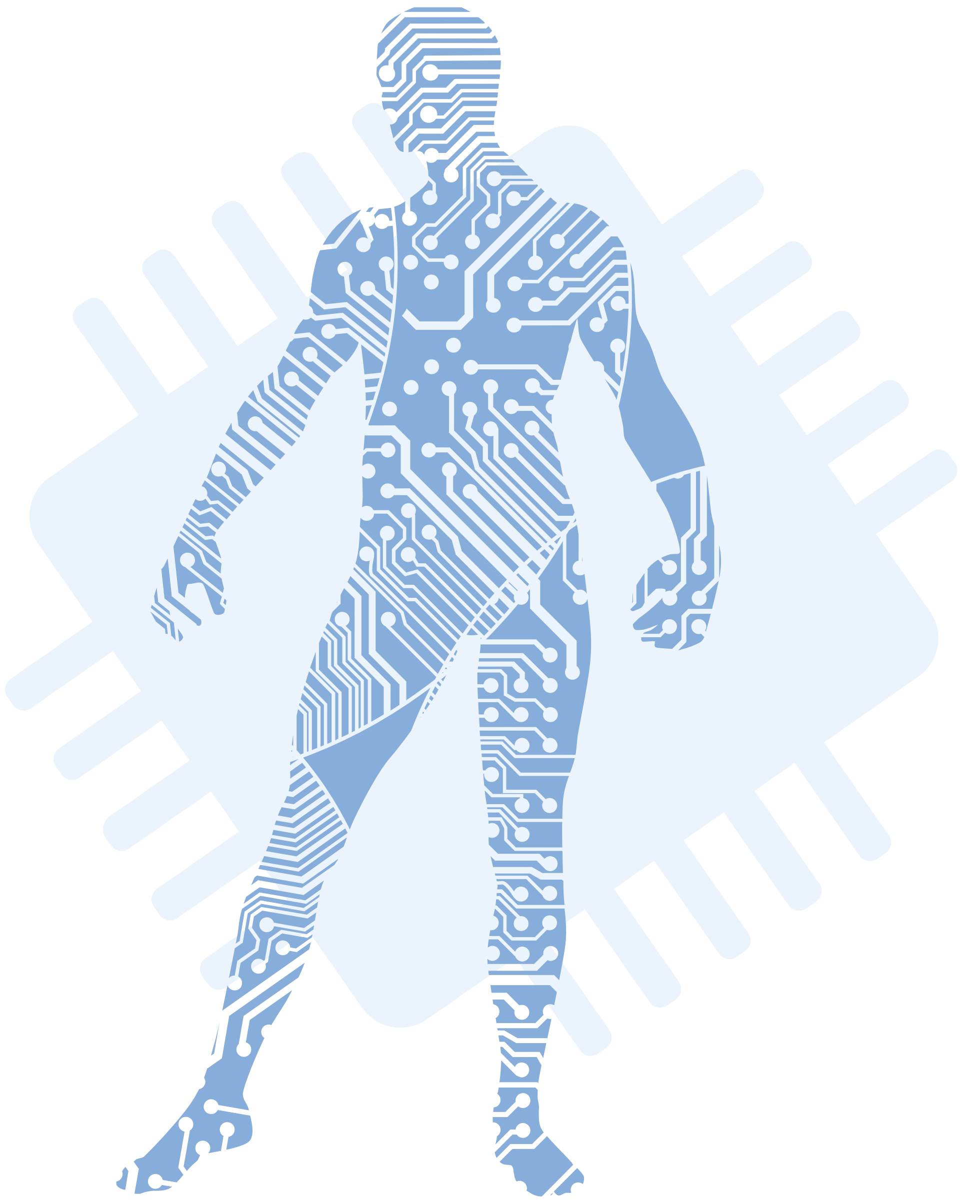
sarah_almetwally4@mans.edu.eg

Agenda

01 SVM

02 Math of SVM

03 Kernel of SVM



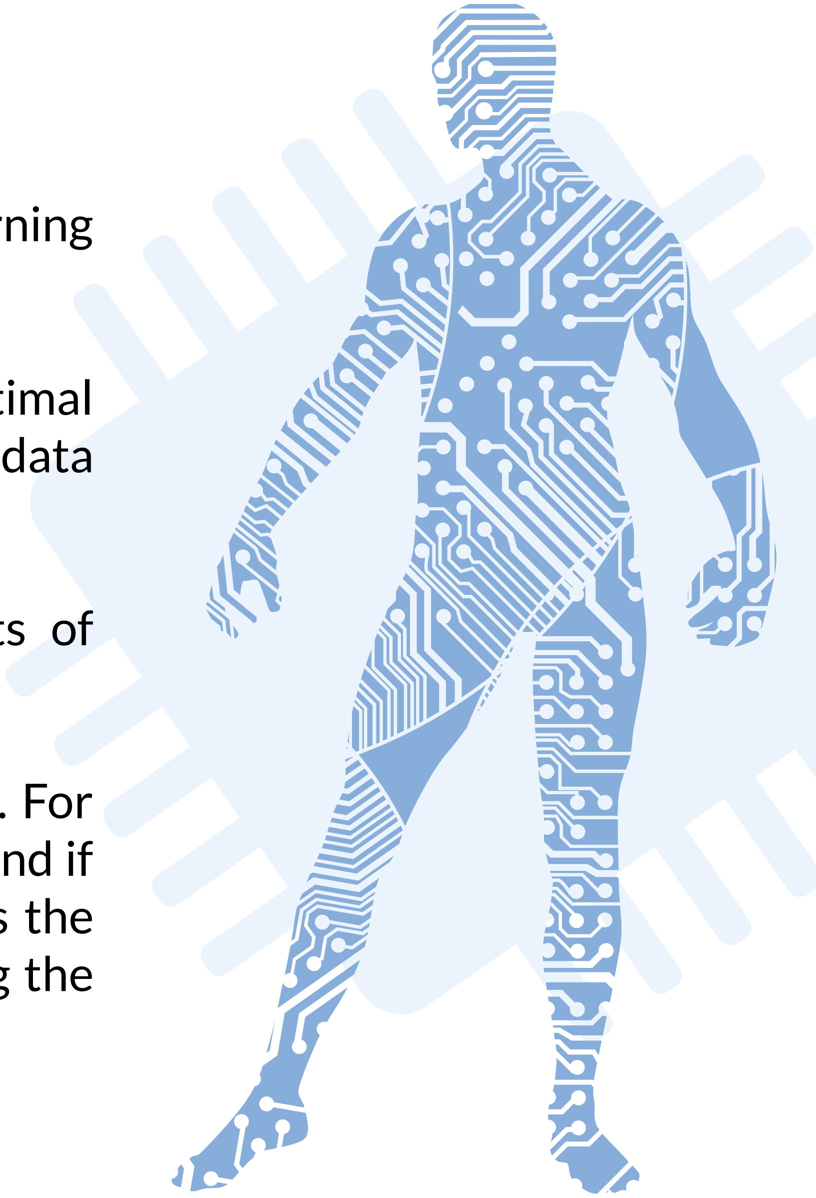
SVM

- ❑ A Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression tasks.

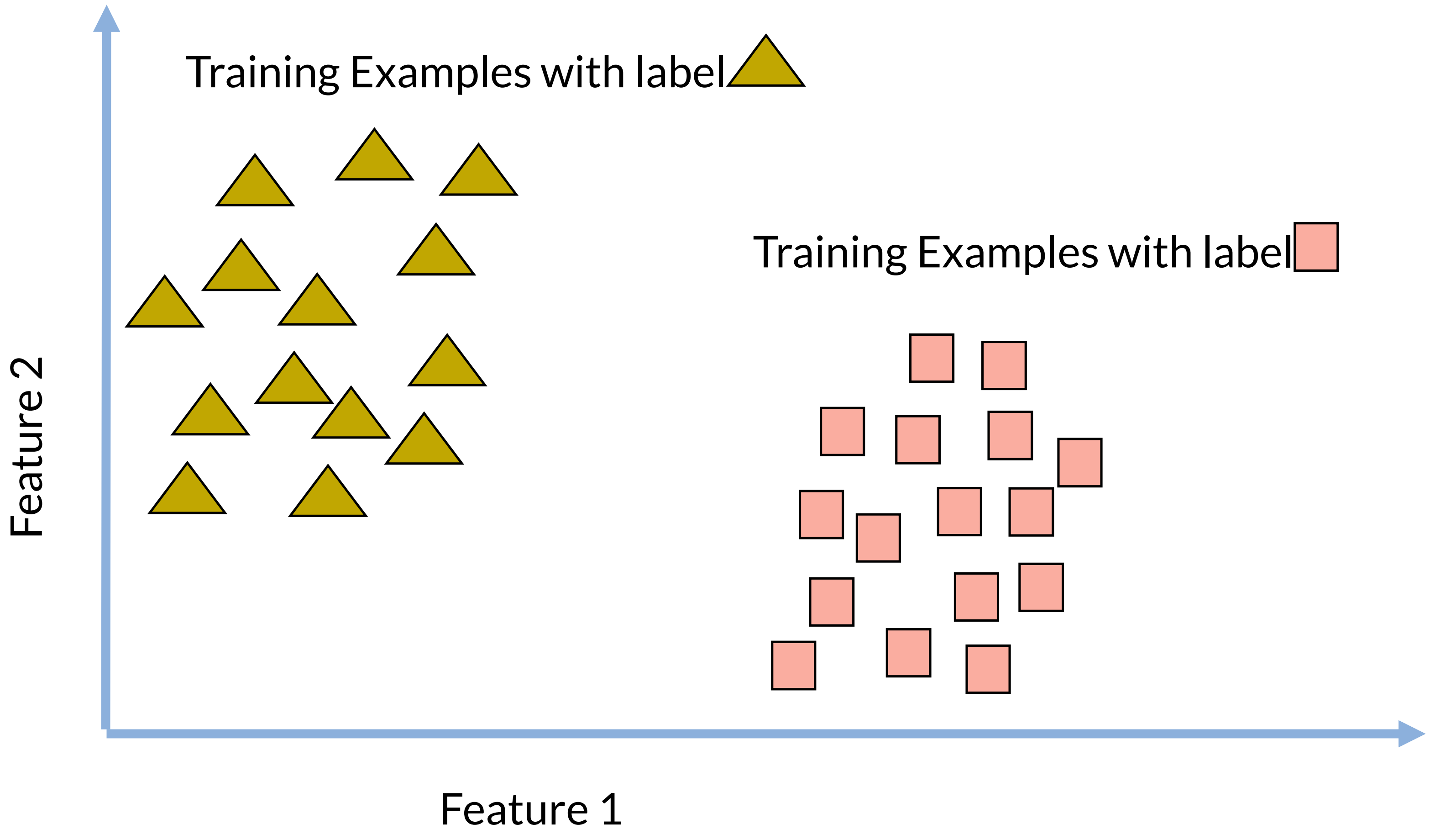
- ❑ The primary objective of the SVM algorithm is to identify the optimal hyperplane in an N-dimensional space that can effectively separate data points into different classes in the feature space.

- ❑ The algorithm ensures that the margin between the closest points of different classes, known as support vectors, is maximized.

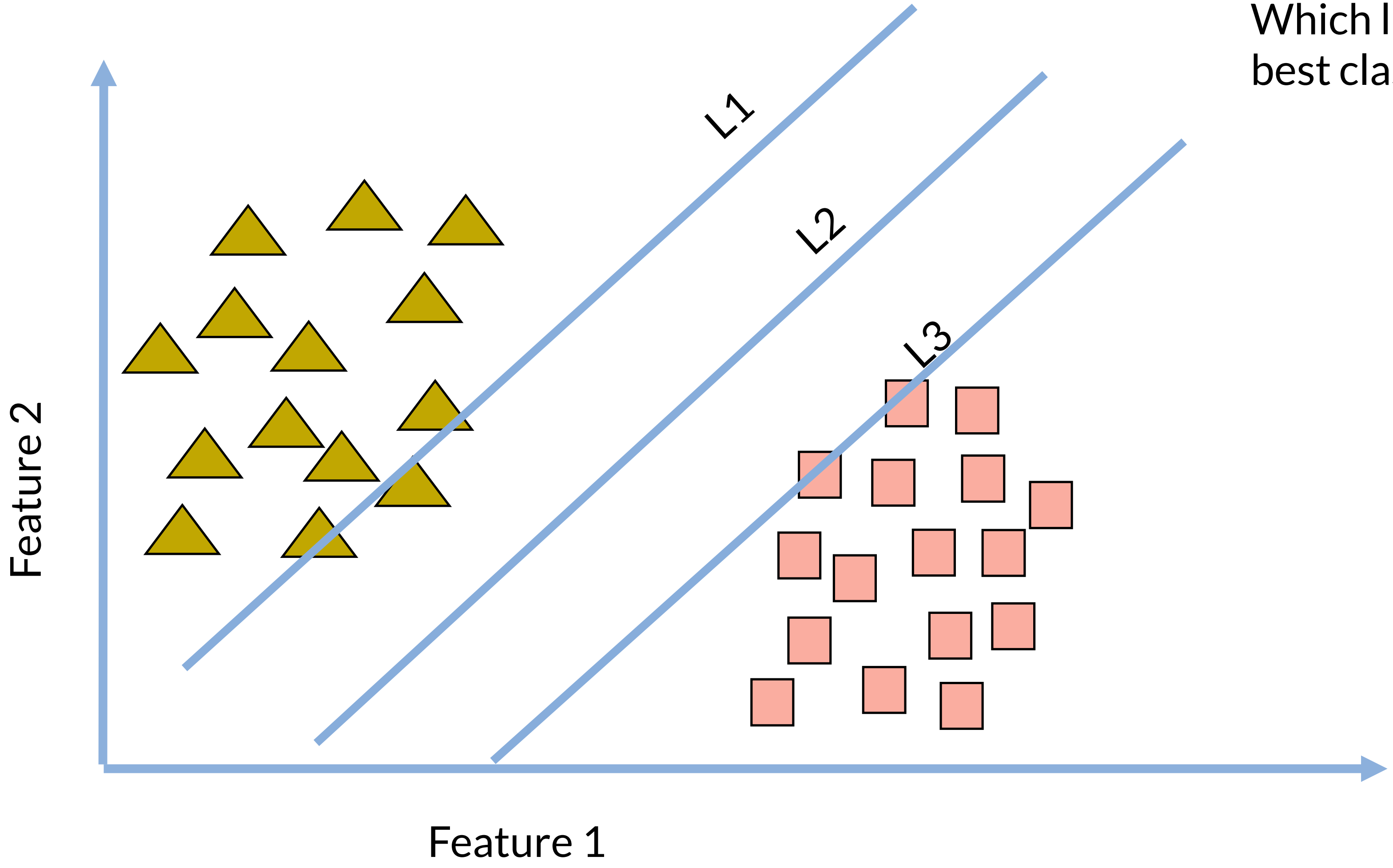
- ❑ The dimension of the hyperplane depends on the number of features. For instance, if there are two input features, the hyperplane is simply a line, and if there are three input features, the hyperplane becomes a 2-D plane. As the number of features increases beyond three, the complexity of visualizing the hyperplane also increases.



SVM

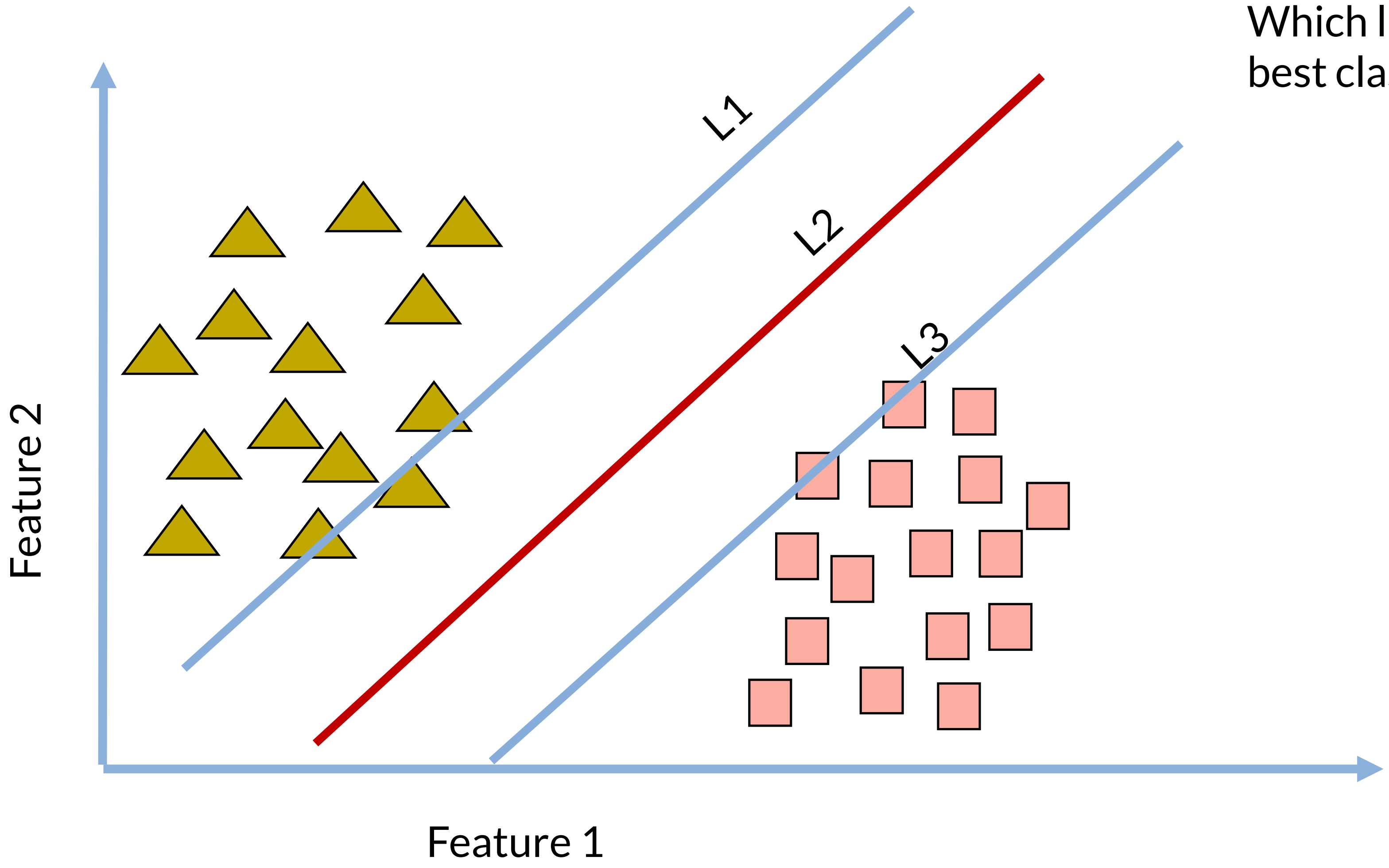

Training Examples with label
Training Examples with label
Feature 2
Feature 1

SVM



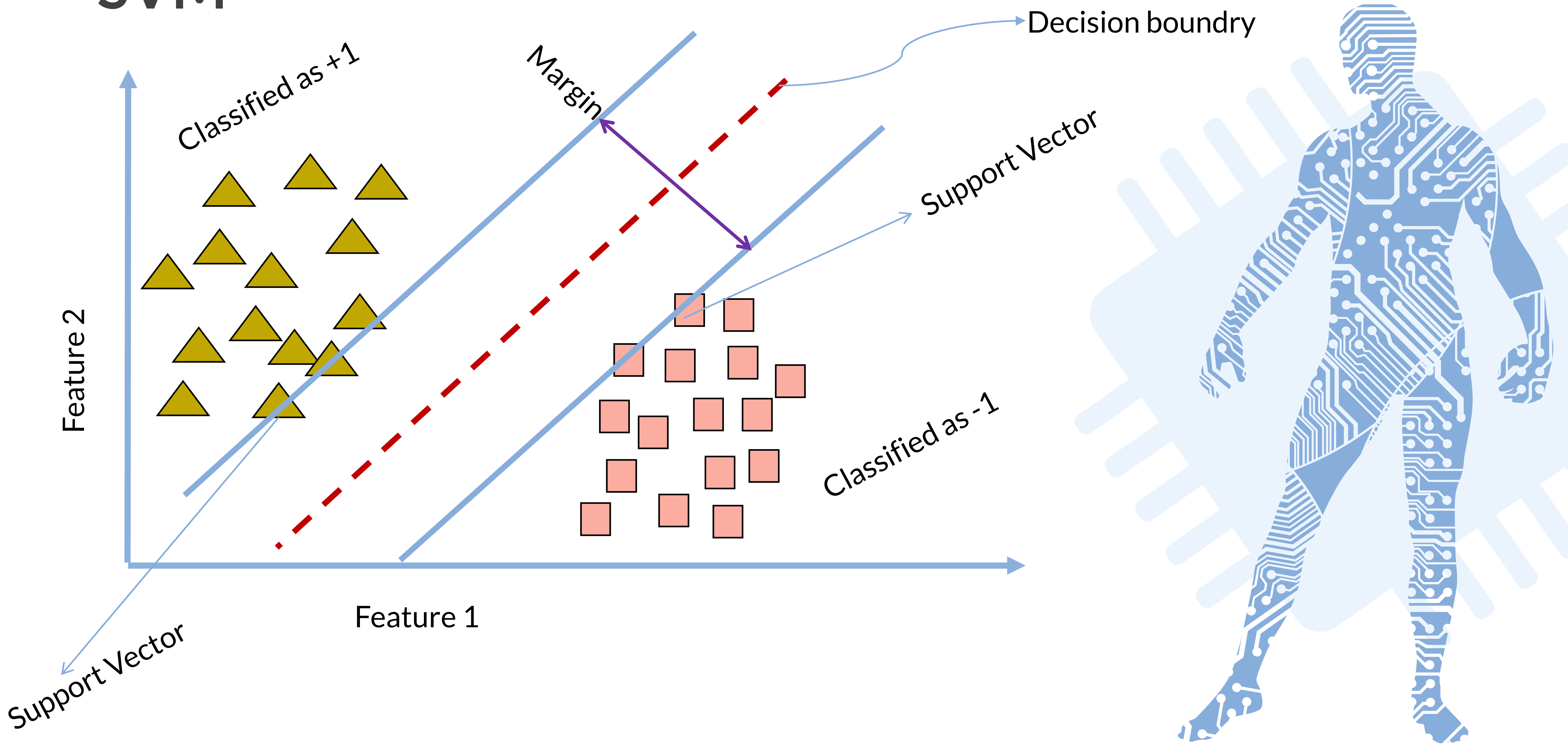
Which line is used as the best classifier line ?

SVM



Which line is used as the best classifier line ?

SVM


Decision boundry
Classified as +1
Margin
Support Vector
Feature 2
Classified as -1
Feature 1
Support Vector

SVM

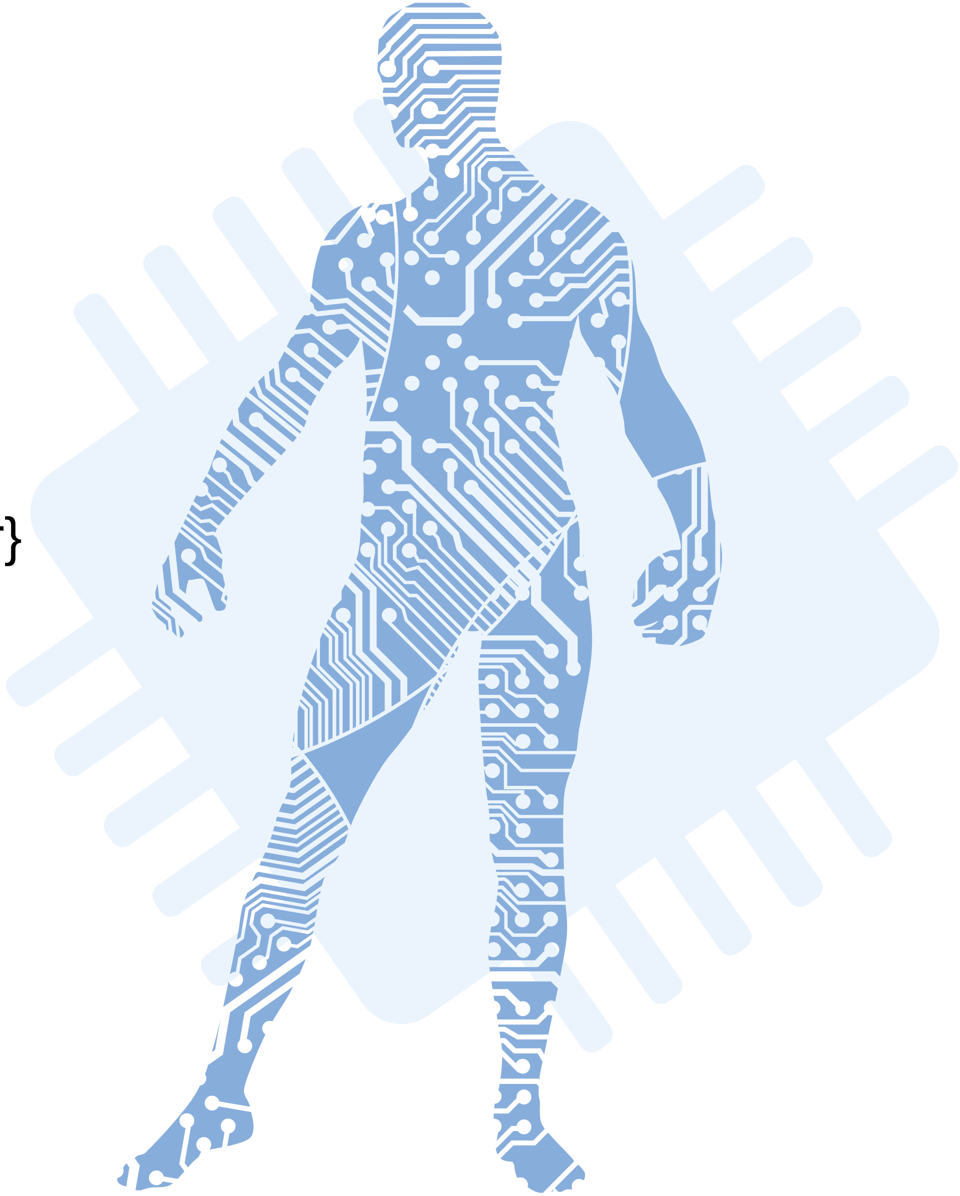
Training Data

$$\begin{bmatrix} \overrightarrow{X_1} \\ \overrightarrow{X_2} \\ - \\ - \\ \overrightarrow{X_n} \end{bmatrix} \qquad \begin{bmatrix} Y_1 \\ Y_2 \\ - \\ - \\ Y_n \end{bmatrix}$$

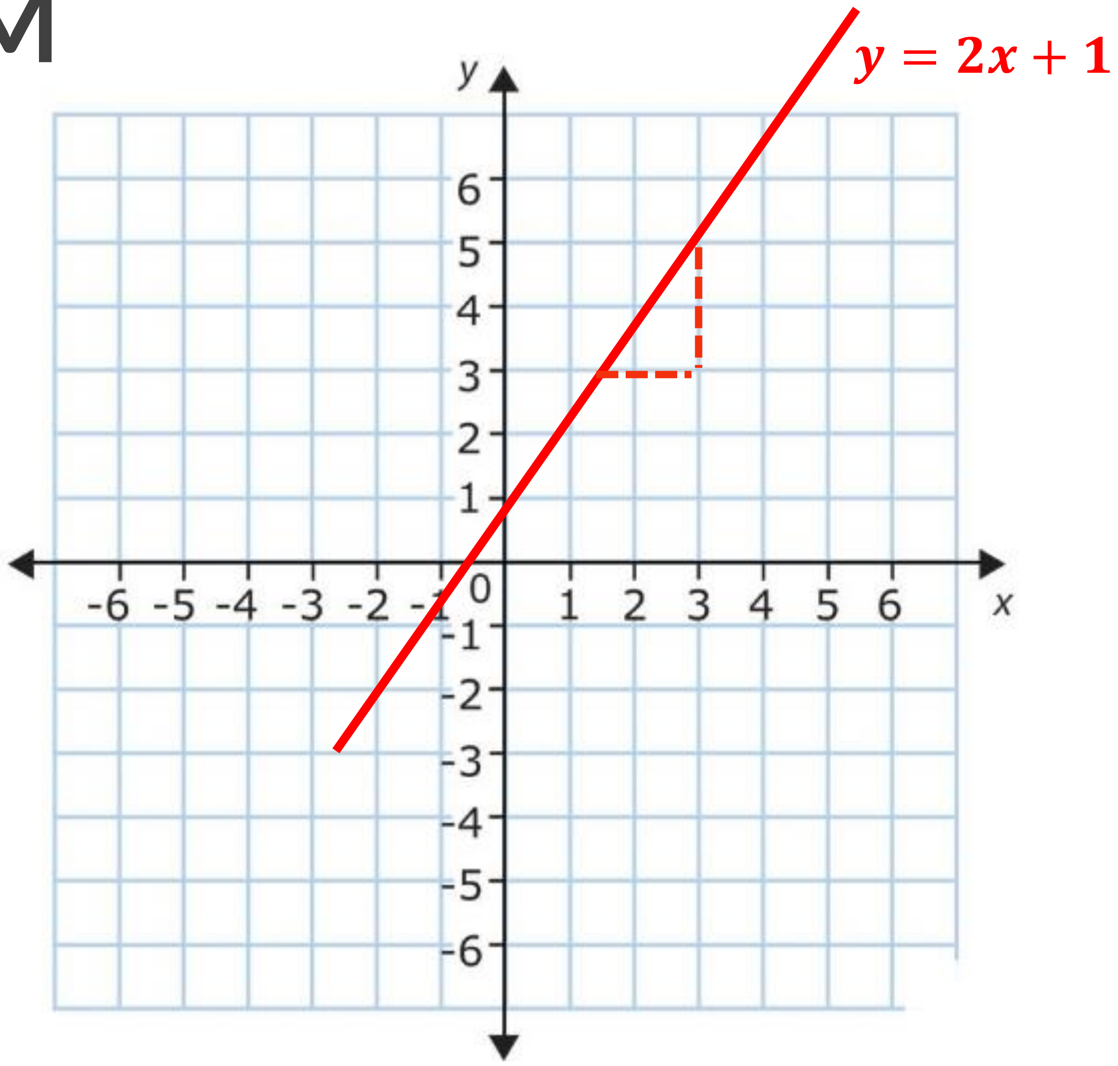
Patients Classes

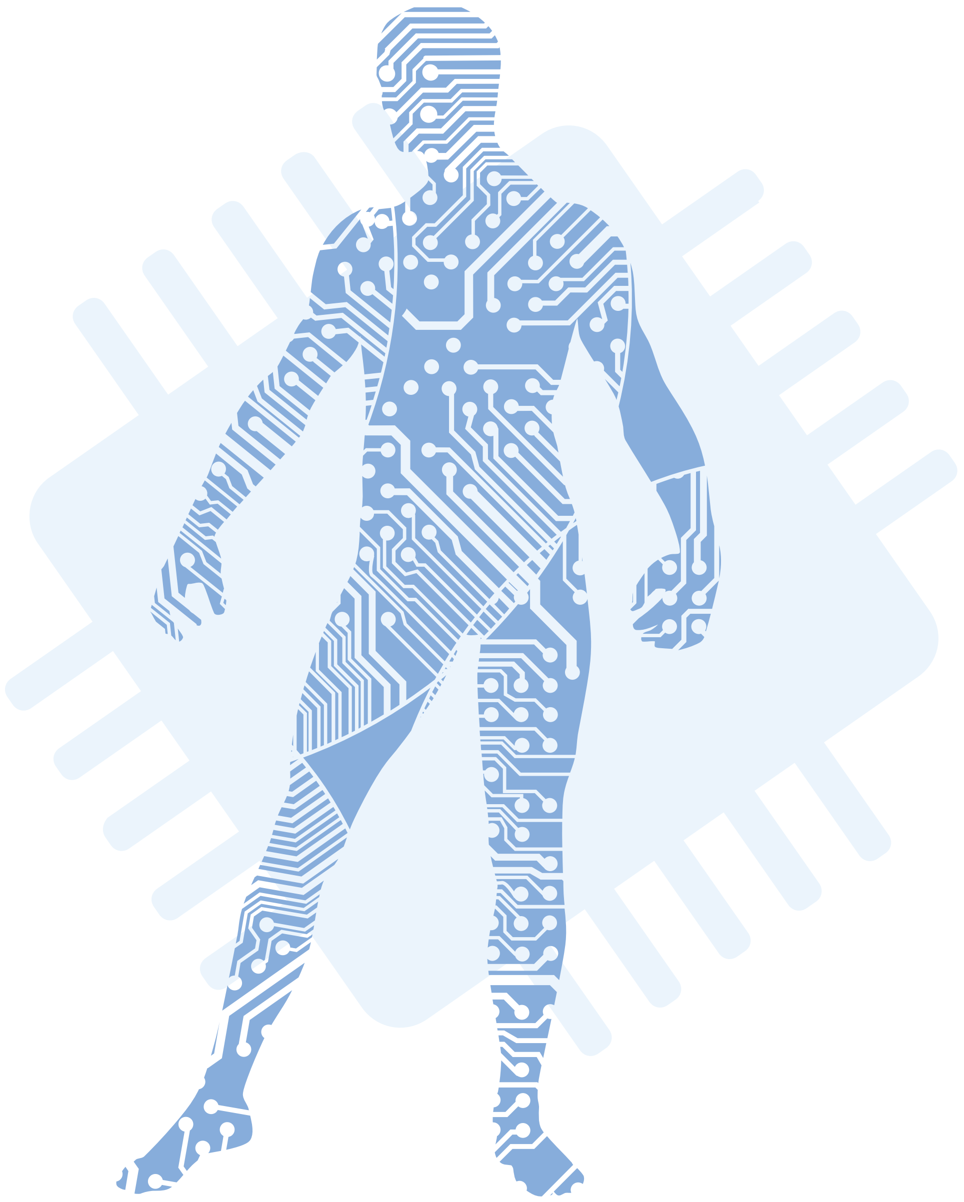
$\{1, -1\}$

{cancer, non-cancer}



SVM





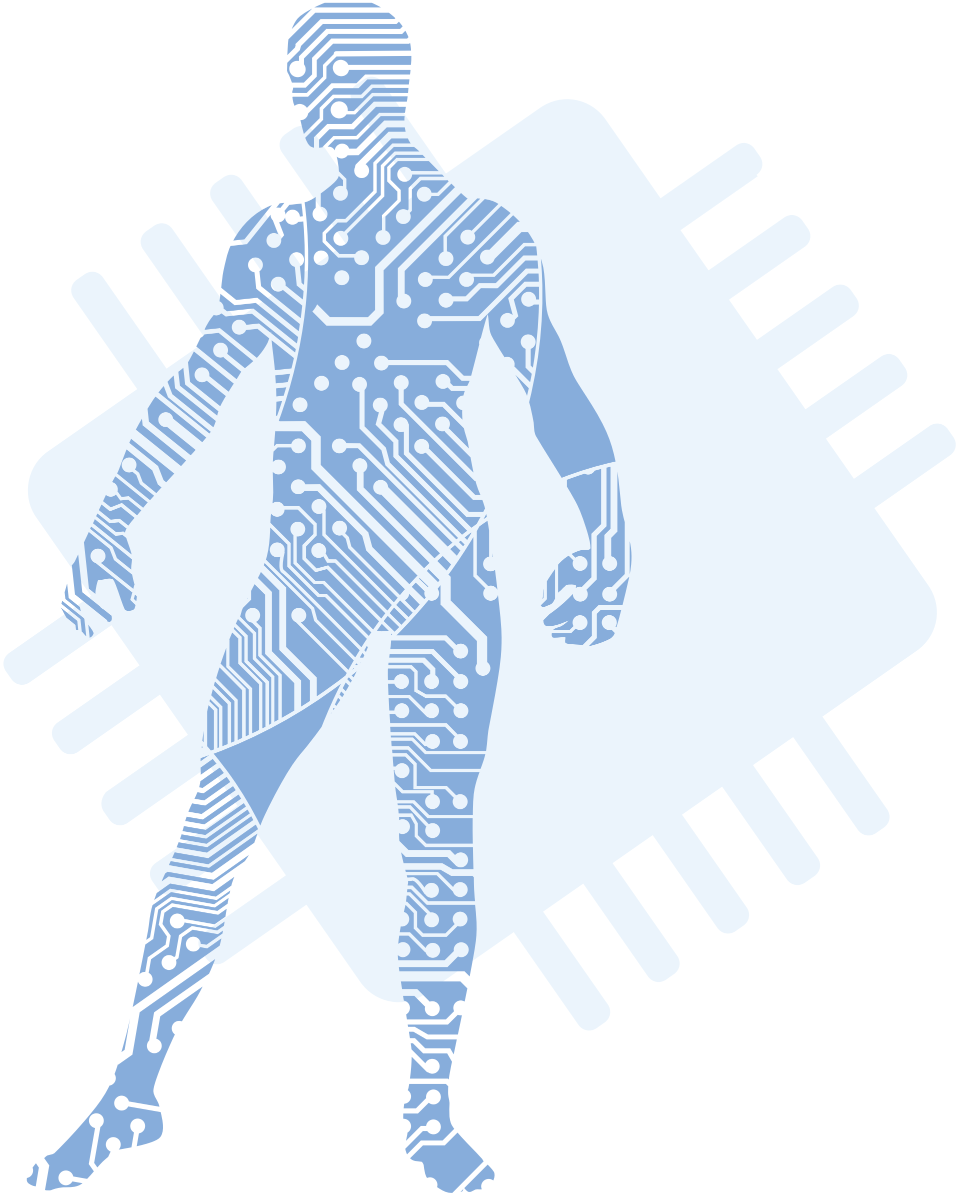
SVM

$$y = mx + b$$

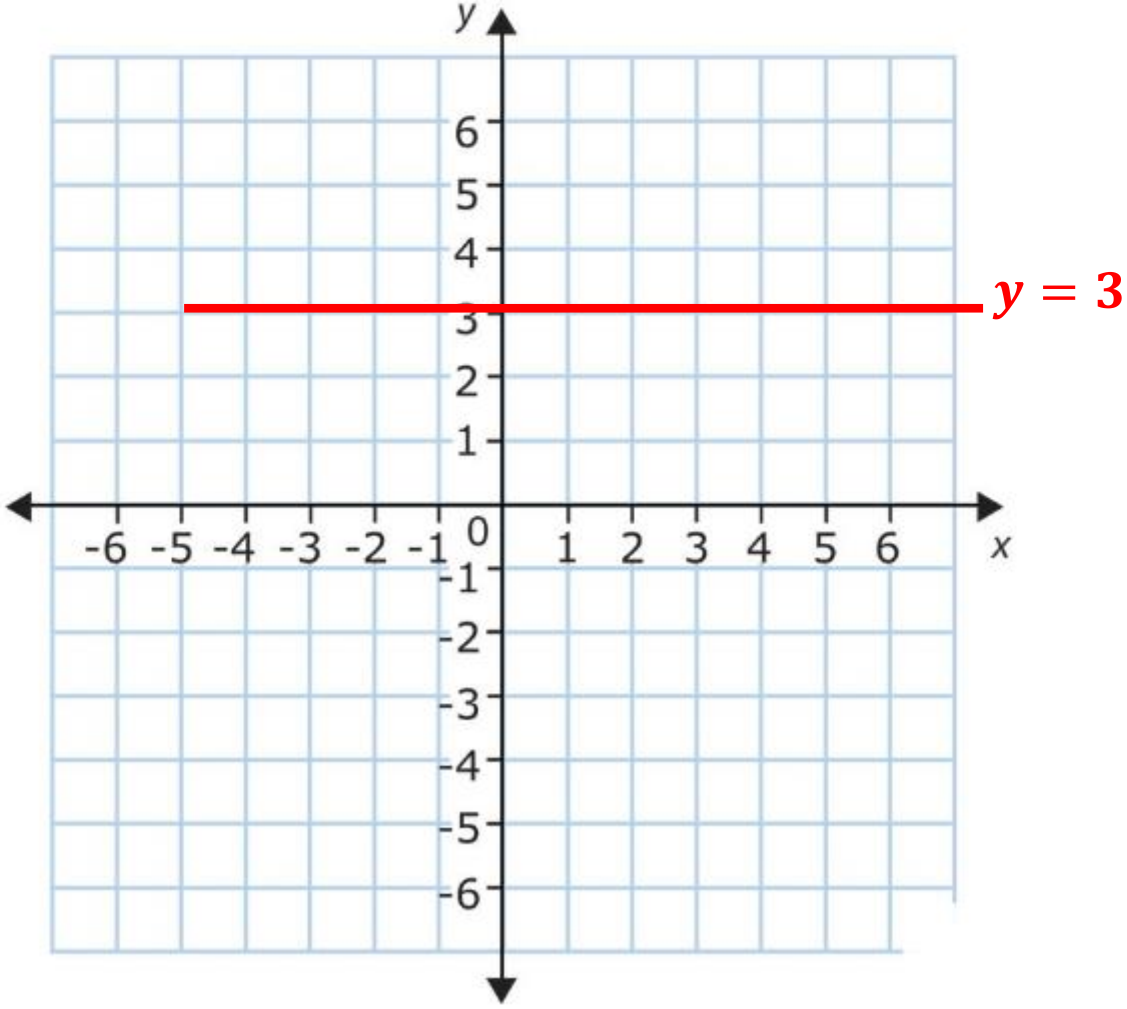
Equation of Straight Line

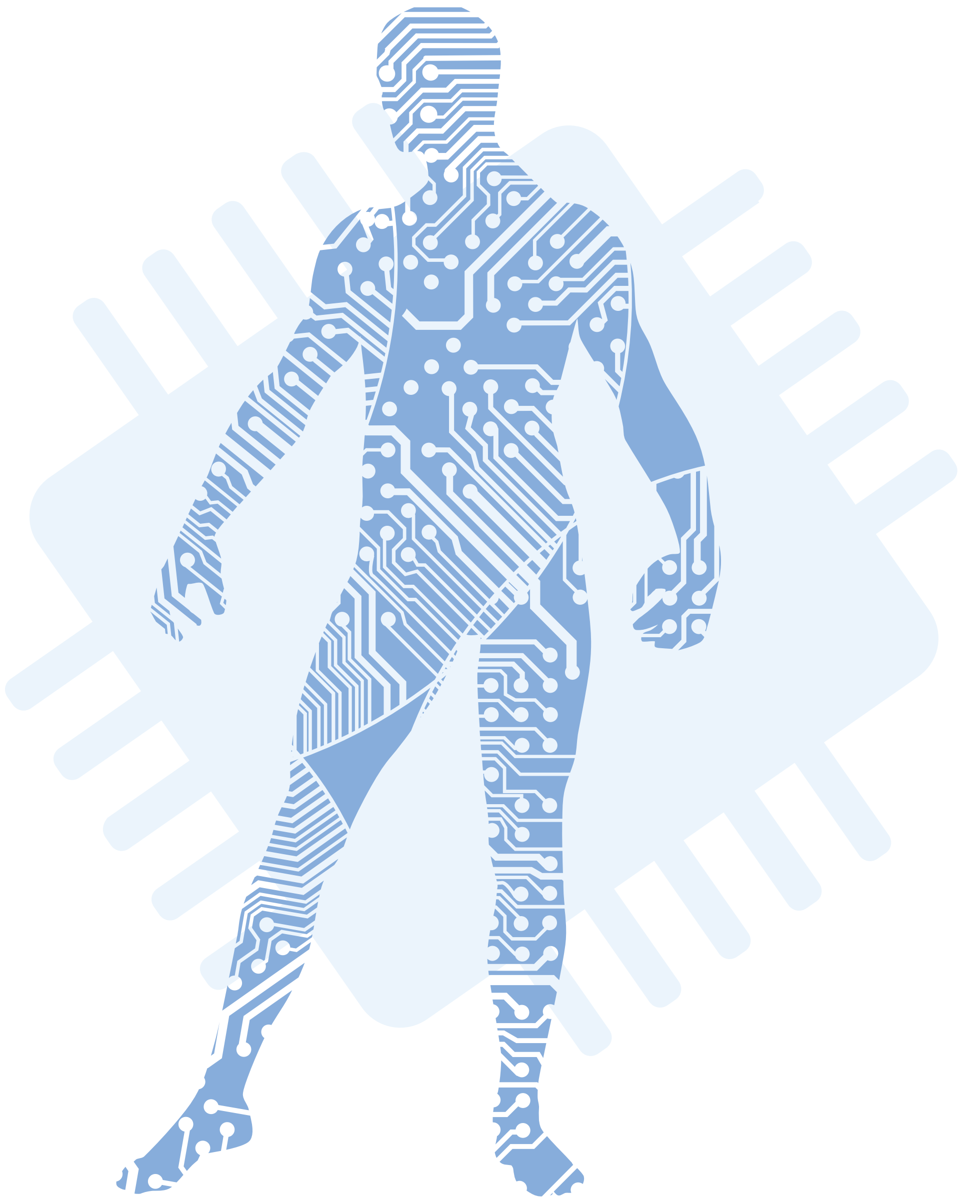
$$W^T x + b = 0$$

Equation of decision boundary line in SVM

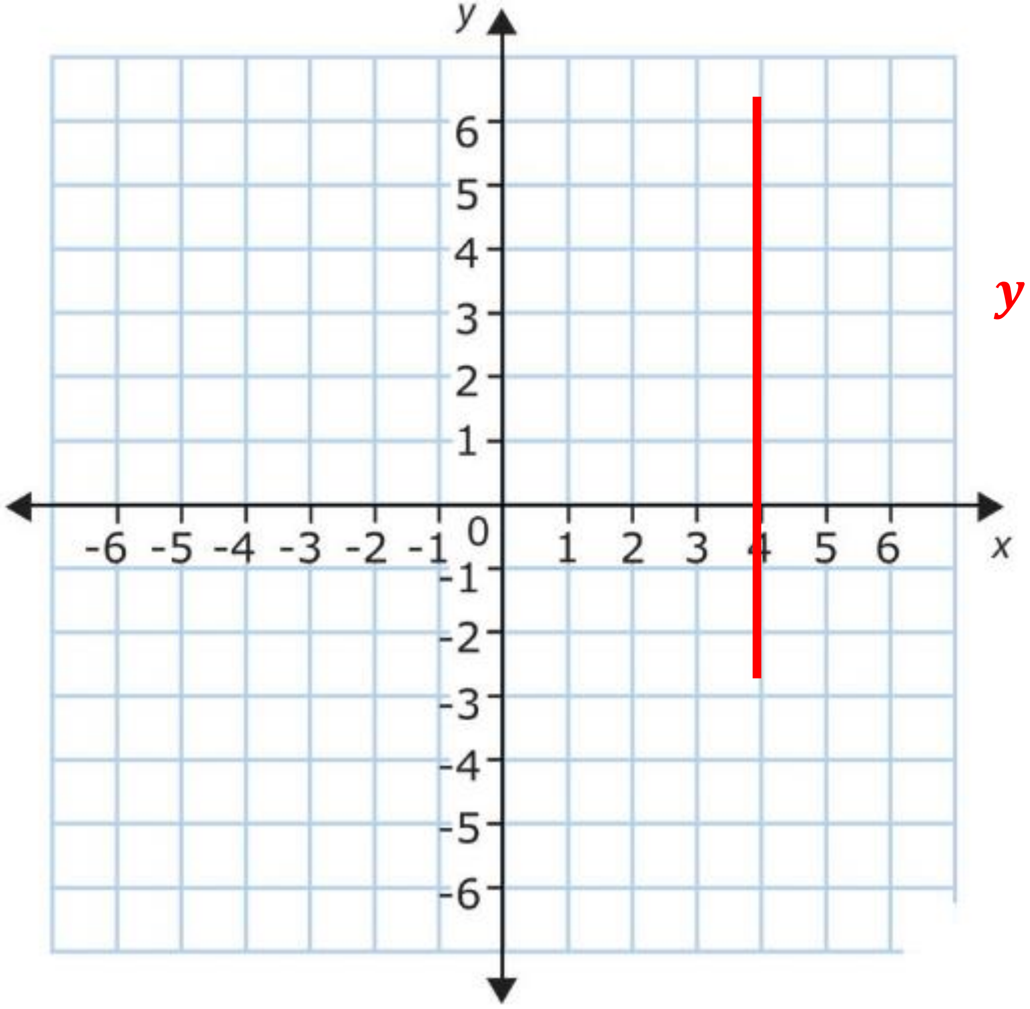


SVM

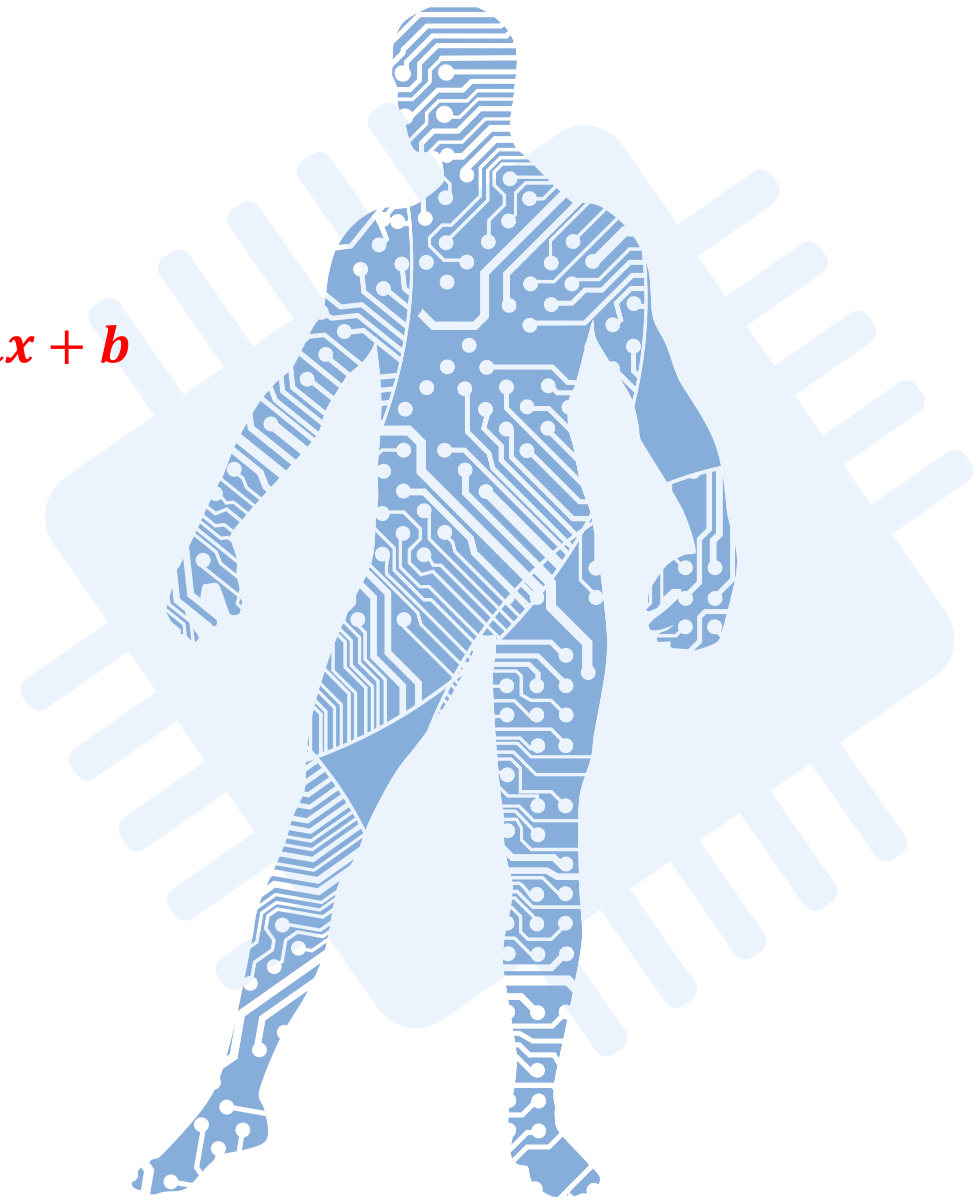




SVM

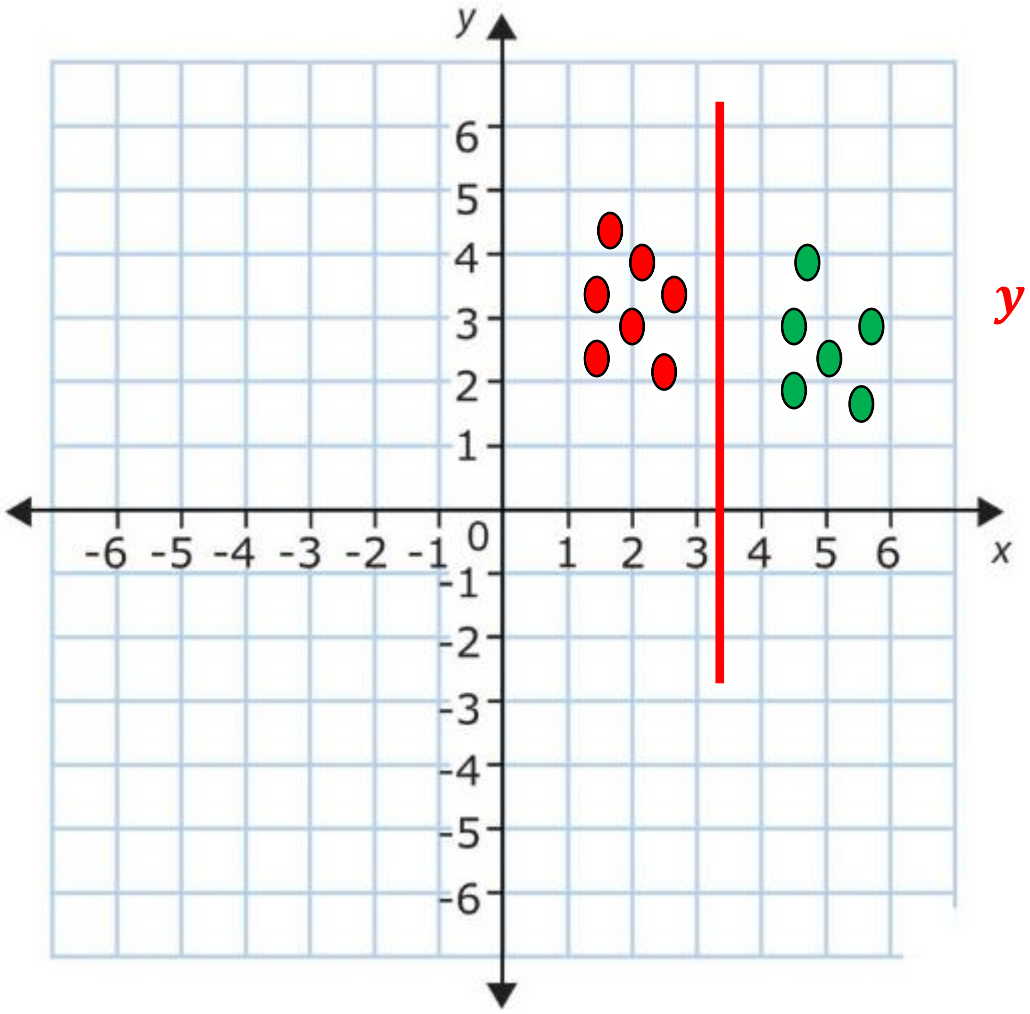


$$y = mx + b$$

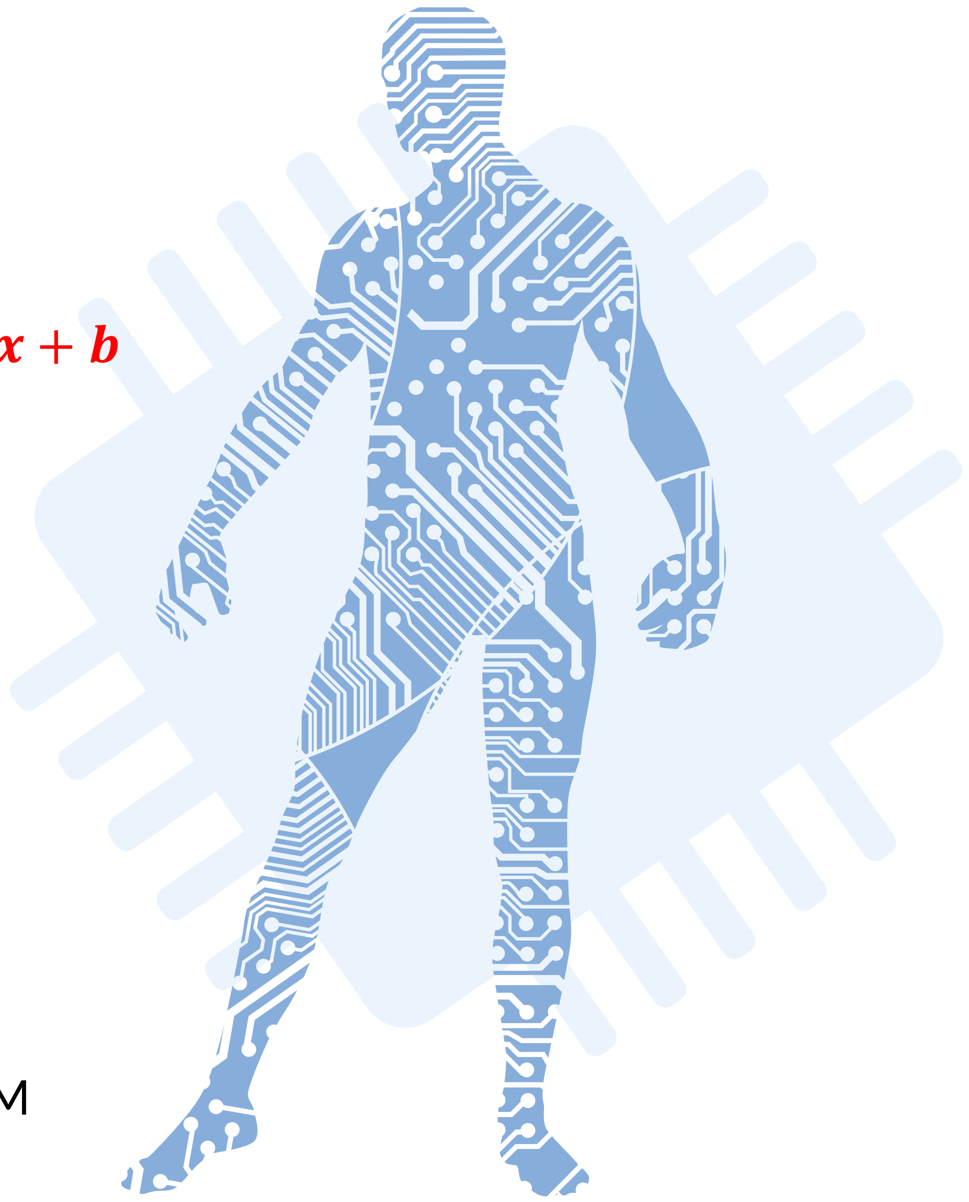


- The equation of straight line can not be used to describe the vertical line because the slope will be infinitely large

SVM



$$y = mx + b$$



- The vertical lines can be used as a decision boundary line in SVM to classify the data points into two groups. So the equation of straight line is not suitable to use in SVM.

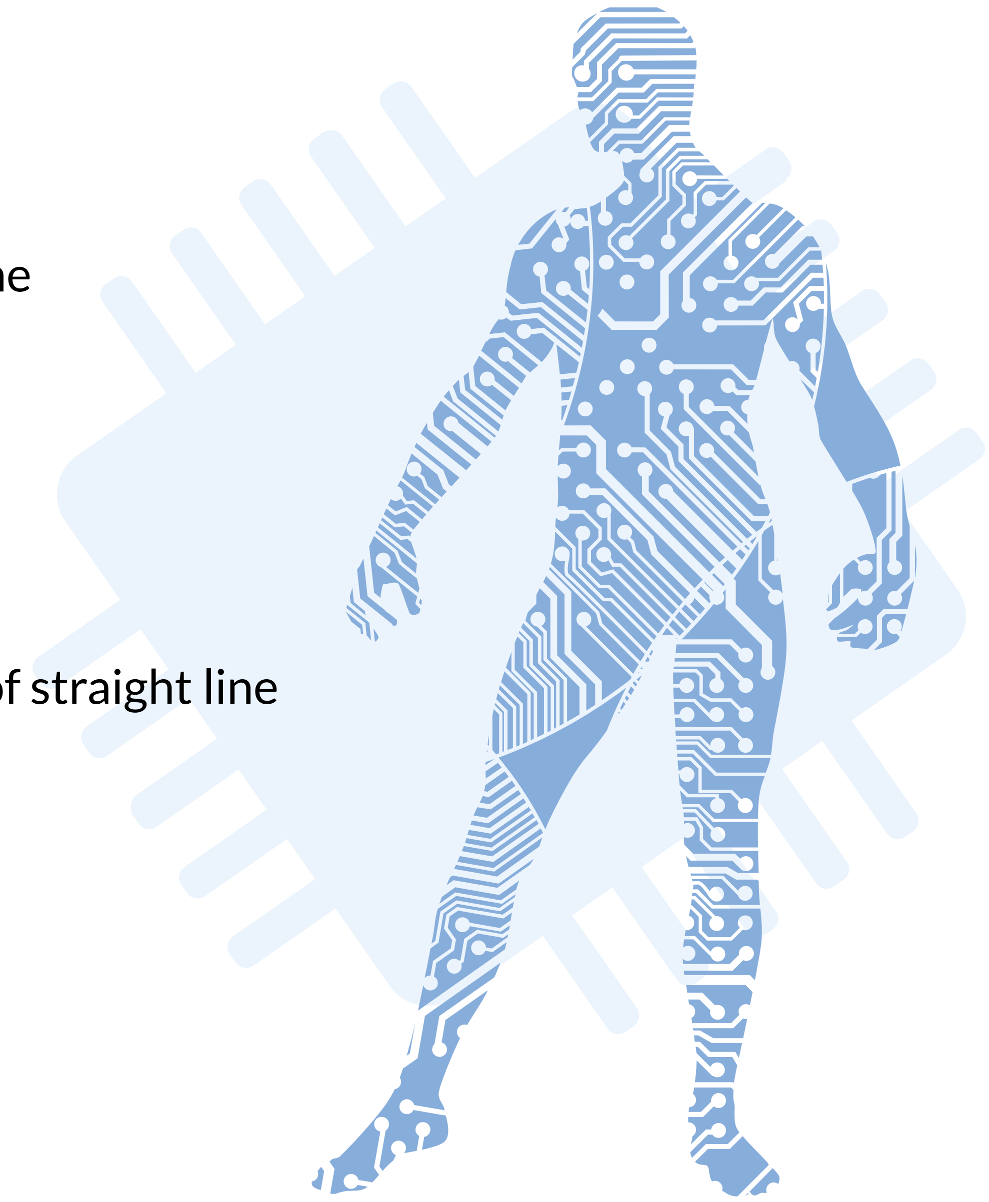
SVM

$$y = mx + b$$

Equation of Straight Line

$$Ax + By + C = 0$$

General form equation of straight line



SVM

$$y = mx + b$$

Equation of Straight Line

$$Ax + By + C = 0$$

General form equation of straight line

$$y = -\frac{A}{B}x - \frac{C}{B}$$

Slope and intercept

Note: they are the same

$$y = 0.5x + 1$$

$$y - 0.5x - 1 = 0$$

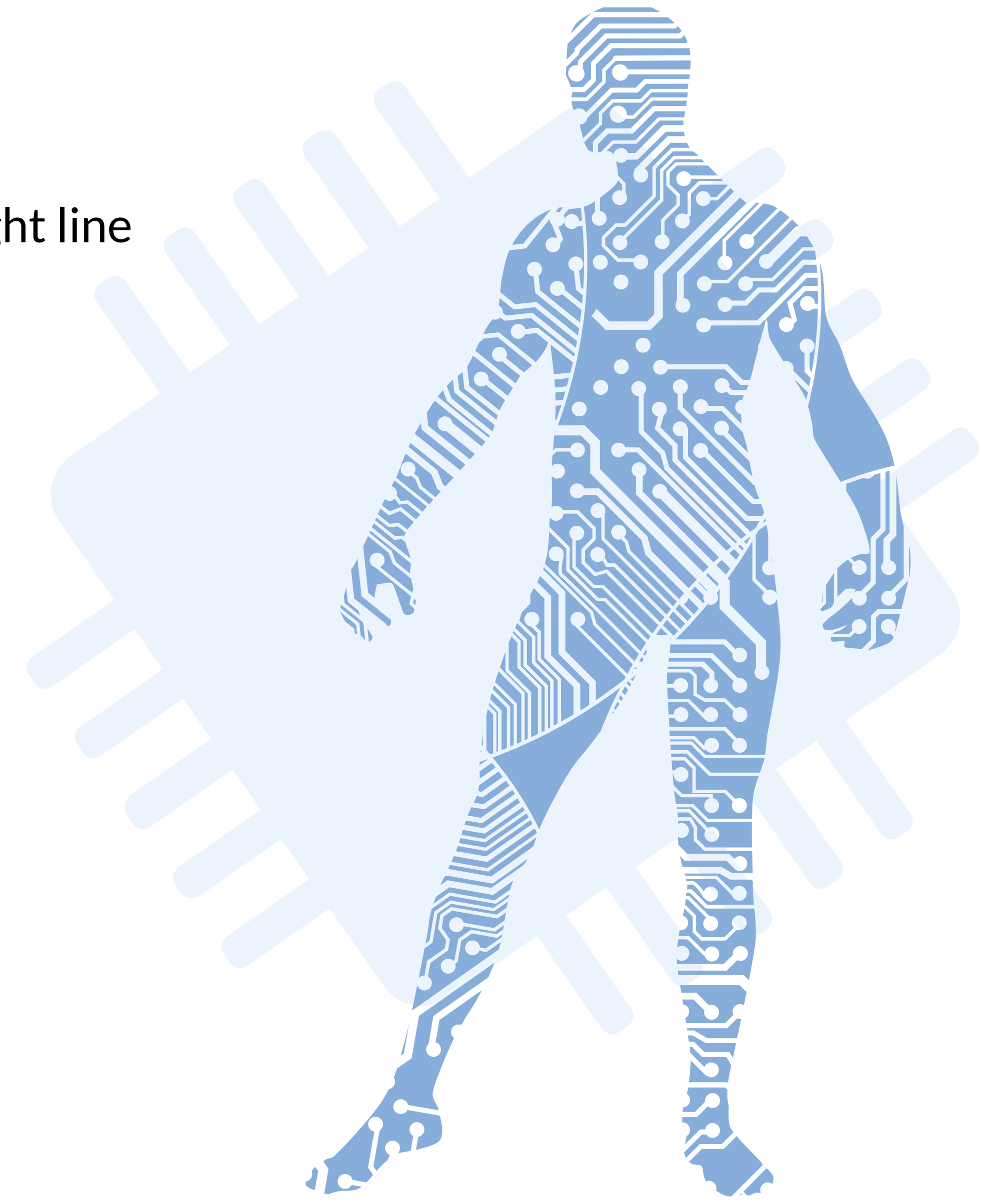
$$-0.5x + y - 1 = 0 \qquad \times 4$$

$$-2x + 4y - 4 = 0$$

Solve it for y

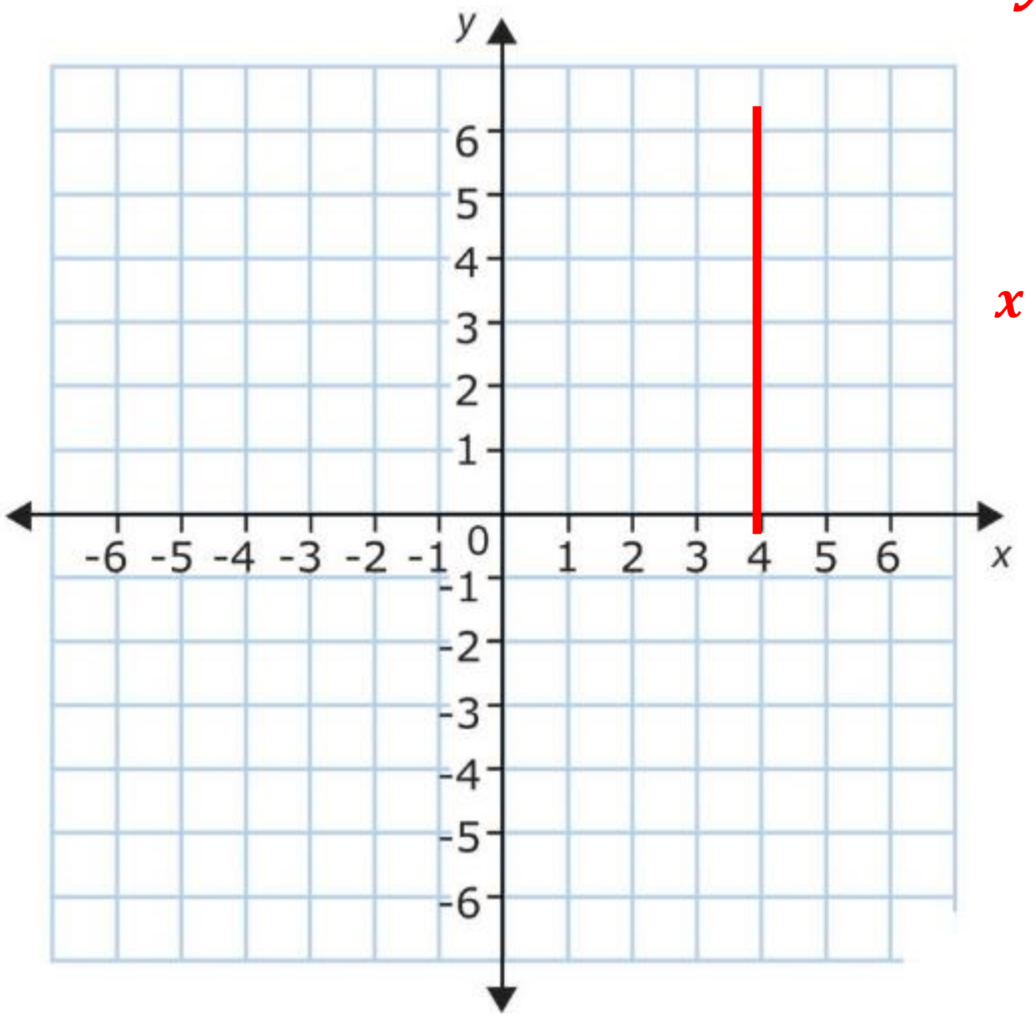
$$4y = 2x + 4$$

$$y = 0.5x + 1$$

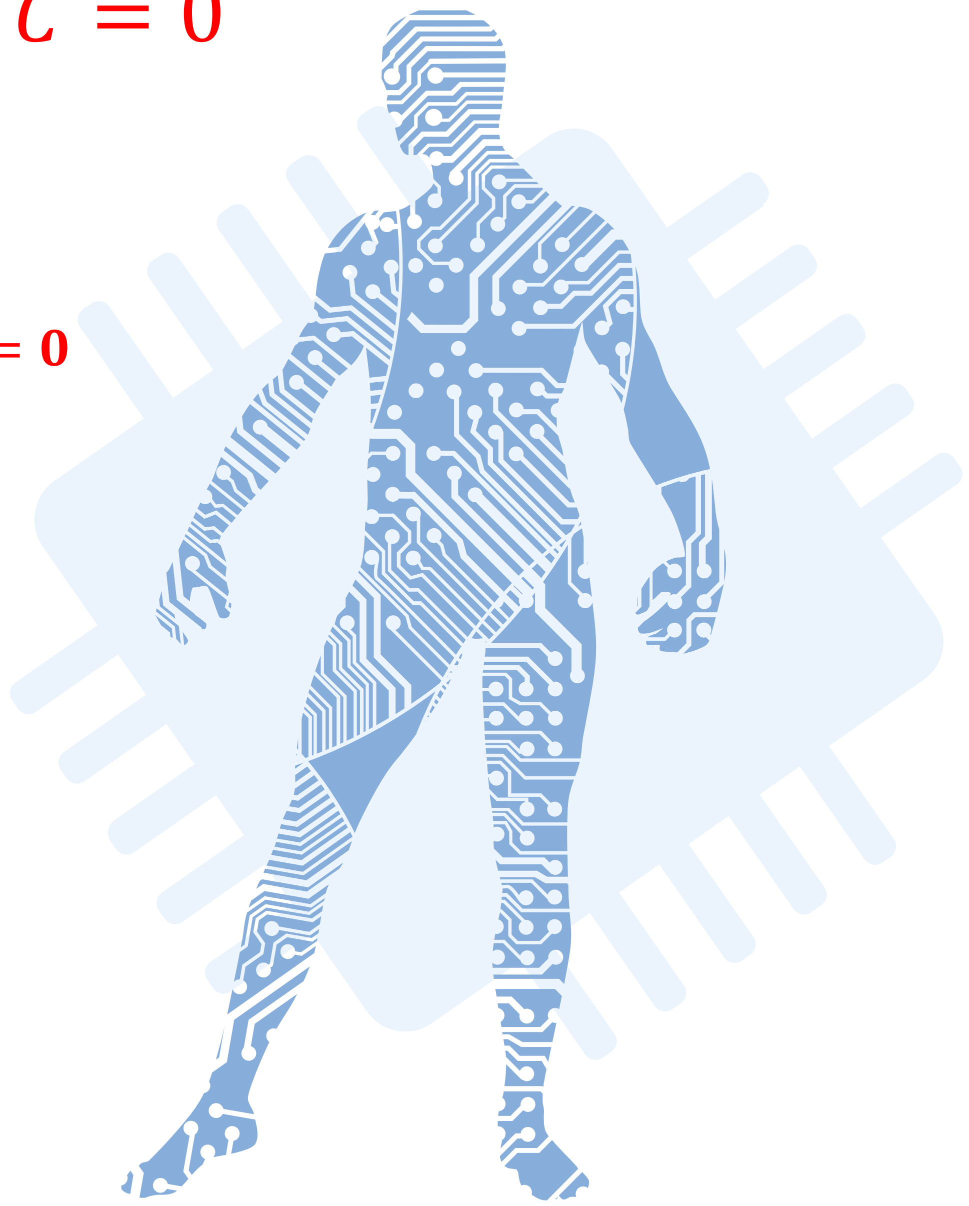


SVM

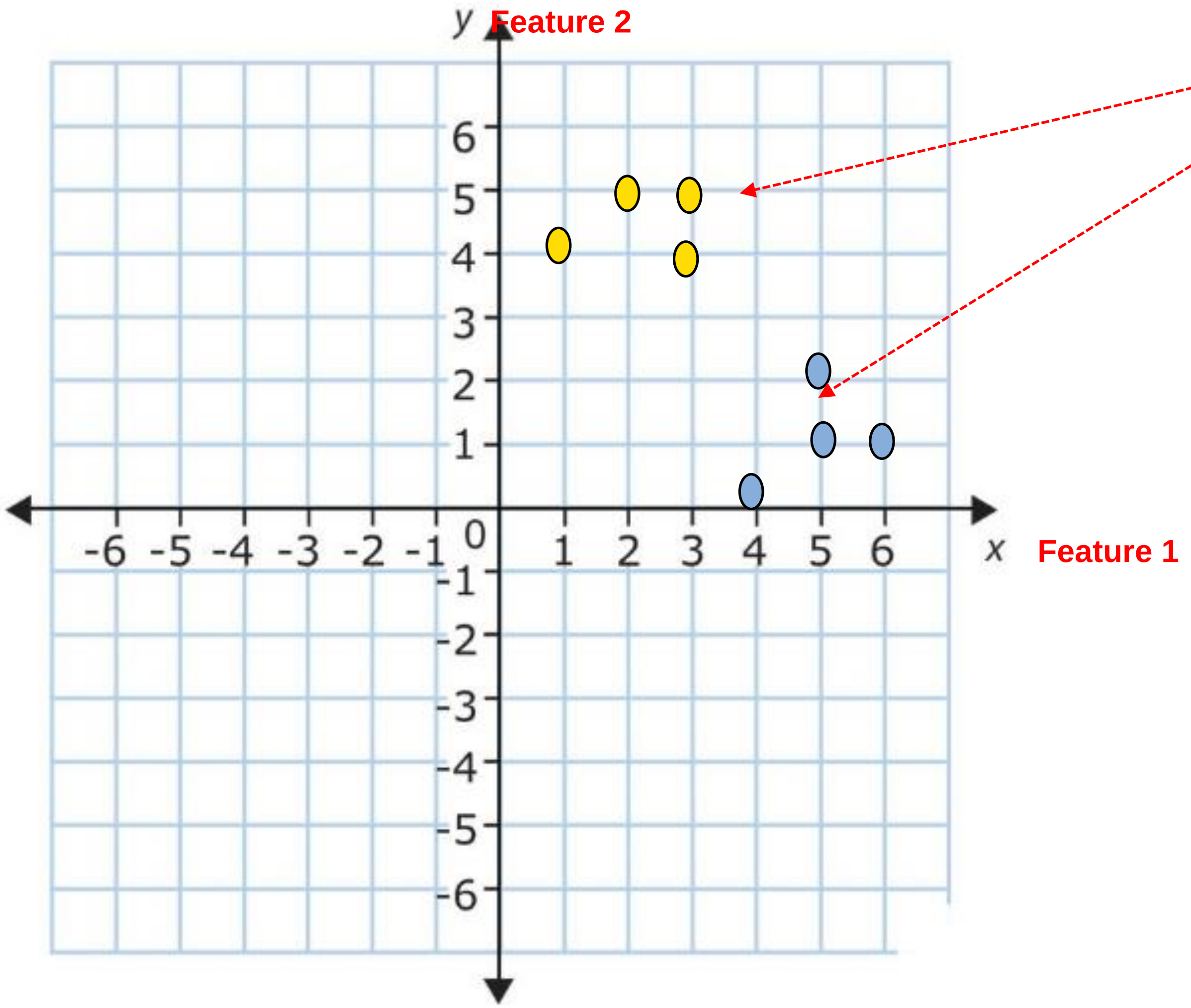
$$Ax + By + C = 0$$



$$\boldsymbol{x - 4 = 0}$$



SVM

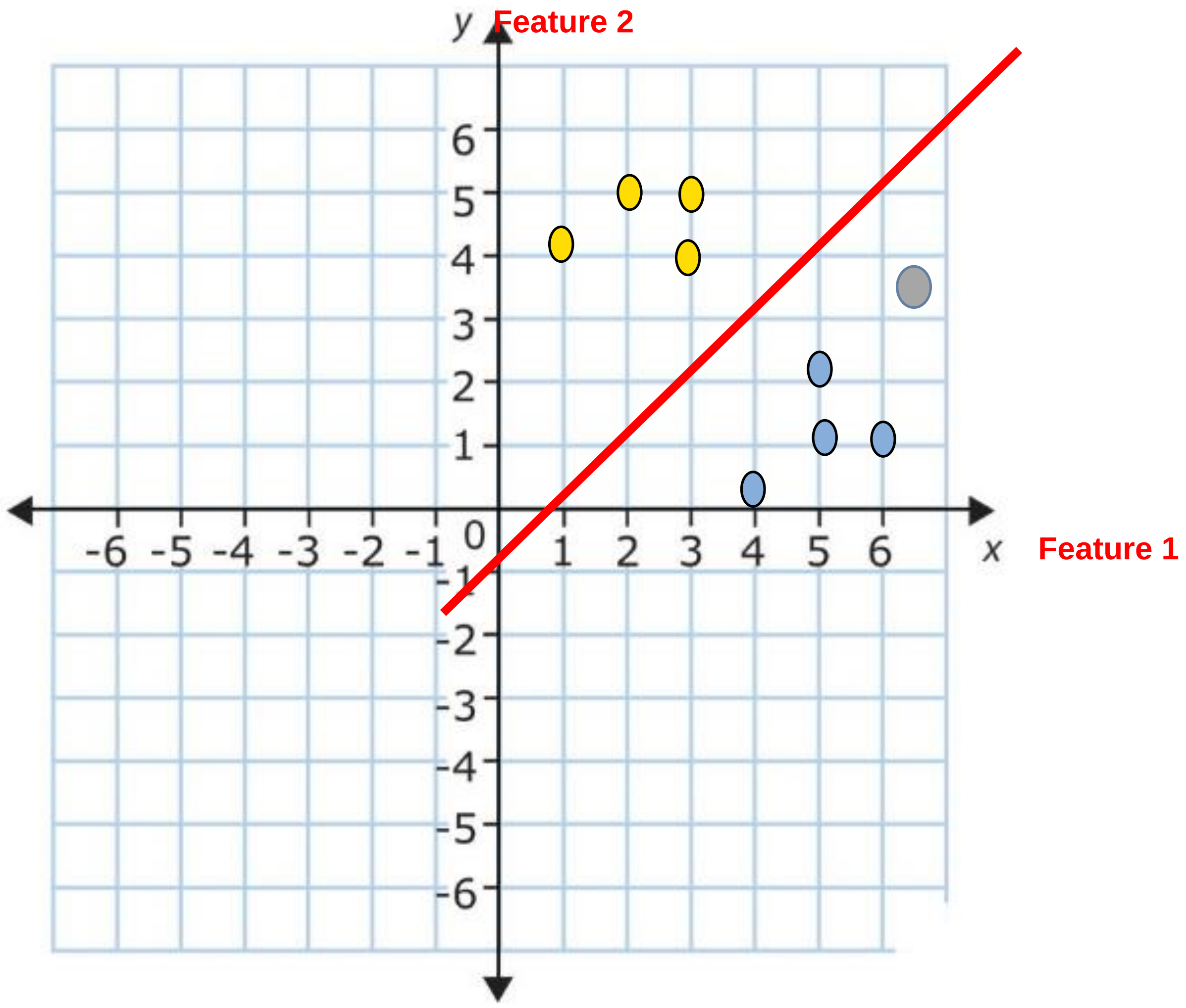


Training Data

Feature x	Feature y	Label
1	4	Cancer
2	5	Cancer
3	5	Cancer
3	4	Cancer
6	1	Normal
4	0	Normal
5	2	Normal
5	1	Normal

Two features, two dimensional space with training model represents by a line that separates the data points accurately
Three features, three dimensional space, which creates a plane or hyper-plane

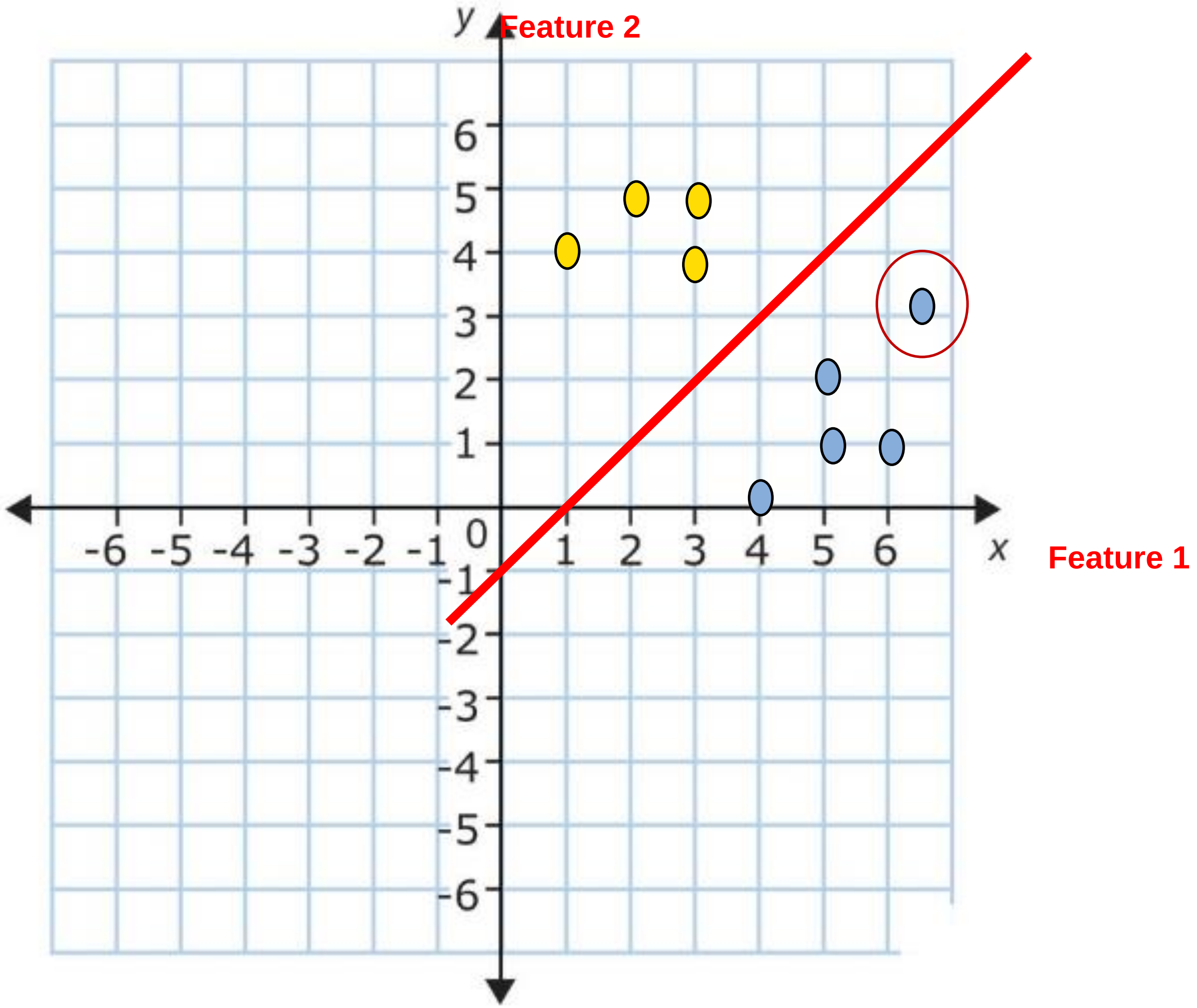
SVM



Feature x	Feature y	Label
1	4	Cancer
2	5	Cancer
3	5	Cancer
3	4	Cancer
6	1	Normal
4	0	Normal
5	2	Normal
5	1	Normal

By using the line that represents the trained model, a new data point (gray point) could be classified
If the data point below the line, it will be classified as Normal

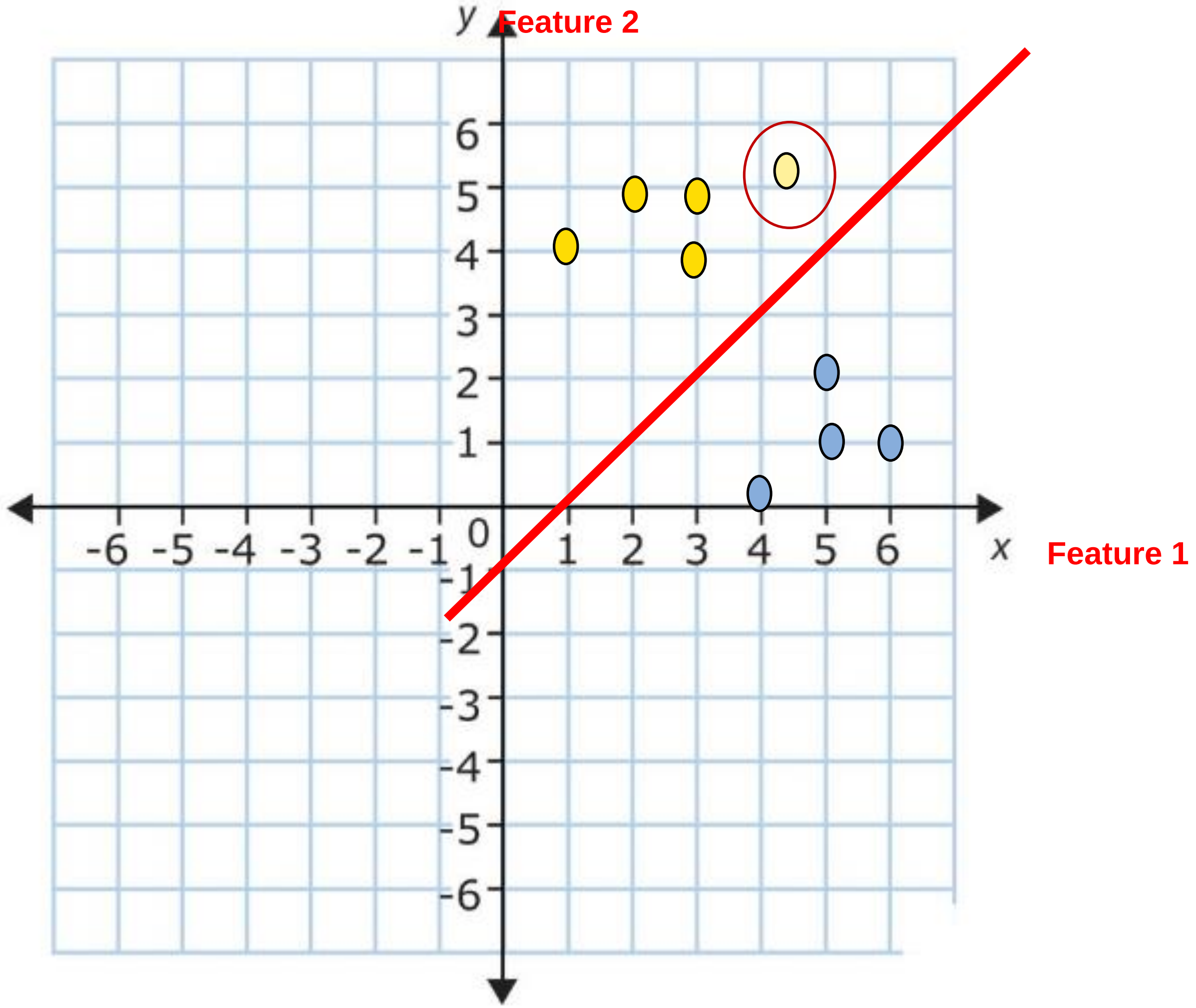
SVM



Feature x	Feature y	Label
1	4	Cancer
2	5	Cancer
3	5	Cancer
3	4	Cancer
6	1	Normal
4	0	Normal
5	2	Normal
5	1	Normal

By using the line that represents the trained model, a new data point (gray point) could be classified
If the data point below the line, it will be classified as Normal

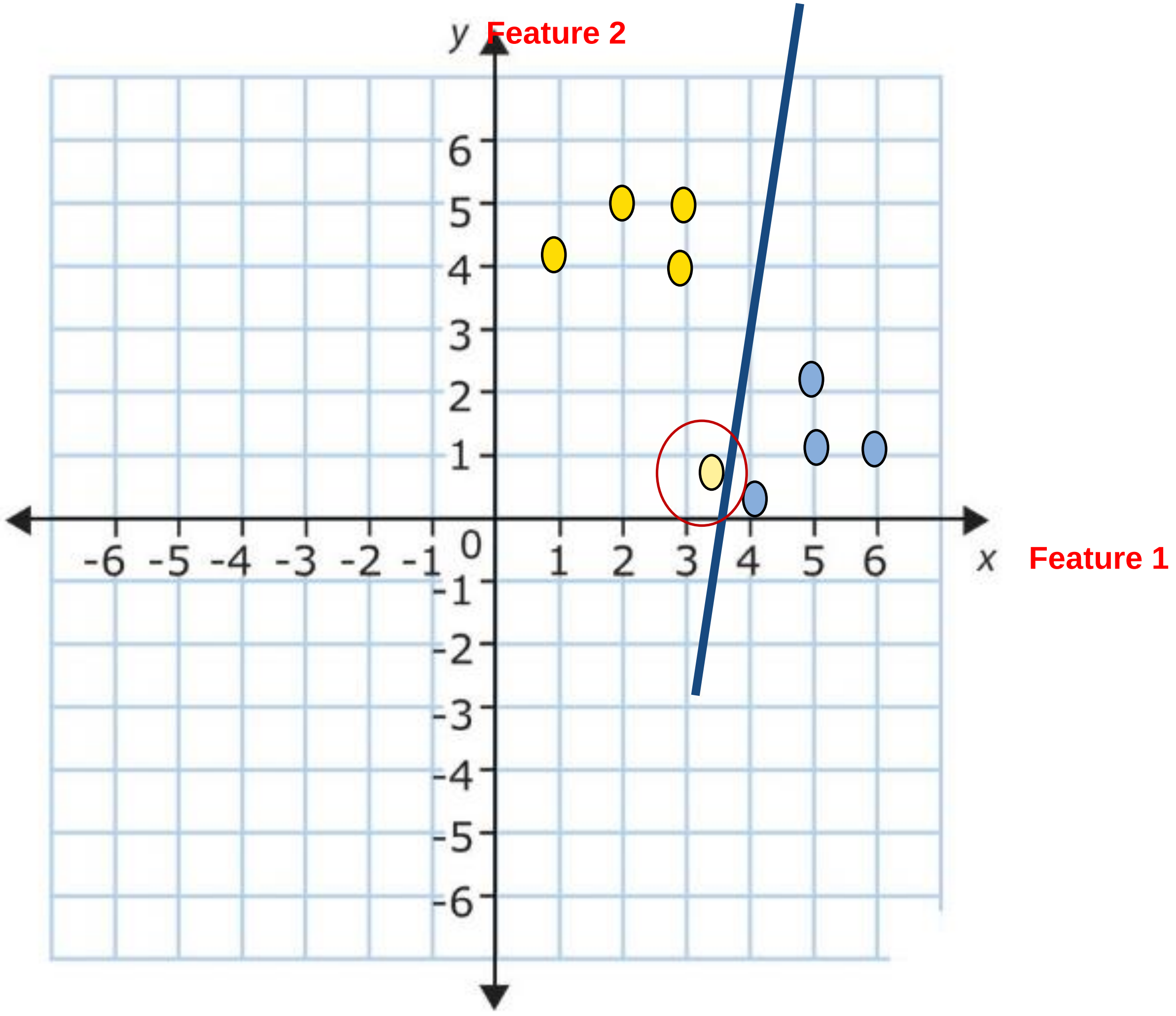
SVM



Feature x	Feature y	Label
1	4	Cancer
2	5	Cancer
3	5	Cancer
3	4	Cancer
6	1	Normal
4	0	Normal
5	2	Normal
5	1	Normal

By using the line that represents the trained model, a new data point (gray point) could be classified
If the data point above the line, it will be classified as Cancer

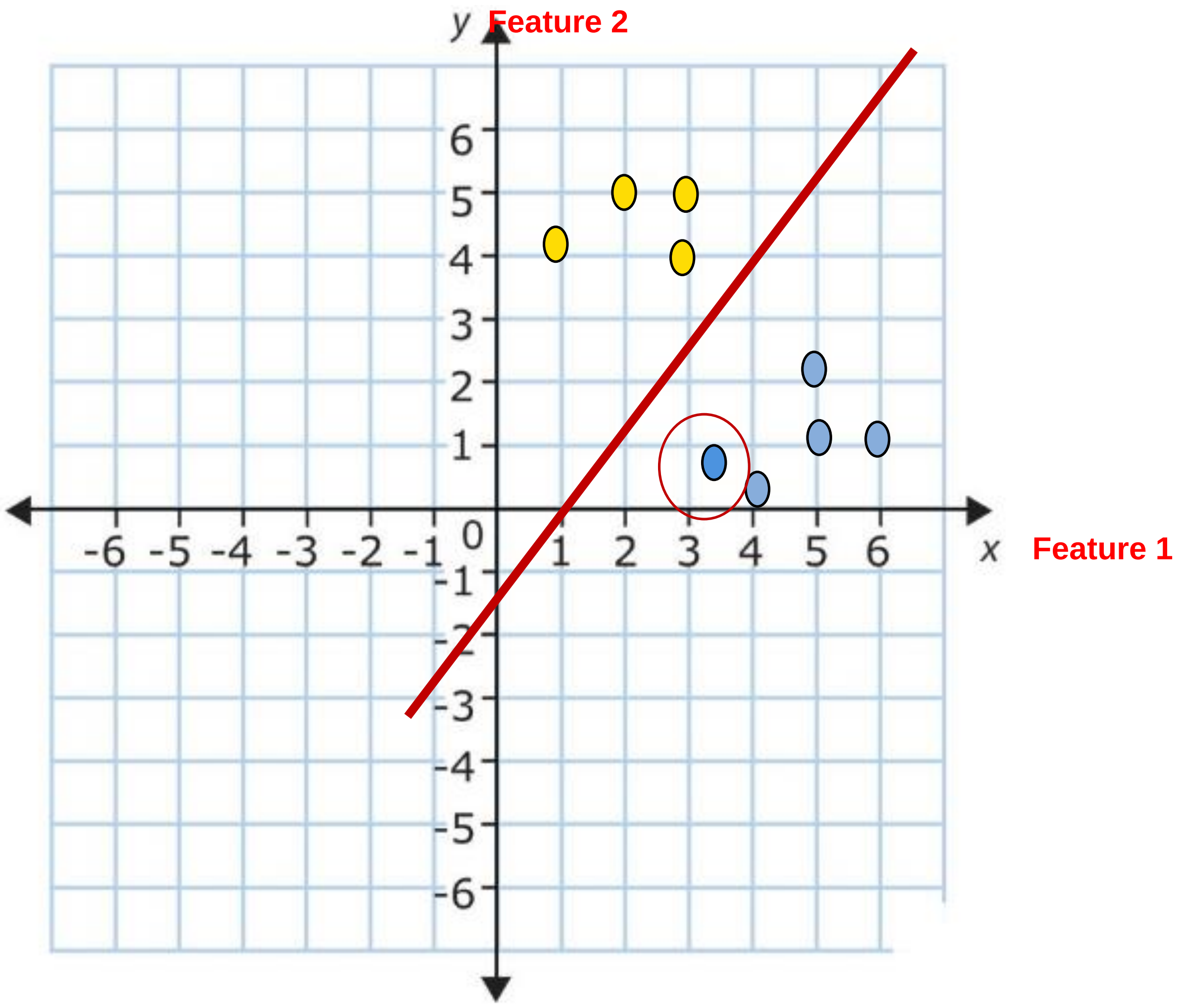
SVM



Feature x	Feature y	Label
1	4	Cancer
2	5	Cancer
3	5	Cancer
3	4	Cancer
6	1	Normal
4	0	Normal
5	2	Normal
5	1	Normal

By using a blue line, the data point inside the circle is classified as a Cancer but it is closer to healthy individuals

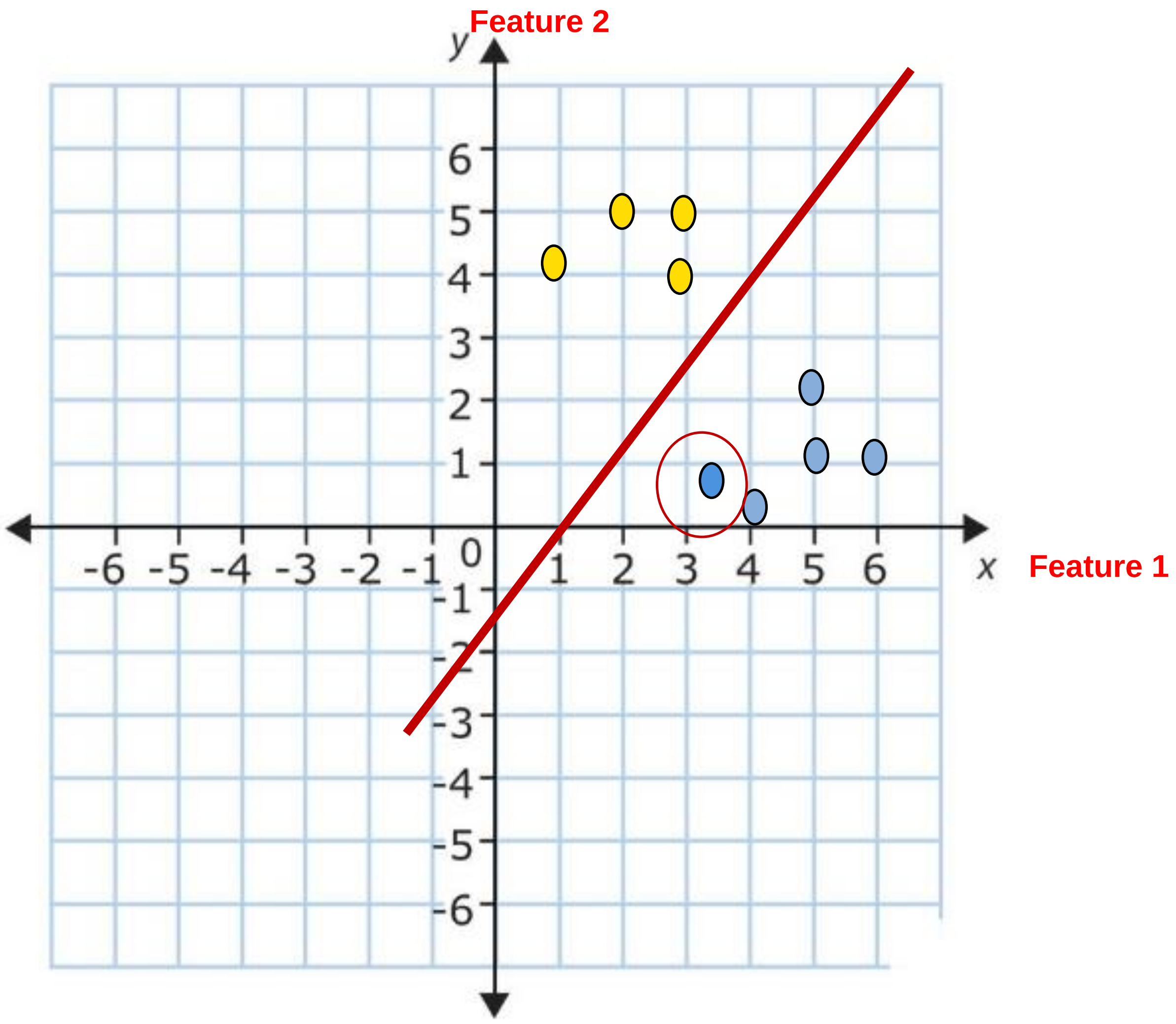
SVM



Feature x	Feature y	Label
1	4	Cancer
2	5	Cancer
3	5	Cancer
3	4	Cancer
6	1	Normal
4	0	Normal
5	2	Normal
5	1	Normal

By using a red line, the data point inside the circle is classified as a normal

SVM

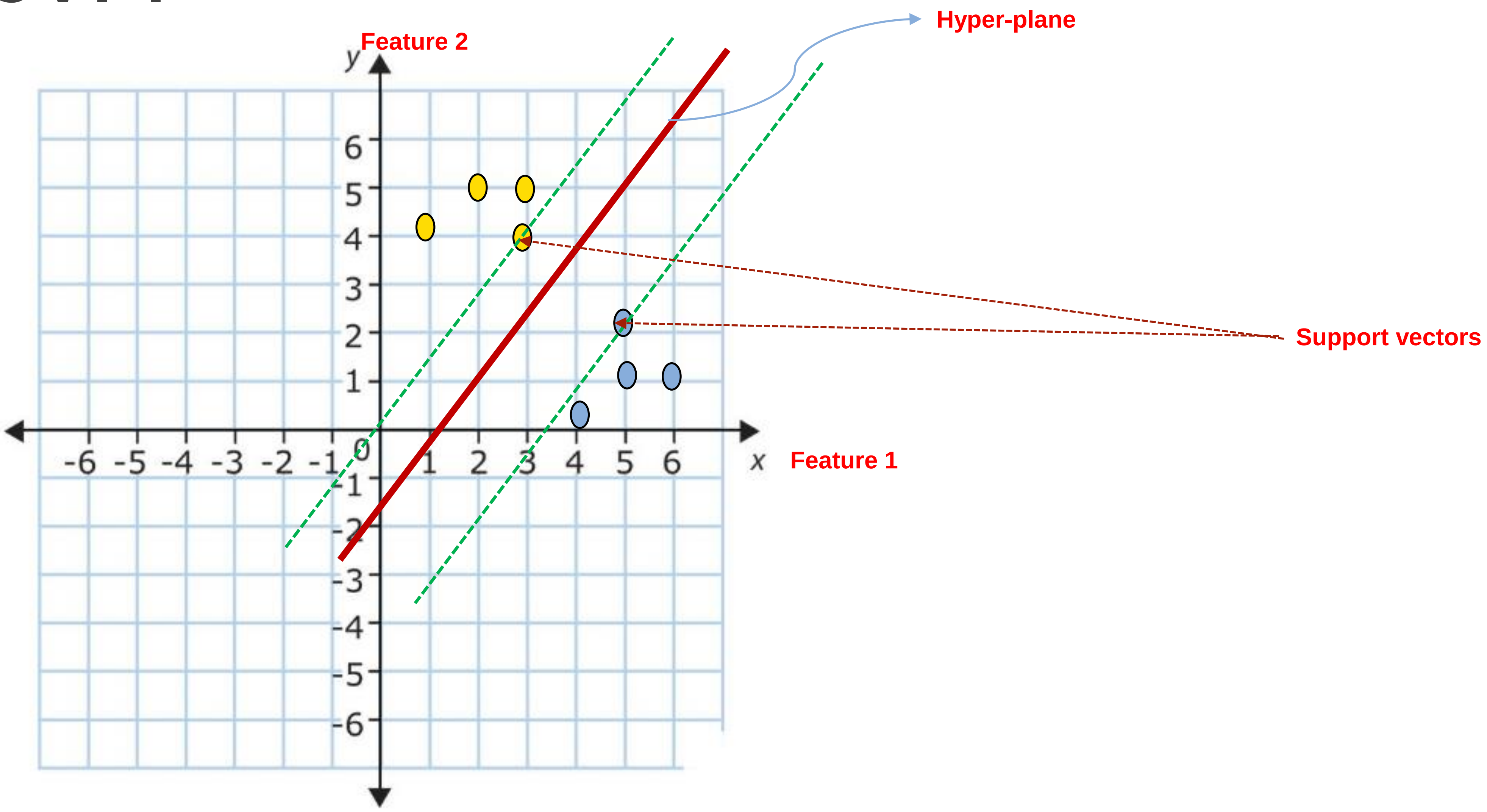


$$y = x - 1$$

$$-x + y + 1 = 0$$

$$-4x + 4y + 4 = 0$$

SVM



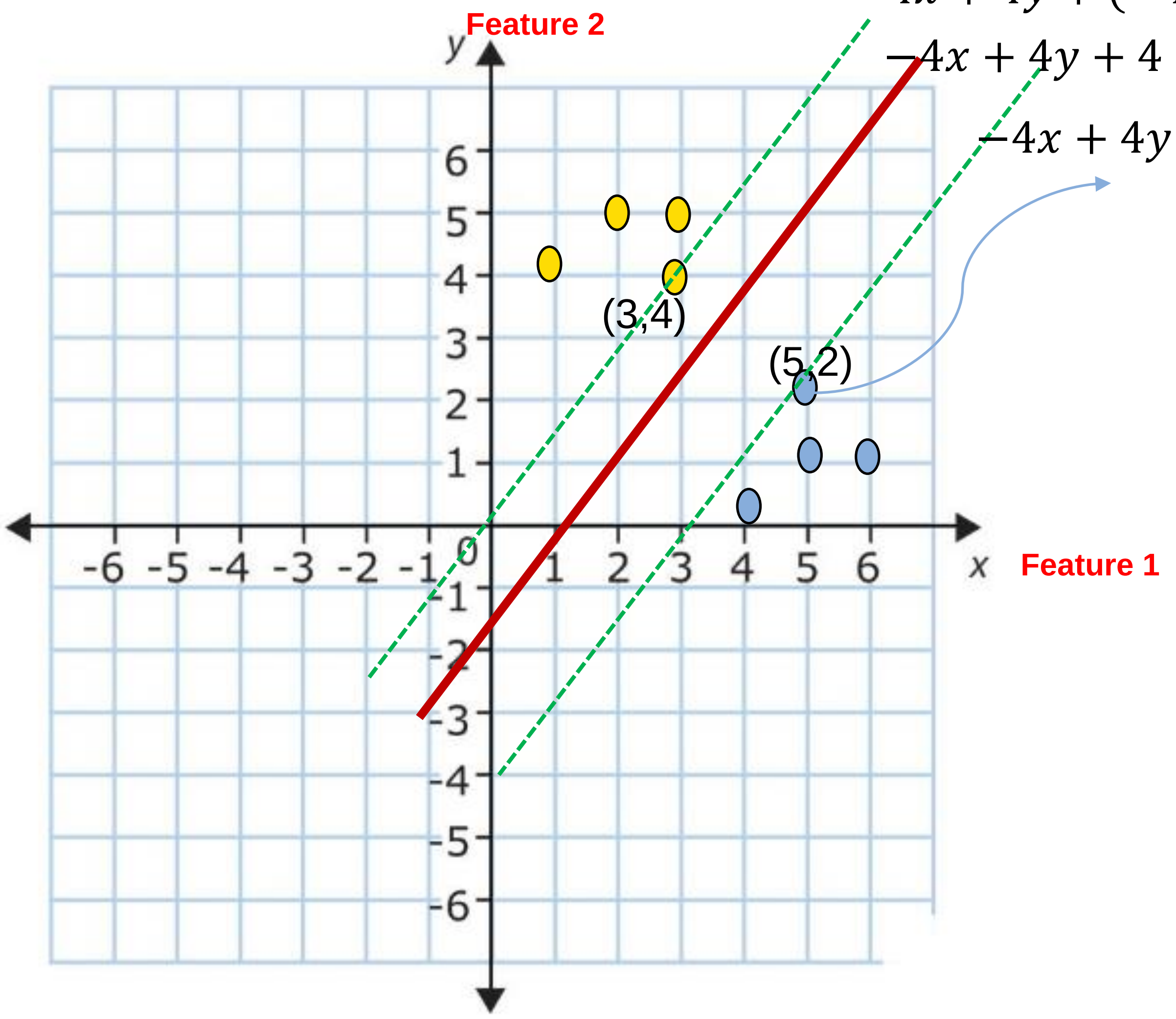
Draw a parallel two lines that intercept the two closest data points to the Hyper-plane

SVM

$$-4x + 4y + (-4) = 0$$

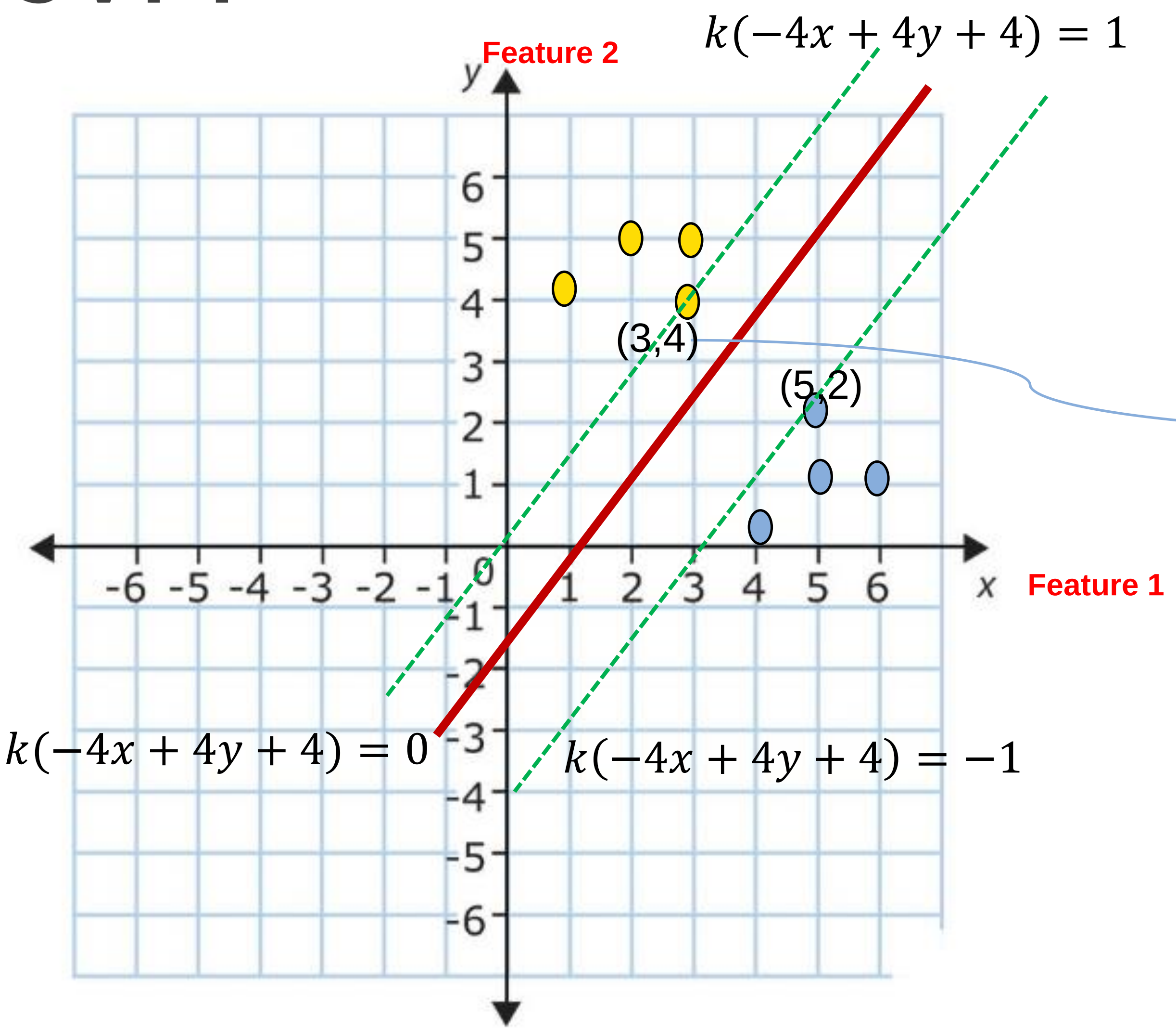
$$-4x + 4y + 4 = 0$$

$$-4x + 4y + 12 = 0$$



Draw a parallel two lines that intercept the two closest data points to the Hyper-plane

SVM

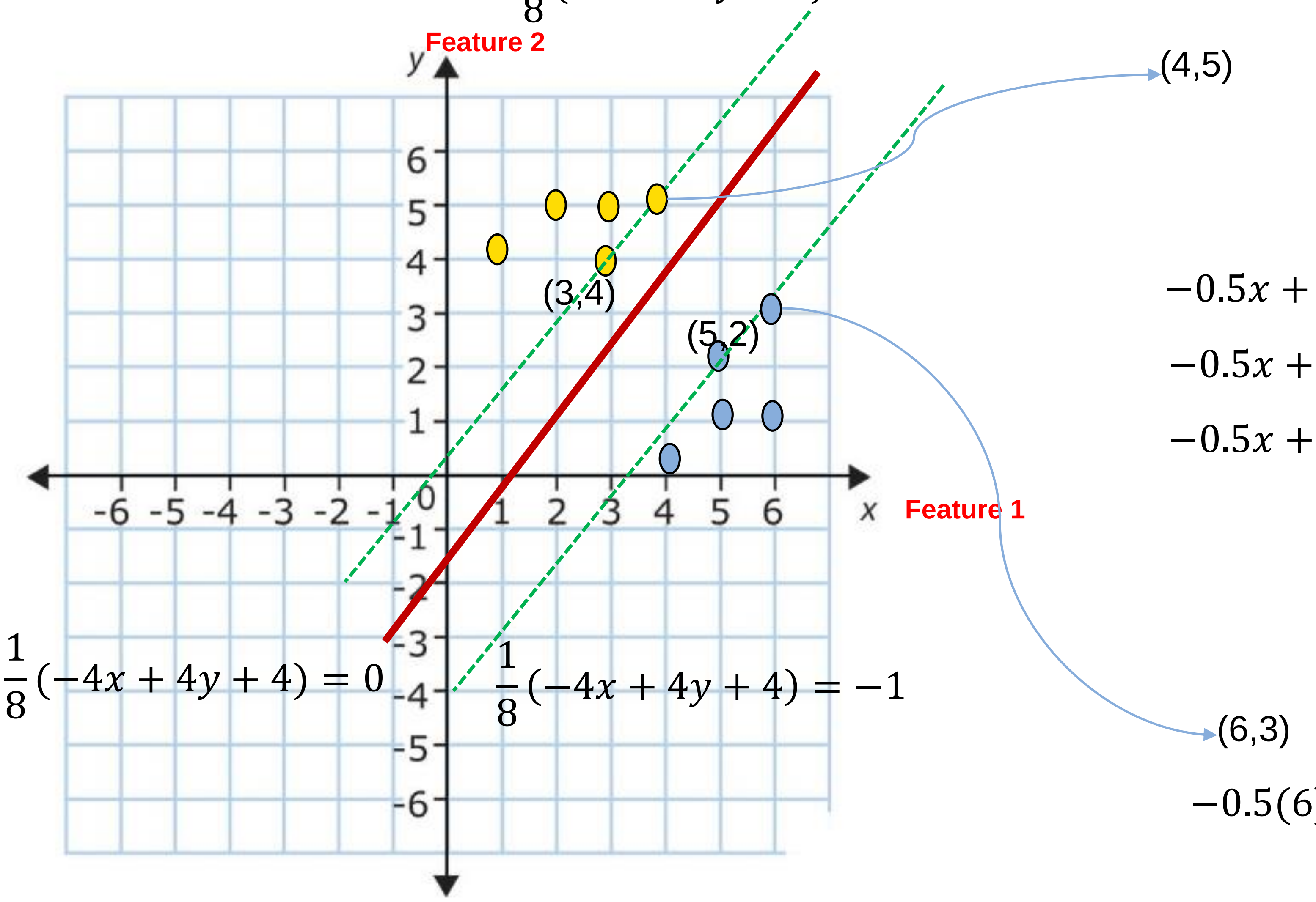


$k(-4(3) + 4(4) + 4) = 1 \quad k = \frac{1}{8}$

Can we normalize the equations

SVM

$$\frac{1}{8}(-4x + 4y + 4) = 1$$



$$\frac{1}{8}(-4x + 4y + 4) = 0$$

$$\frac{1}{8}(-4x + 4y + 4) = -1$$

$$-0.5(4) + 0.5(5) + 0.5 = +1$$

$$-0.5x + 0.5y + 0.5 = 1$$

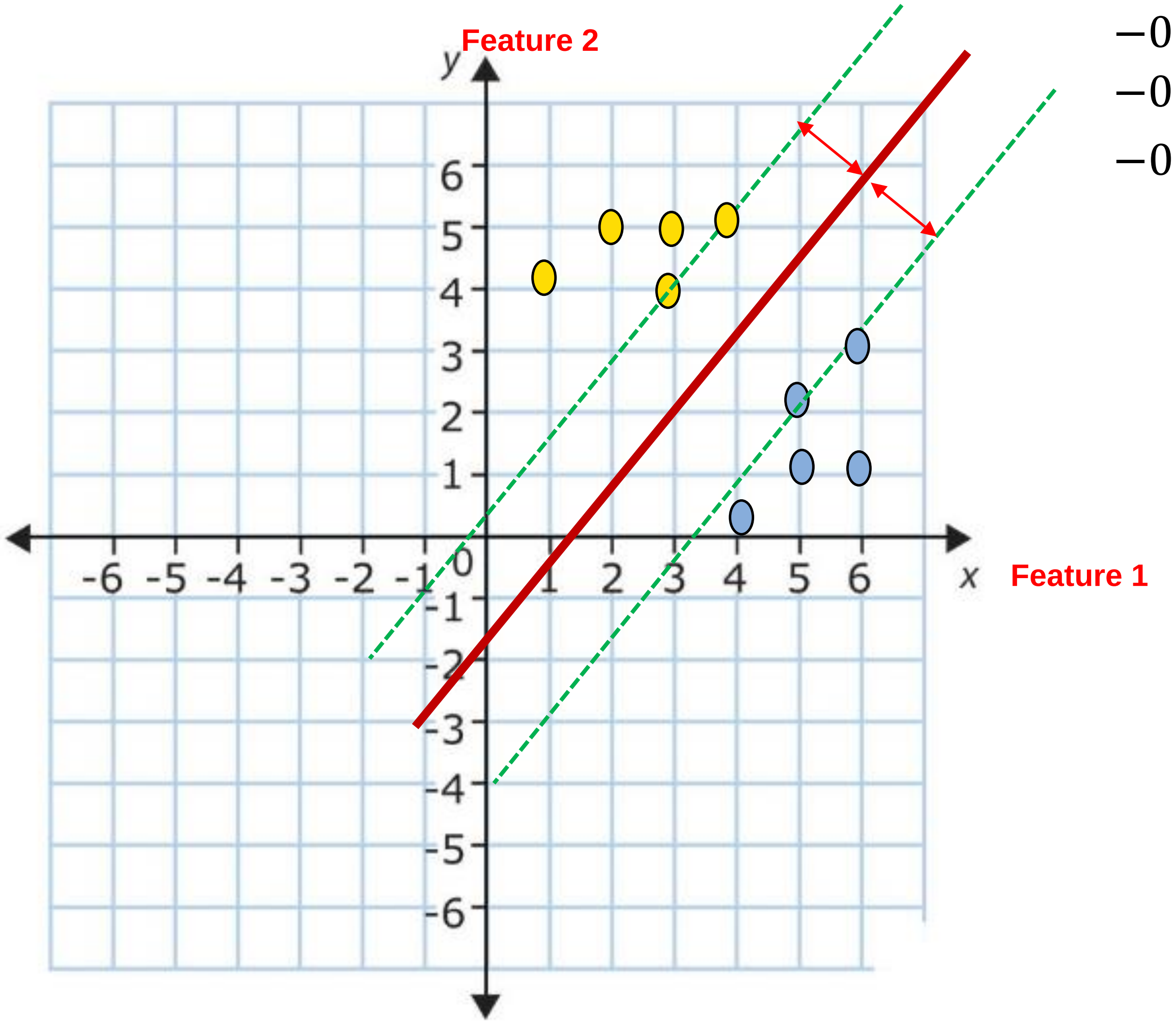
$$-0.5x + 0.5y + 0.5 = 0$$

$$-0.5x + 0.5y + 0.5 = -1$$

$$-0.5(6) + 0.5(3) + 0.5 = -1$$

SVM Standard form of Equations

SVM



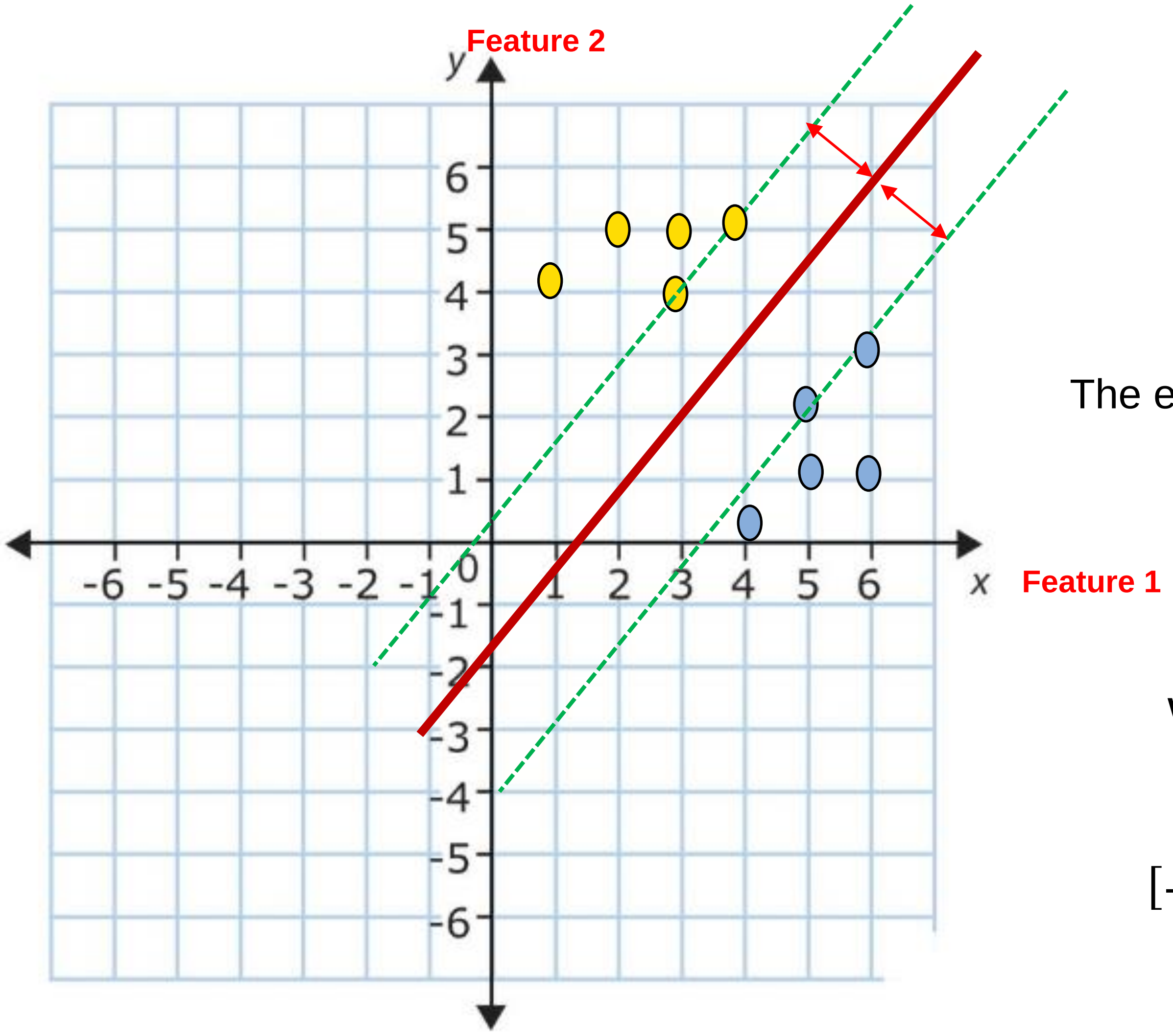
$$-0.5x + 0.5y + 0.5 = 1$$
$$-0.5x + 0.5y + 0.5 = 0$$
$$-0.5x + 0.5y + 0.5 = -1$$

$$b = 0.5$$

$$-0.5x + 0.5y + b = 1$$
$$-0.5x + 0.5y + b = 0$$
$$-0.5x + 0.5y + b = -1$$

The red line in the center between two lines, this is the reason to use 1 and -1 in the above equations.

SVM



$$b = 0.5$$

$$-0.5x + 0.5y + b = 1$$

$$-0.5x + 0.5y + b = 0$$

$$-0.5x + 0.5y + b = -1$$

The equation of hyper-plane in SVM usually expressed like this:

$$W^T x + b = 0$$

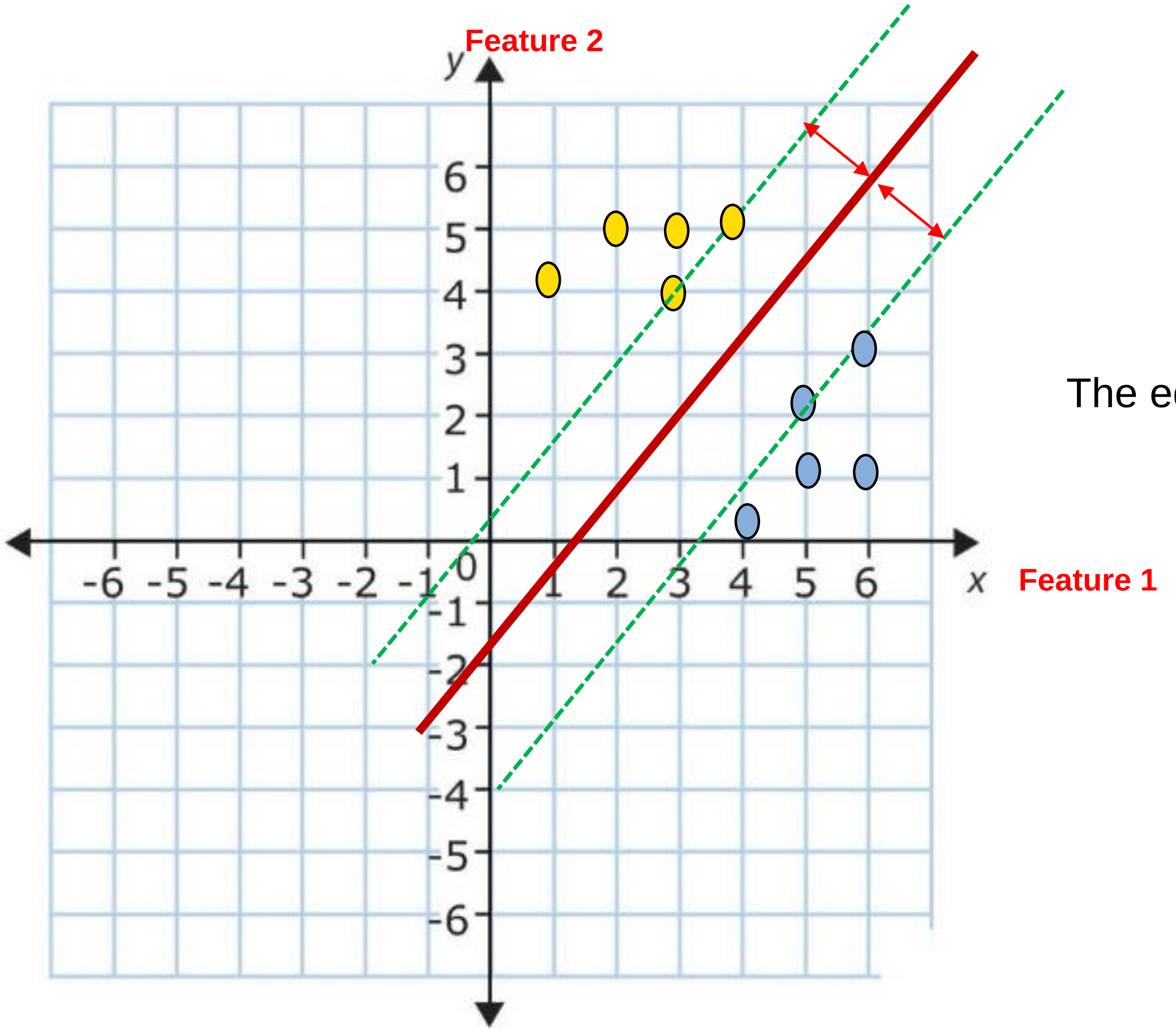
Weights

Features

$$\begin{bmatrix} -0.5 & 0.5 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} + b = 0$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

SVM



$$b = 0.5$$

$$-0.5x + 0.5y + b = 1$$

$$-0.5x + 0.5y + b = 0$$

$$-0.5x + 0.5y + b = -1$$

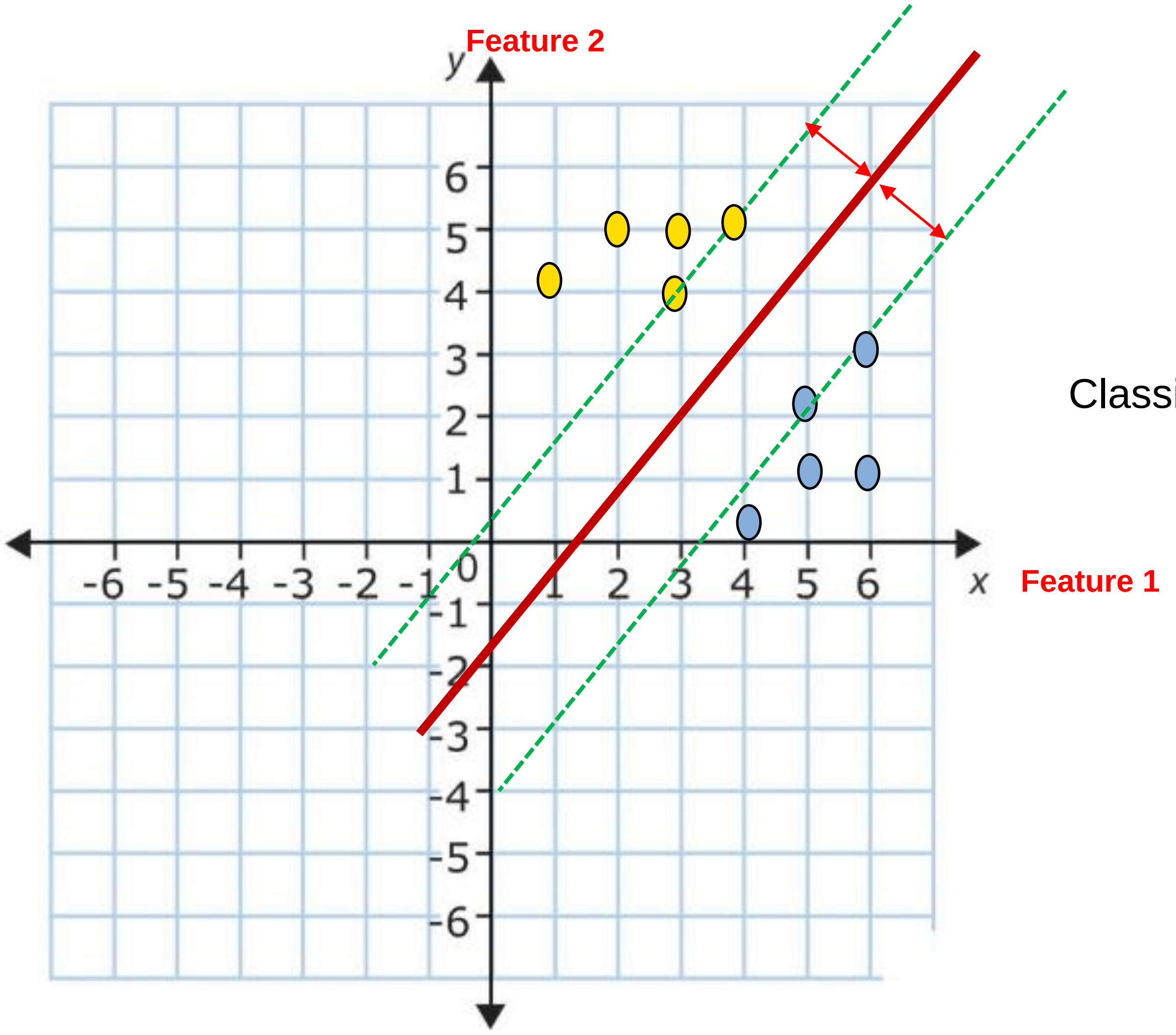
The equation of hyper-plane in SVM usually expressed like this:

$$W^T x + b = 0$$

$$W^T x + b = +1$$

$$W^T x + b = -1$$

SVM



$$b = 0.5$$

$$-0.5x + 0.5y + b = 1$$

$$-0.5x + 0.5y + b = 0$$

$$-0.5x + 0.5y + b = -1$$

Classify a data-point (6,3)

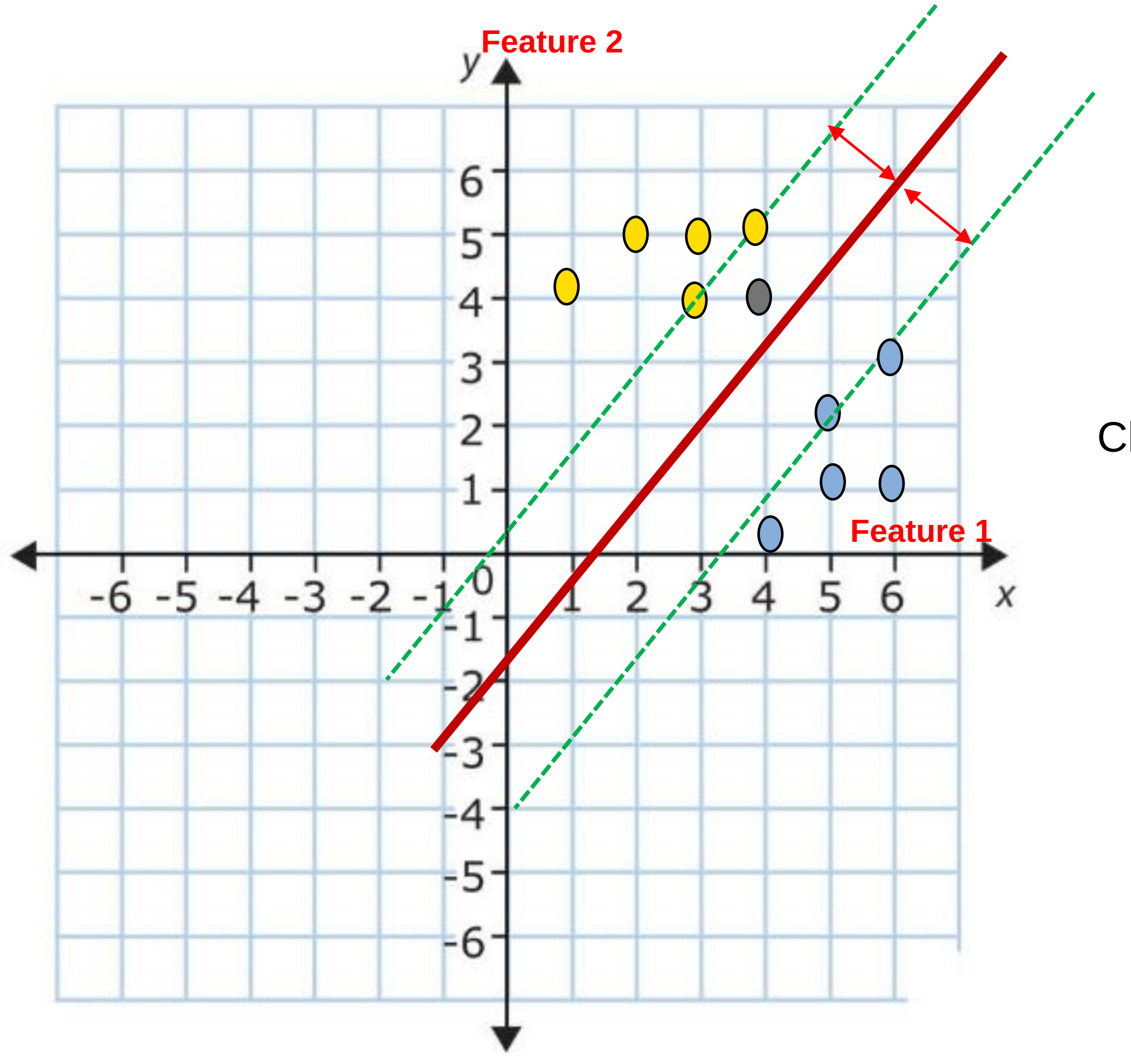
By using a hyper-plane equation:

$$-0.5(6) + 0.5(3) + 0.5 = 0$$

If it results in -1, then belongs to normal class

If it results in +1, then belongs to cancer class

SVM



$$b = 0.5$$

$$-0.5x + 0.5y + b = 1$$

$$-0.5x + 0.5y + b = 0$$

$$-0.5x + 0.5y + b = -1$$

Classify a data-point (4,4)

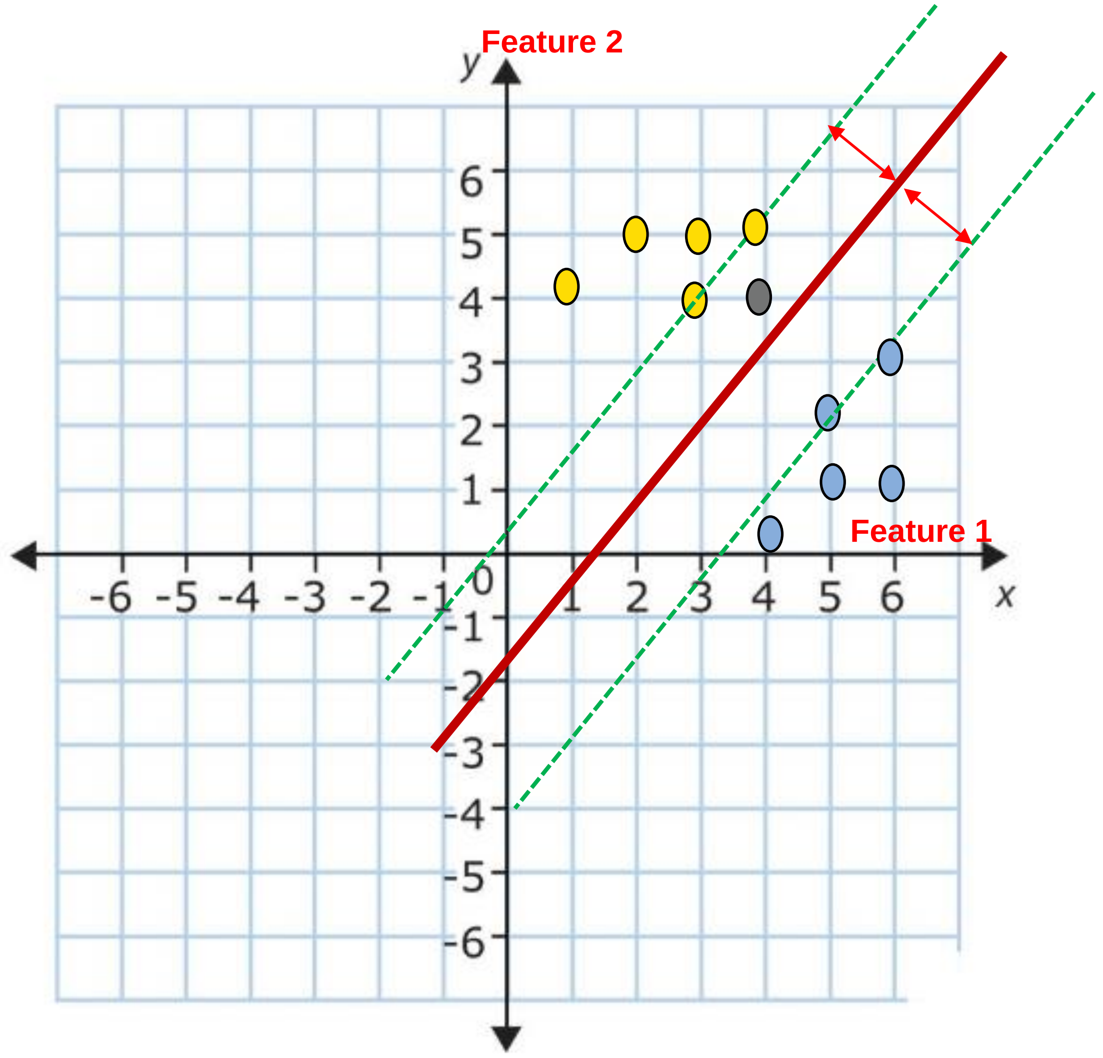
By using a hyper-plane equation:

$$-0.5(4) + 0.5(4) + 0.5 = 0.5$$

If it results in a negative value, then belongs to normal class

If it results in a positive value, then belongs to cancer class

SVM



$$b = 0.5$$

$$-0.5x + 0.5y + b = 1$$

$$-0.5x + 0.5y + b = 0$$

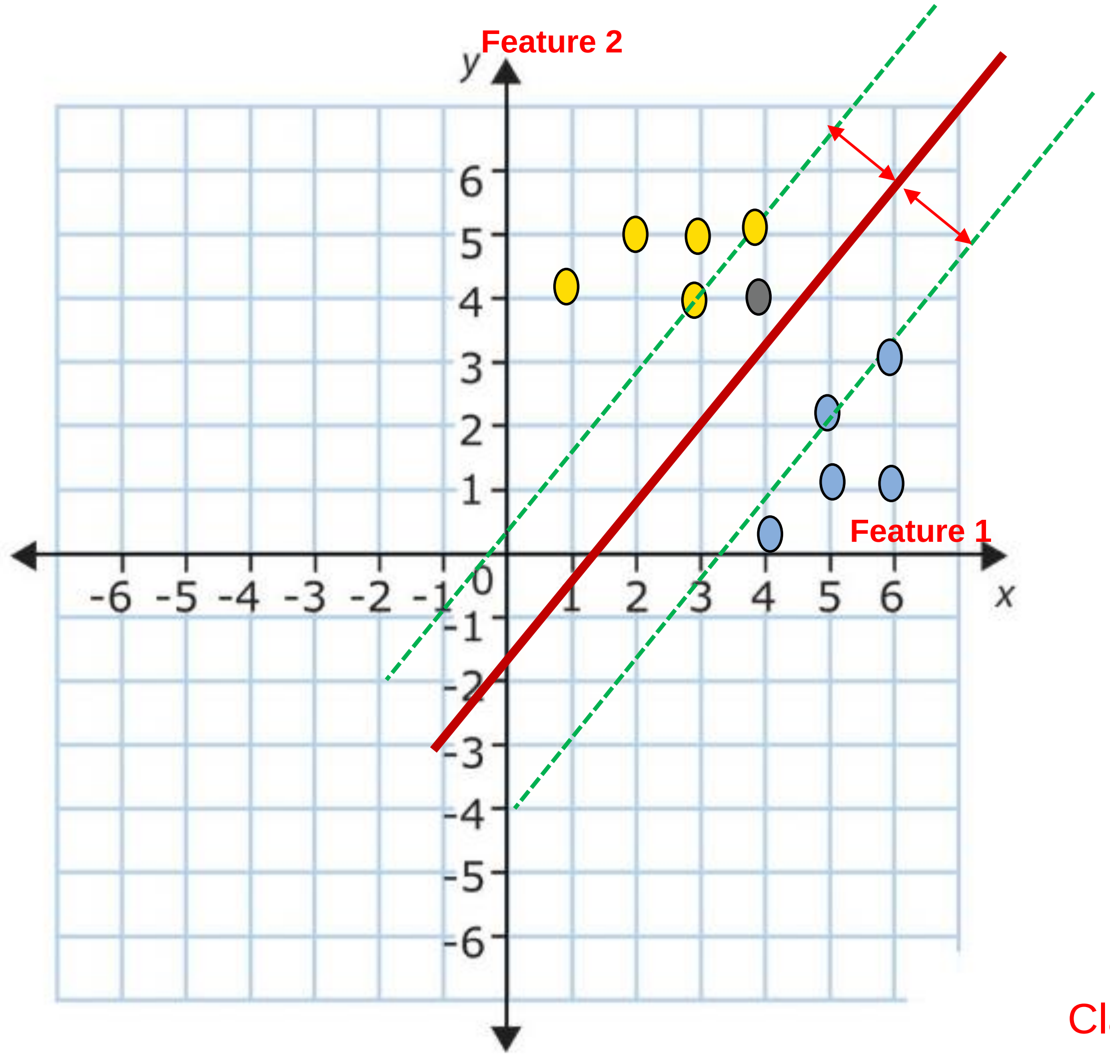
$$-0.5x + 0.5y + b = -1$$

If it results in a negative value, then belongs to normal class
If it results in a positive value, then belongs to cancer class

Feature x	Feature y	Label
1	4	Cancer (+1)
2	5	Cancer (+1)
3	5	Cancer (+1)
3	4	Cancer (+1)
6	1	Normal (-1)
4	0	Normal (-1)
5	2	Normal (-1)
5	1	Normal (-1)

$$Y_i = [1 \quad 1 \quad 1 \quad 1 \quad -1 \quad -1 \quad -1 \quad -1]$$

SVM



$$b = 0.5$$

$$-0.5x + 0.5y + b = 1$$

$$-0.5x + 0.5y + b = 0$$

$$-0.5x + 0.5y + b = -1$$

$$Y_i = [1 \quad 1 \quad 1 \quad 1 \quad -1 \quad -1 \quad -1 \quad -1]$$

Feature x	Feature y	Label
1	4	Cancer (+1)
2	5	Cancer (+1)
3	5	Cancer (+1)
3	4	Cancer (+1)
6	1	Normal (-1)
4	0	Normal (-1)
5	2	Normal (-1)
5	1	Normal (-1)

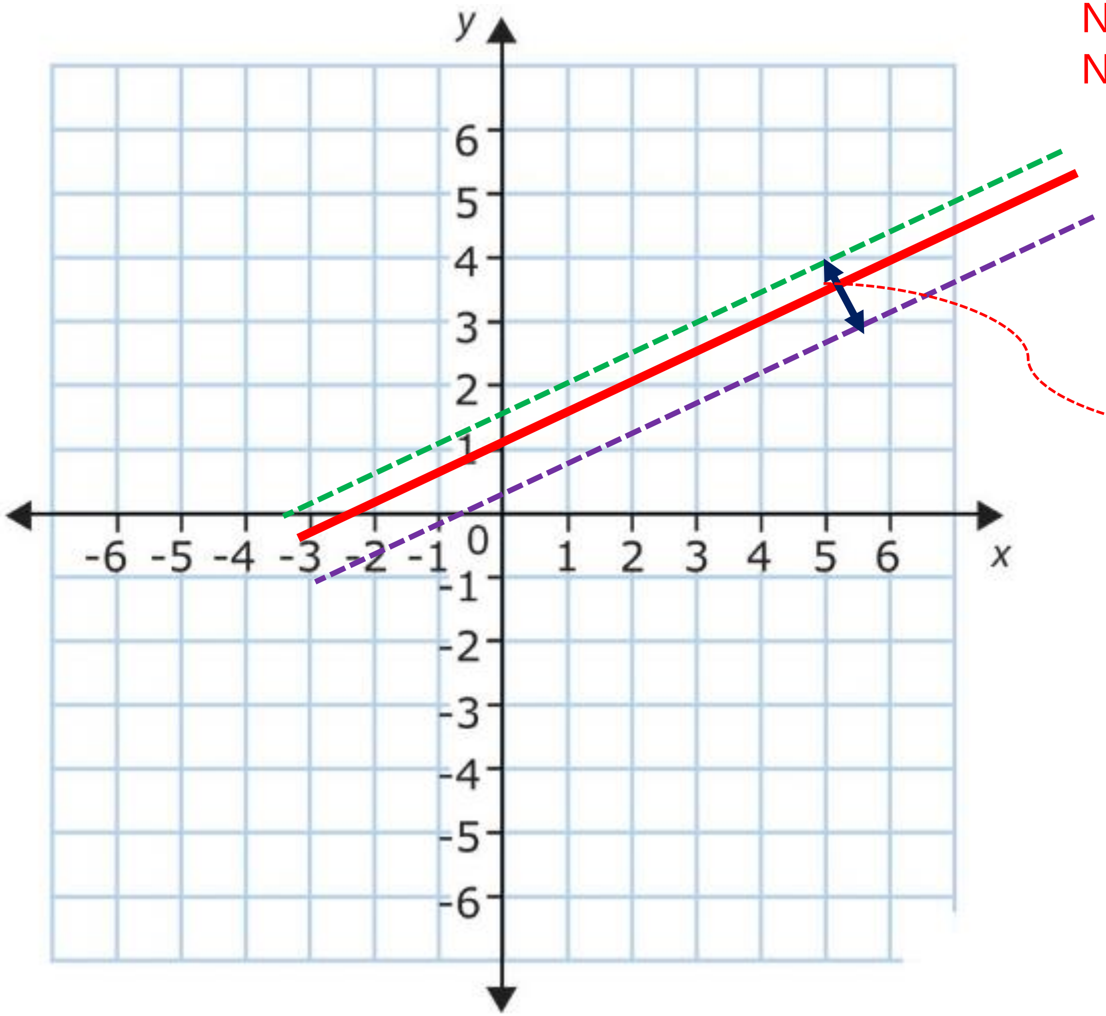
Classifications by SVM is usually defined by the following equation:

$$Y \begin{Bmatrix} +1 & if\ W^T x + b \geq 0 \\ -1 & if\ W^T x + b\ < 0 \end{Bmatrix}$$

SVM

$$Ax + By + C = 0 \qquad W^T x + b = 0$$

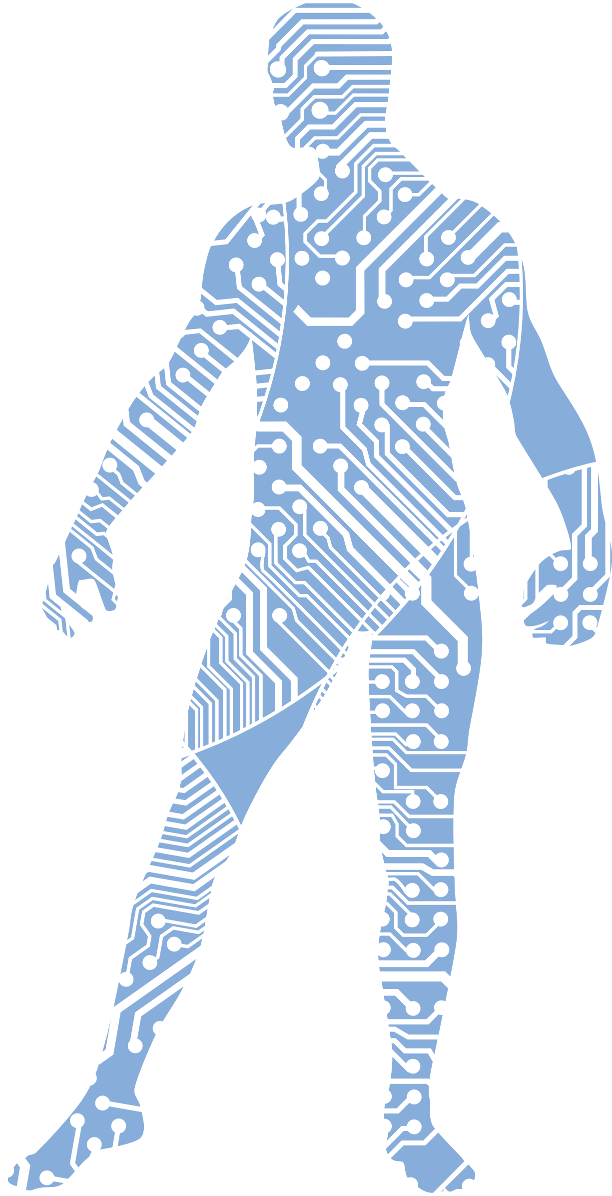
Notes: A and B are weights
Notes: x and y are features



$$d = \frac{|C_2 - C_1|}{\sqrt{A^2 + B^2}}$$

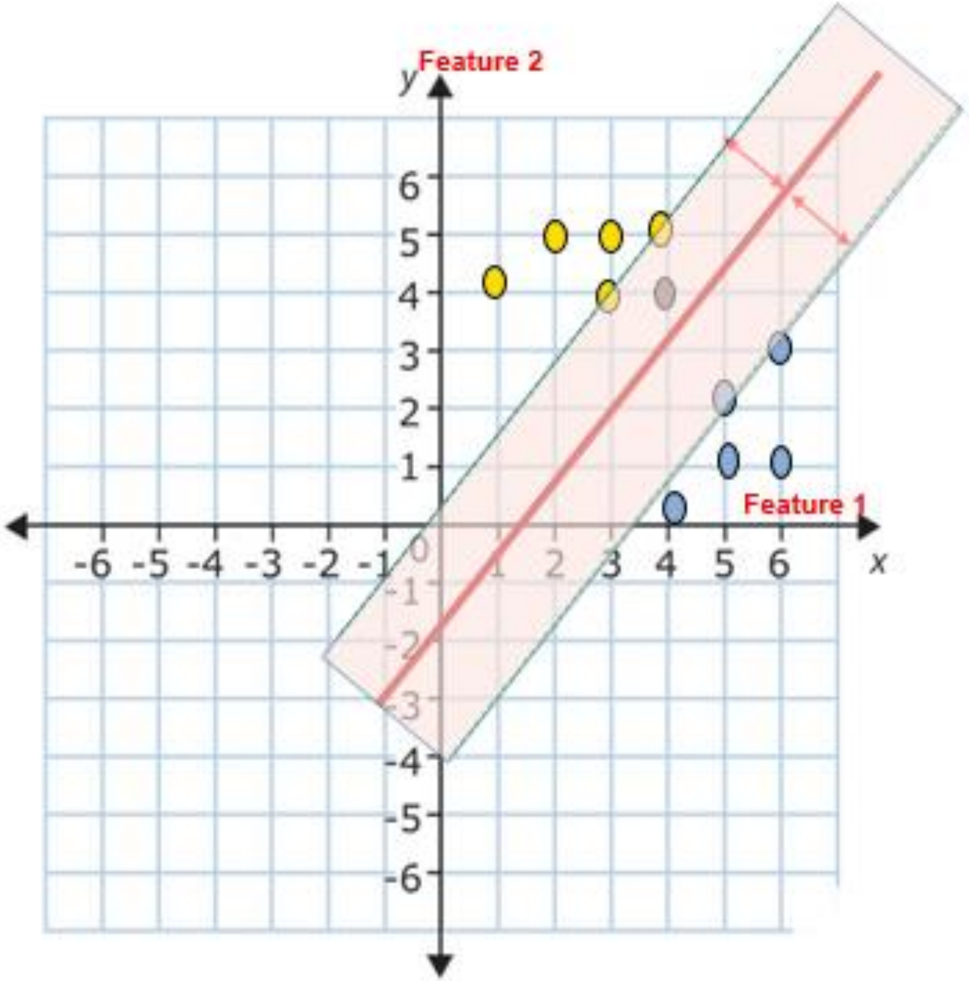
$$d = \frac{2}{\sqrt{A^2 + B^2}}$$

$$d = \frac{2}{\|w\|}$$



SVM tries to find the optimal weights so the distance between two blue lines is maximized
Or minimize the denominator because this will also maximize the distance between the lines.
The area between the two lines is called margin, SVM tries to find a hyper-plane that maximize the width of this margin

SVM



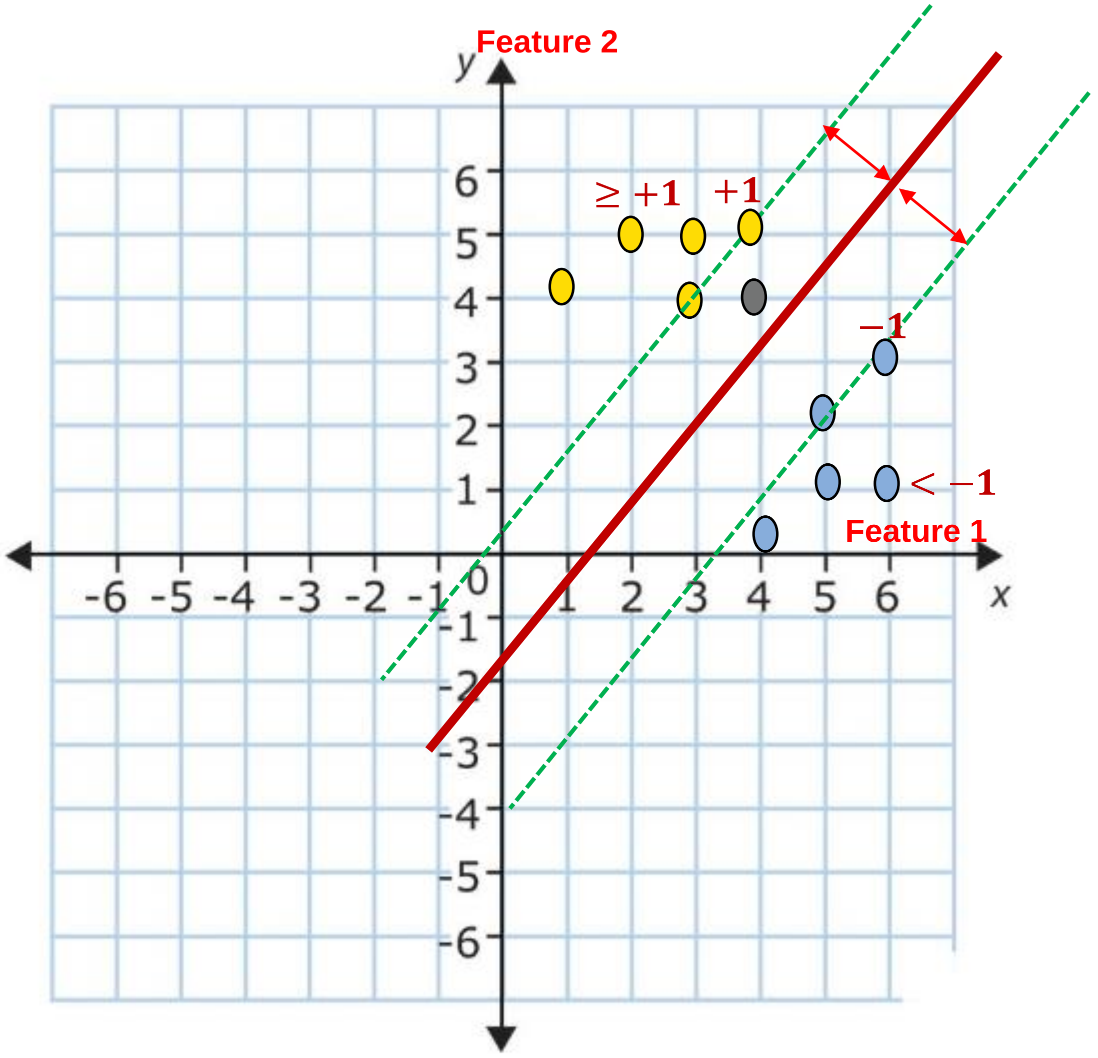
$$d = \frac{2}{\sqrt{A^2 + B^2}}$$

$$d = \frac{2}{\|w\|}$$

$$\max \frac{2}{\|w\|} \; such\ that \;\; W^T x + b \begin{matrix} \geq 1 & if\ Y_i = +1 \\ \leq -1 & if\ Y_i = -1 \end{matrix}$$

The margin should not span beyond the data-points (support vectors) closest to the hyper-plane.
There are two types of Margin: hard and soft.
Soft margin allows for misclassification but with penalties cost.

SVM



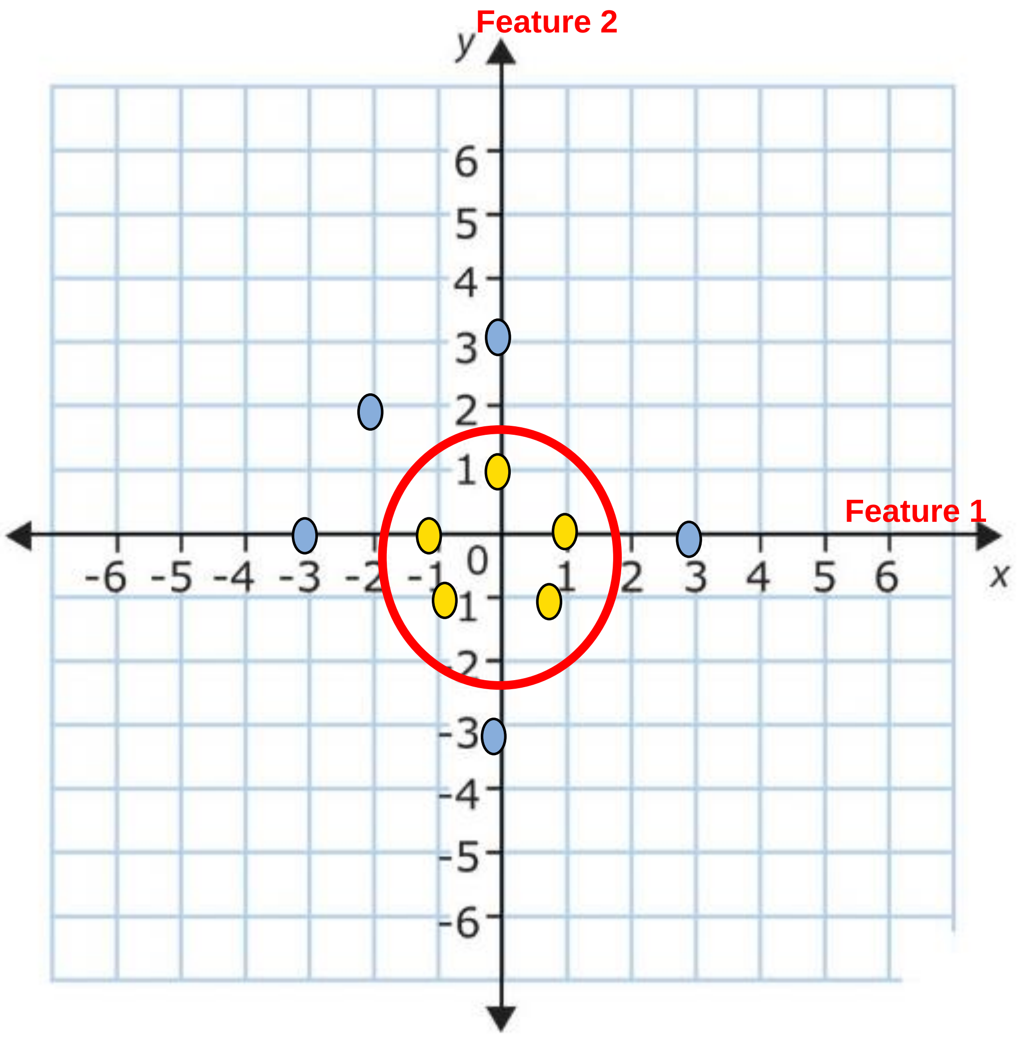
- The points on the dotted lines will result in a classification +1 or -1, according to the red line if it is above or below.
- The points further away from the dotted line result in a classification greater than one or less than one according to the red line if it is above or below

$$\max \frac{2}{\|w\|} \, such\ that \;\; W^T x + b \begin{array}{ll} \geq 1 & if\ Y_i = +1 \\ \leq -1 & if\ Y_i = -1 \end{array}$$

$$\max \frac{2}{\|w\|} \, such\ that \;\; Y_i(W^T x + b) \geq 1$$

$$Y_i = [1 \quad 1 \quad 1 \quad 1 \quad -1 \quad -1 \quad -1 \quad -1]$$

SVM



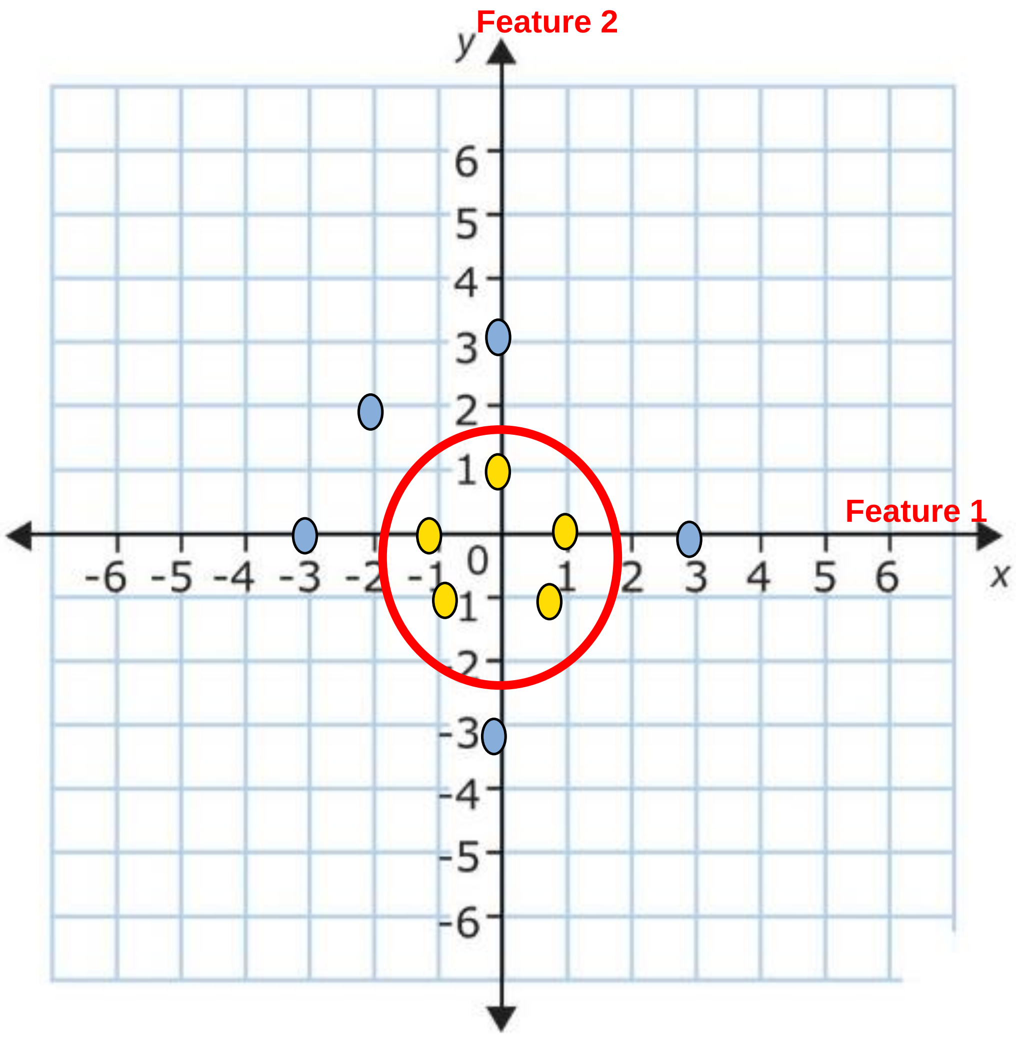
- In some problems or data-points, a straight line can not be used as a classification line (Not a linear function)
- The mathematical function that takes non-linear data and apply some transformation to make it linear is called kernel.
- Kernels are set of functions used to transform data from lower dimension to higher dimension to manipulate data using dot product.

$$\emptyset(x, y) = \left(x^2, \sqrt{2}xy, y^2\right)$$

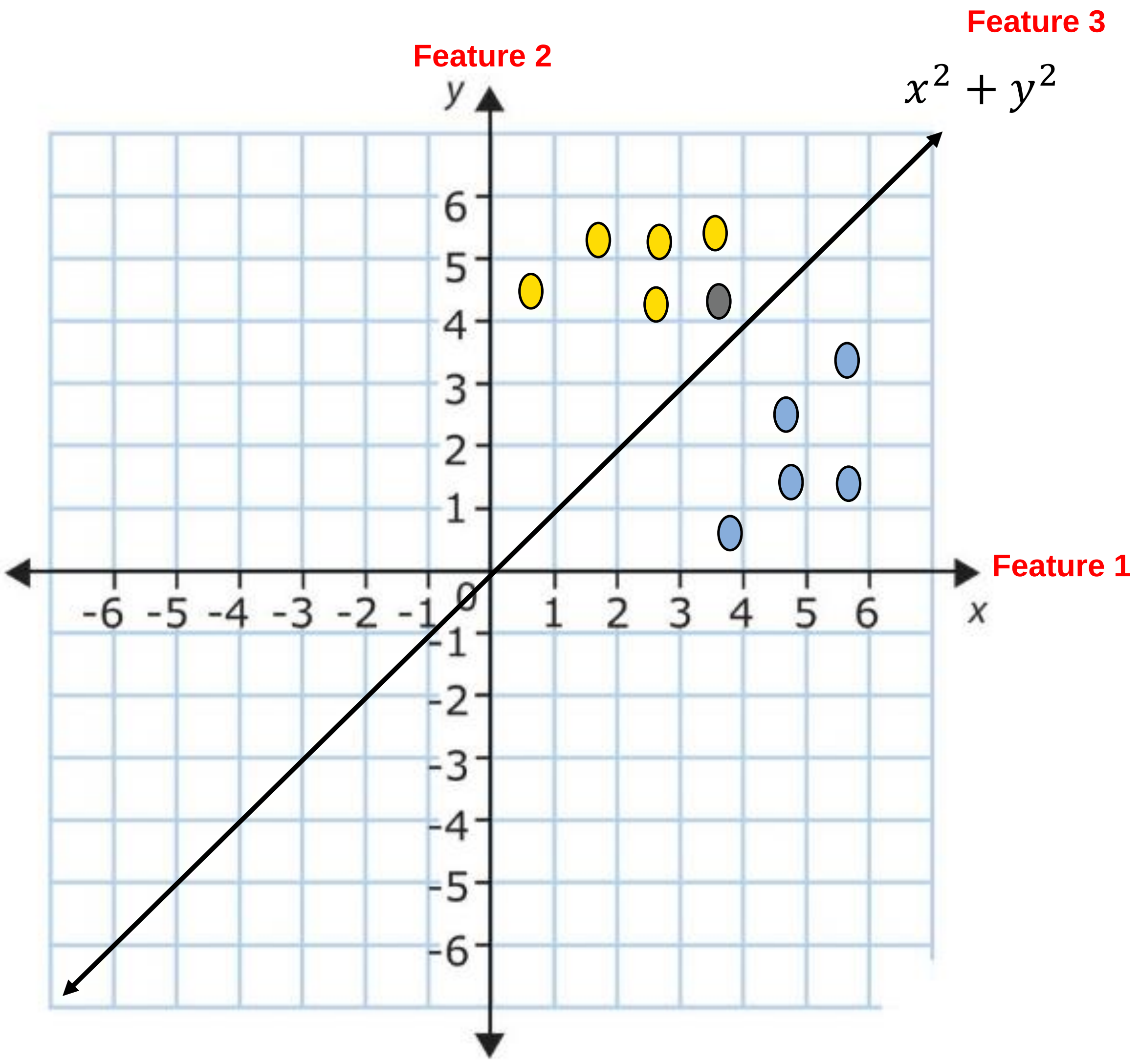
❑ The SVM kernel is a function that takes low-dimensional input space and transforms it into higher-dimensional space, it converts nonseparable problems to separable problems.

❑ For example linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid.

SVM



Low dimensional feature space



High dimensional feature space
Adding new features based on some transformation functions

```python
import numpy as np
import pandas as pd
from sklearn import svm
```

```python
iris = pd.read_csv('iris.csv')

# Splitting the data into features and labels
X = iris.iloc[:, :-1]
y = iris.iloc[:, -1]

# Scaling the features
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
X = scaler.fit_transform(X)
```

```python
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25)
```

```python
clf = svm.SVC(kernel='rbf')
clf.fit(X_train, y_train)
```

```python
# Making predictions on the test set

y_pred = clf.predict(X_test)
```

```python
# Evaluating the classifier accuracy
from sklearn.metrics import accuracy_score

accuracy = accuracy_score(y_test, y_pred)
print('Accuracy:', accuracy)
```

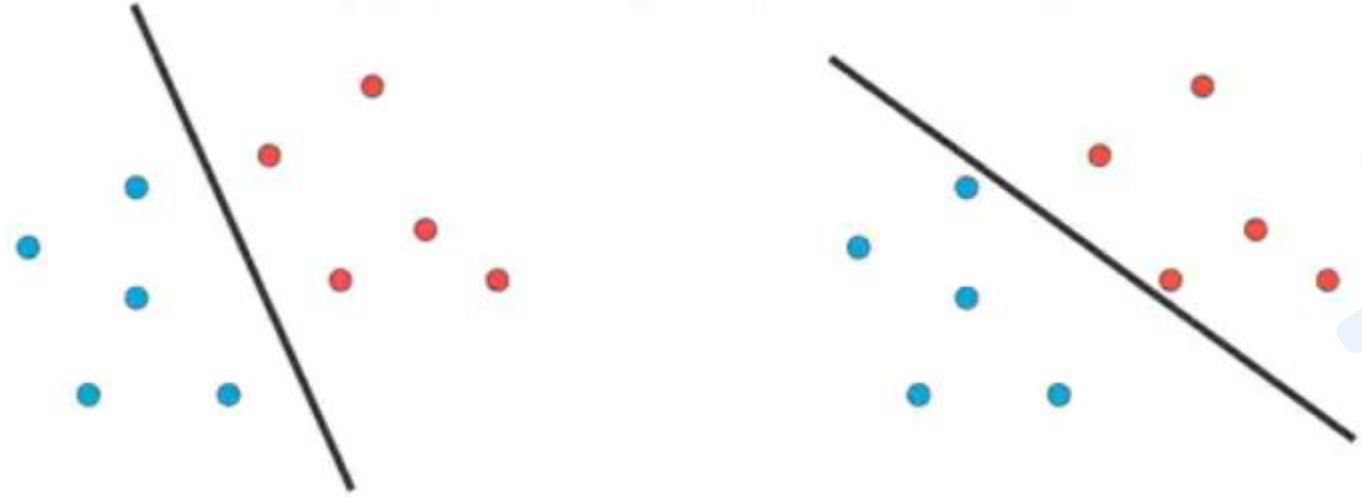
Note! (You should read the following for <u>the exam</u>)

- https://developers.google.com/machine-learning/crash-course/classification/thresholding
- https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall
- https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc
- https://www.youtube.com/watch?v=lBHGPrvXTTM&list=PL4gu8xQu0_5JBO1FKRO5p20wc8DprlOgn

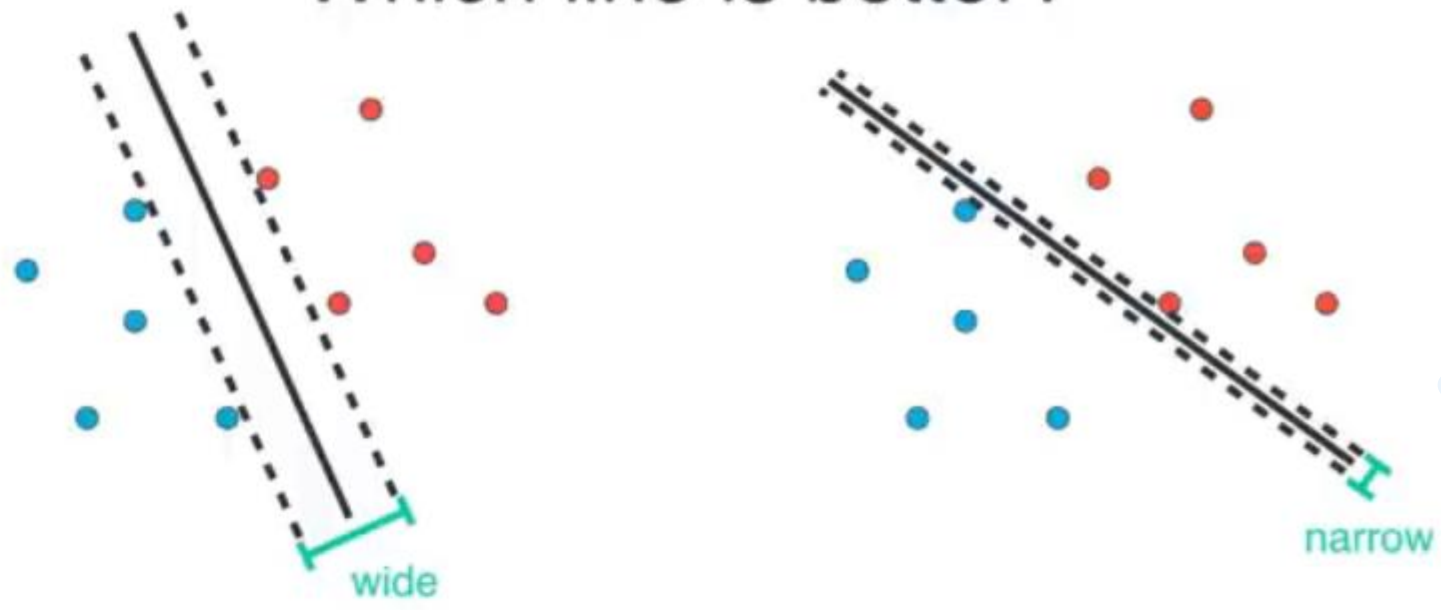
Thank You

Note

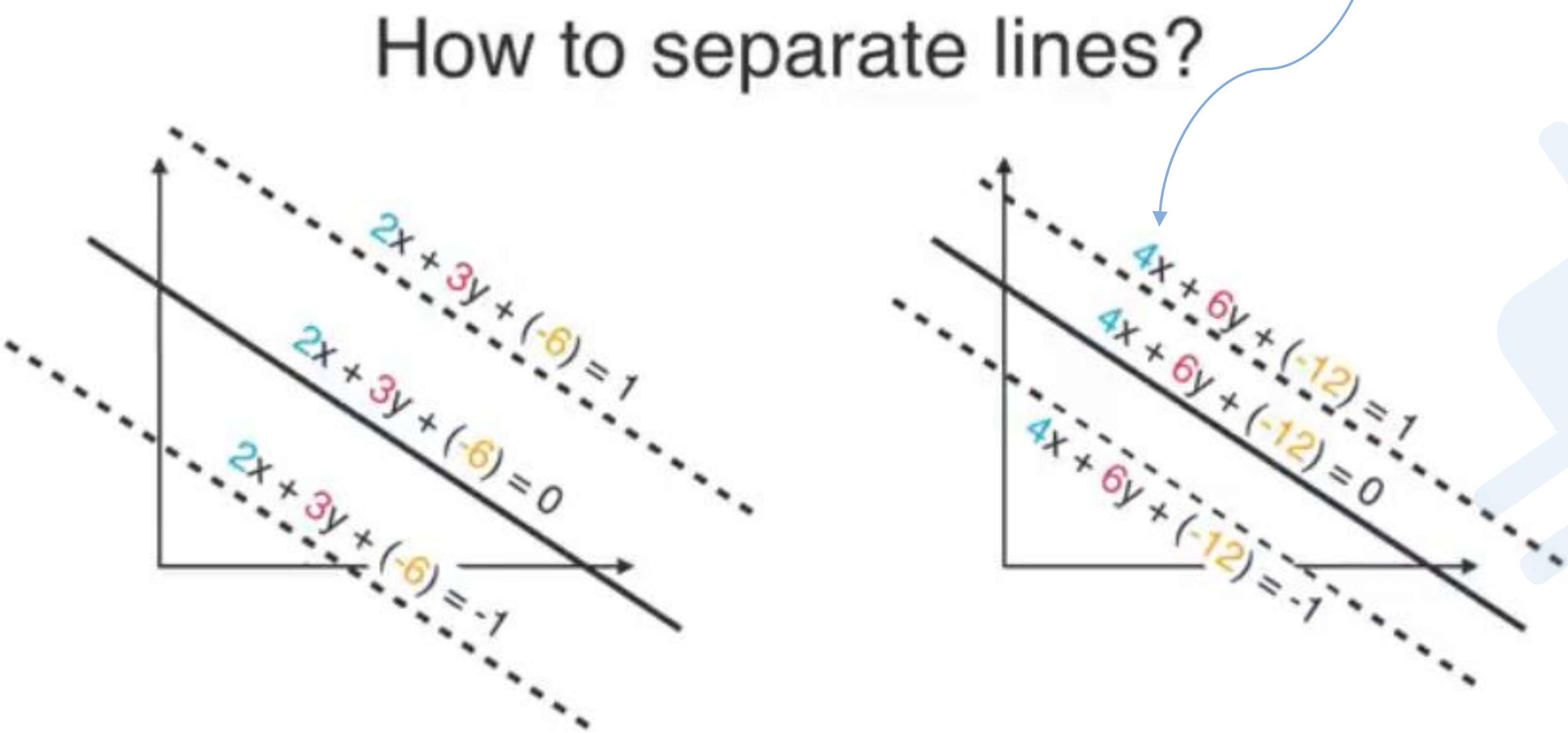
Which line is better?



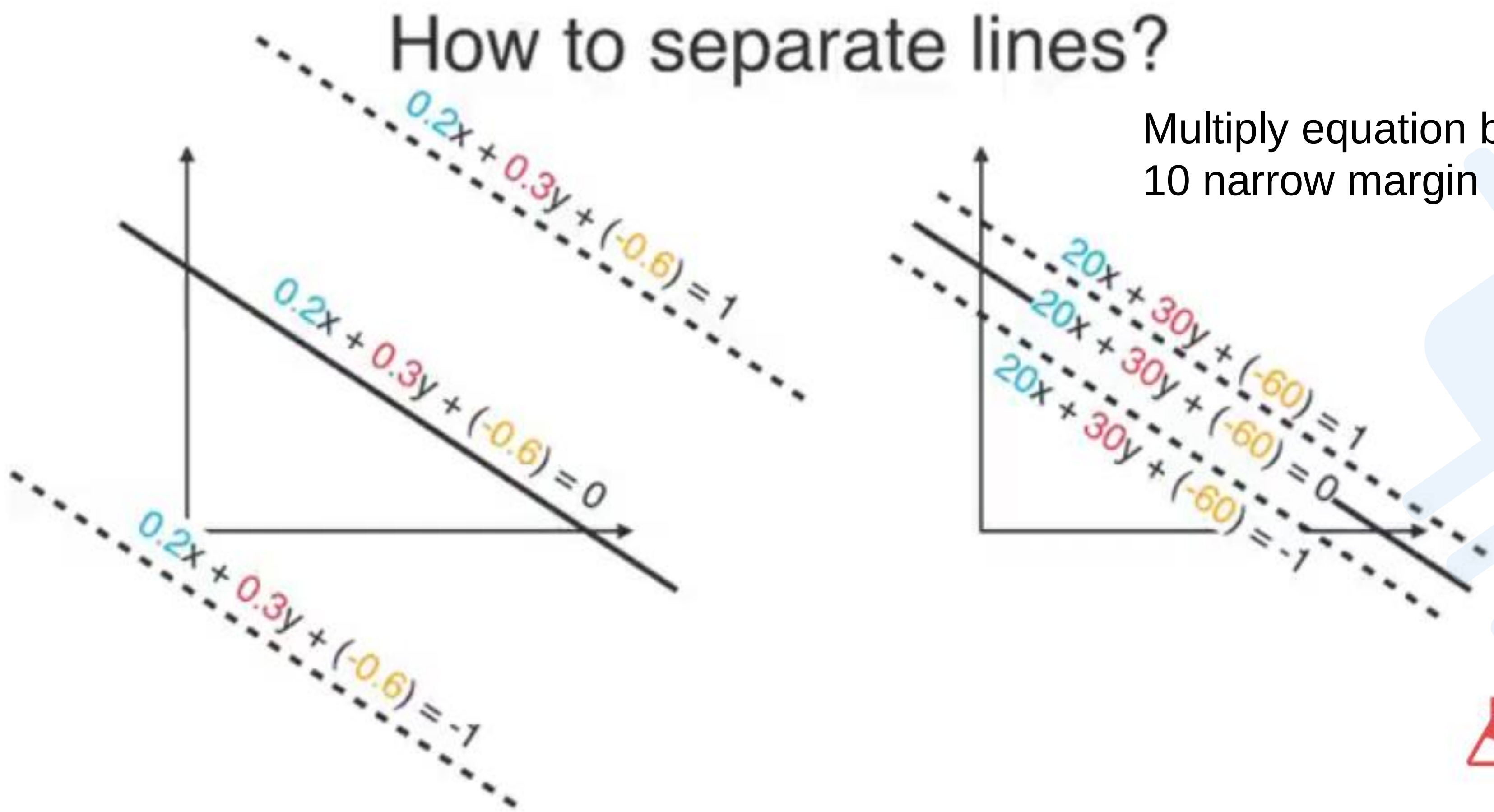
Which line is better?



Multiply equation by a big factor = narrow margin



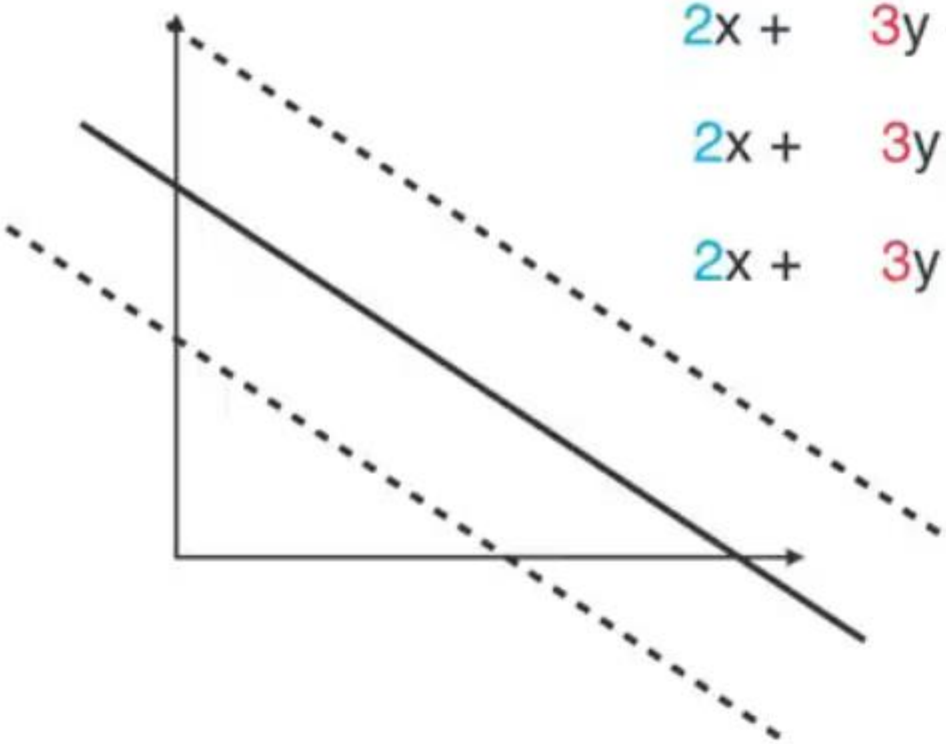
Multiply equation by a small factor (i.e. less than 1) = wider margin



Multiply equation by a big factor = 10 narrow margin

Multiply equation by a small factor (0.1) = wider margin

Expanding rate



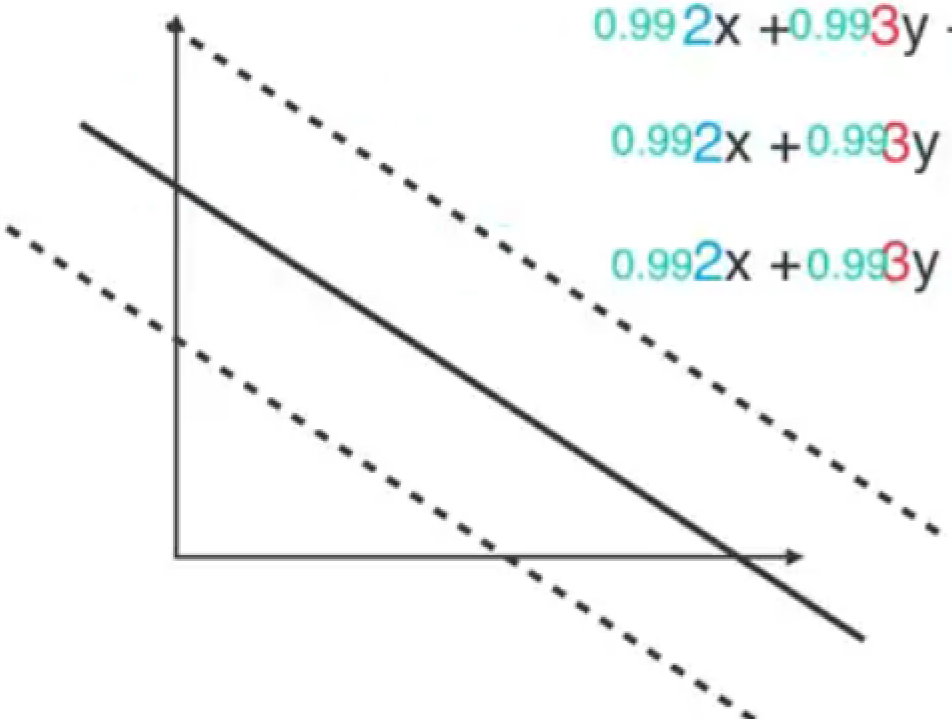
$2x + 3y + (-6) = -1$

$2x + 3y + (-6) = 0$

$2x + 3y + (-6) = 1$

Expanding rate

Expanding rate

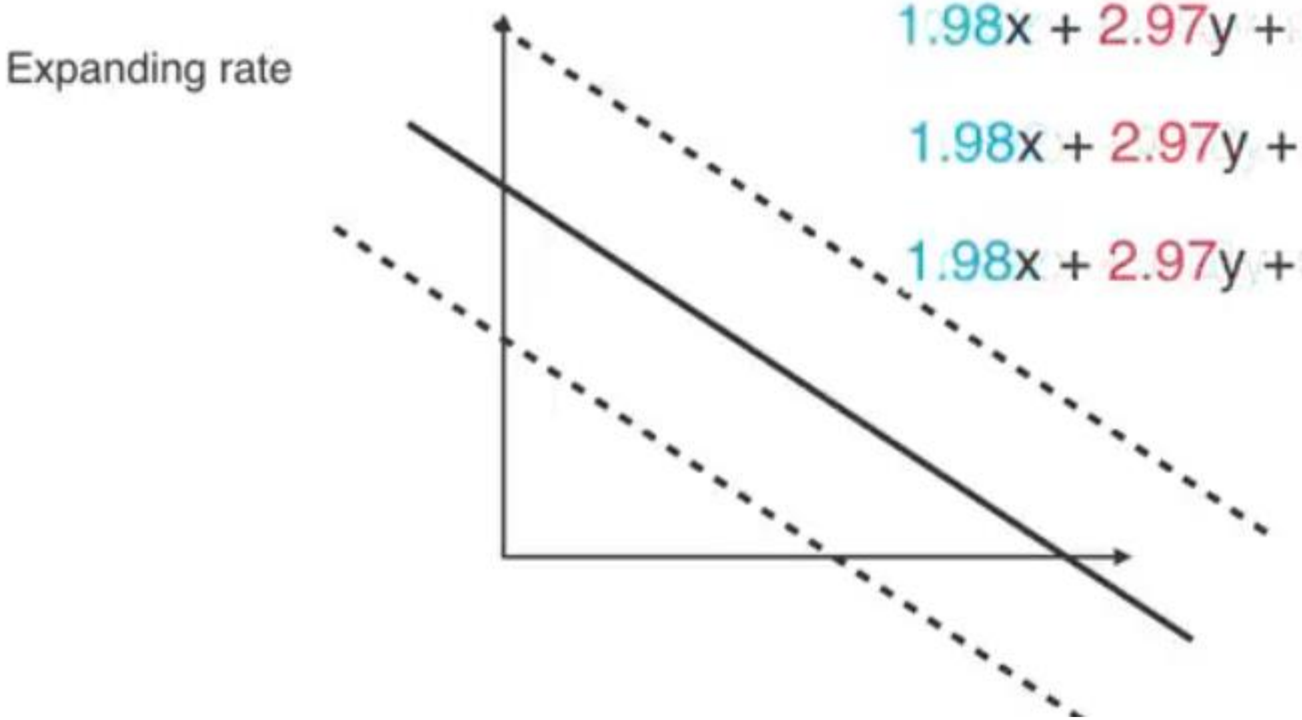


$$0.99\,2x + 0.99\,3y + 0.99(-6) = -1$$

$$0.99\,2x + 0.99\,3y + 0.99(-6) = 0$$

$$0.99\,2x + 0.99\,3y + 0.99(-6) = 1$$

Expanding rate

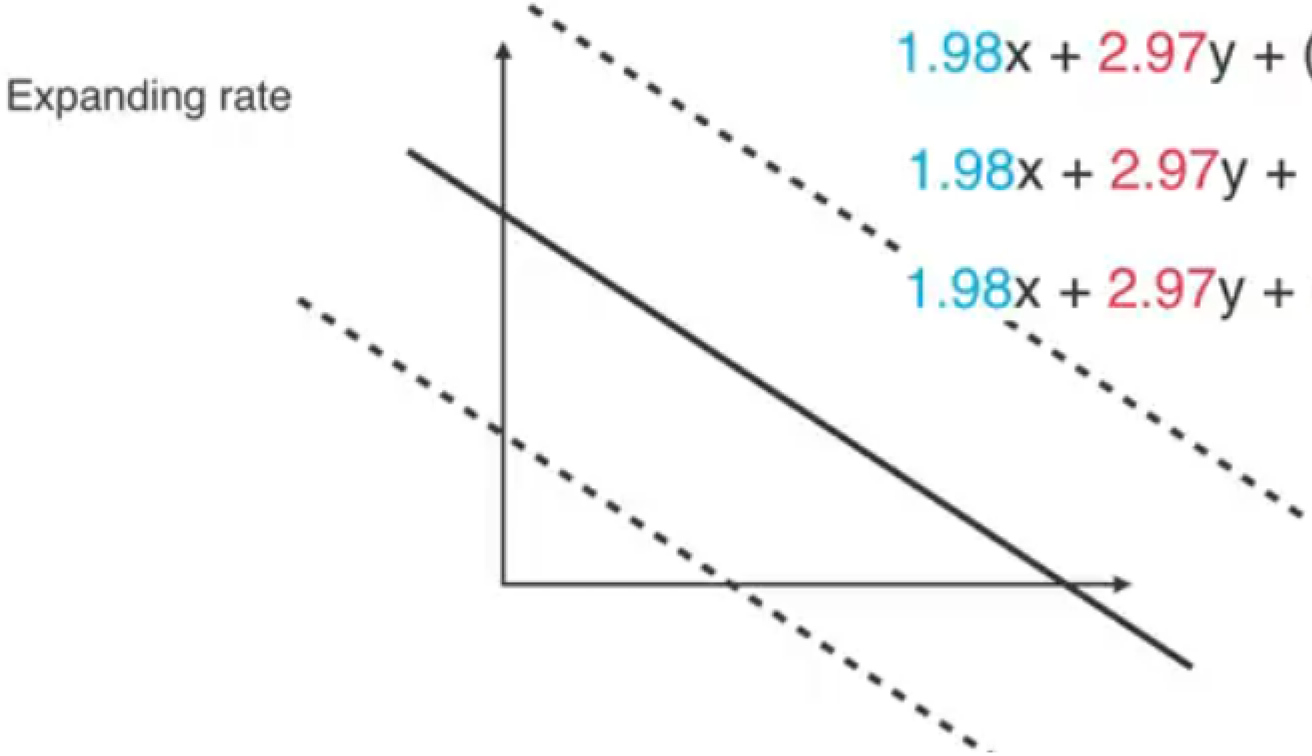


$$1.98x + 2.97y + (-5.94) = -1$$

$$1.98x + 2.97y + (-5.94) = 0$$

$$1.98x + 2.97y + (-5.94) = 1$$

Expanding rate

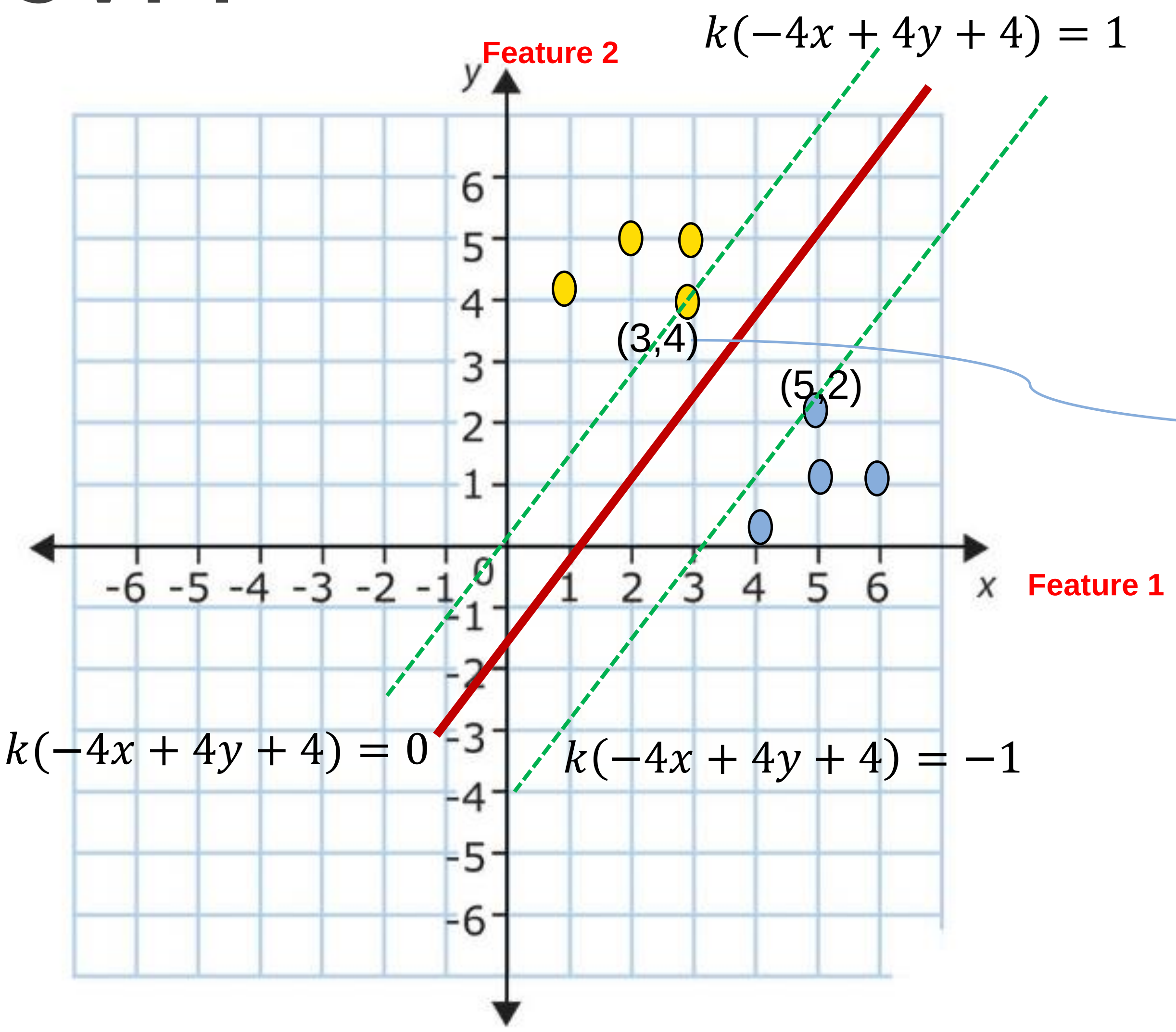


$$1.98x + 2.97y + (-5.94) = -1$$

$$1.98x + 2.97y + (-5.94) = 0$$

$$1.98x + 2.97y + (-5.94) = 1$$

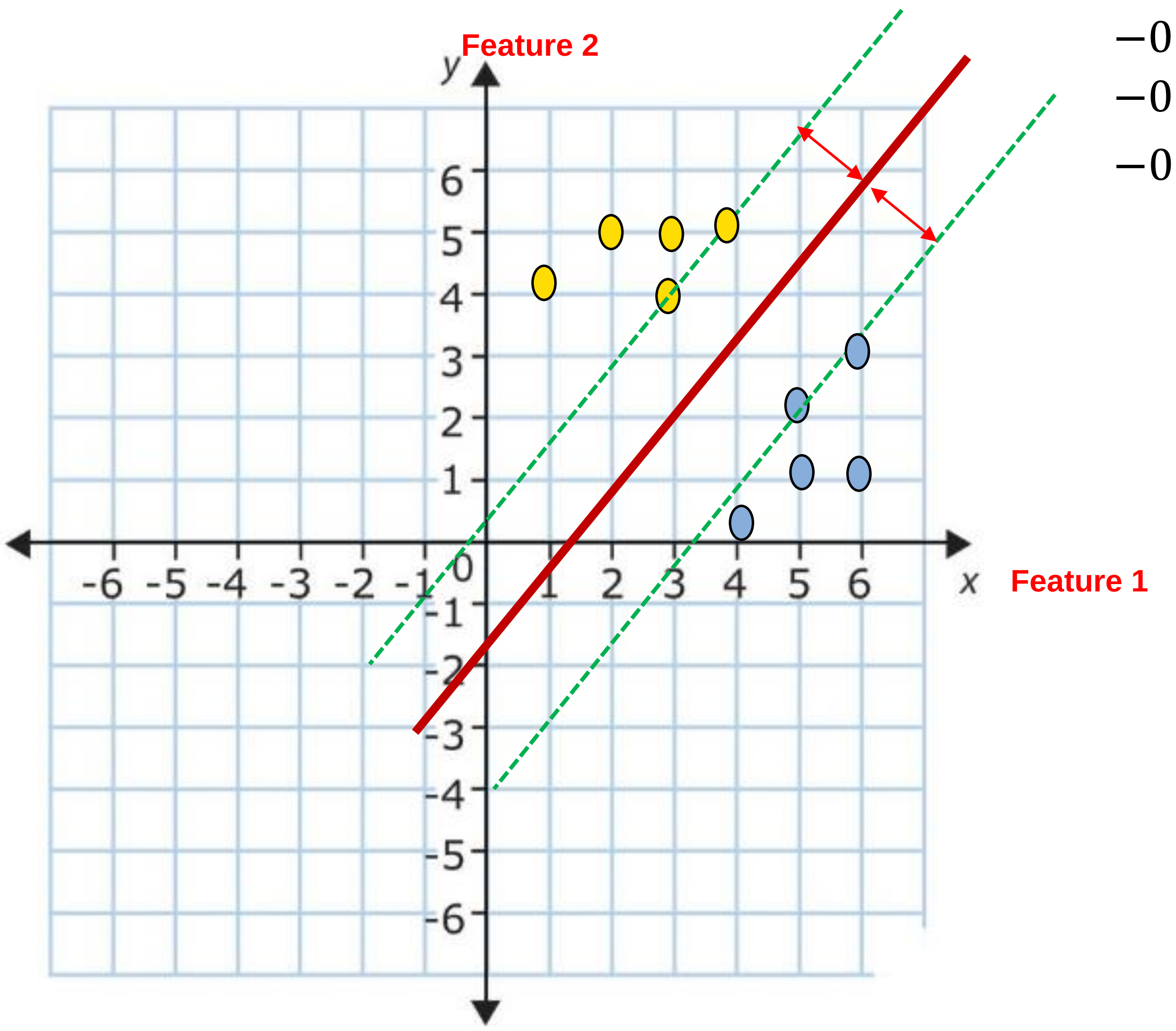
SVM



$k(-4(3)+4(4)+4)=1$ $\quad k=\frac{1}{8}$

Can we normalize the equations

SVM



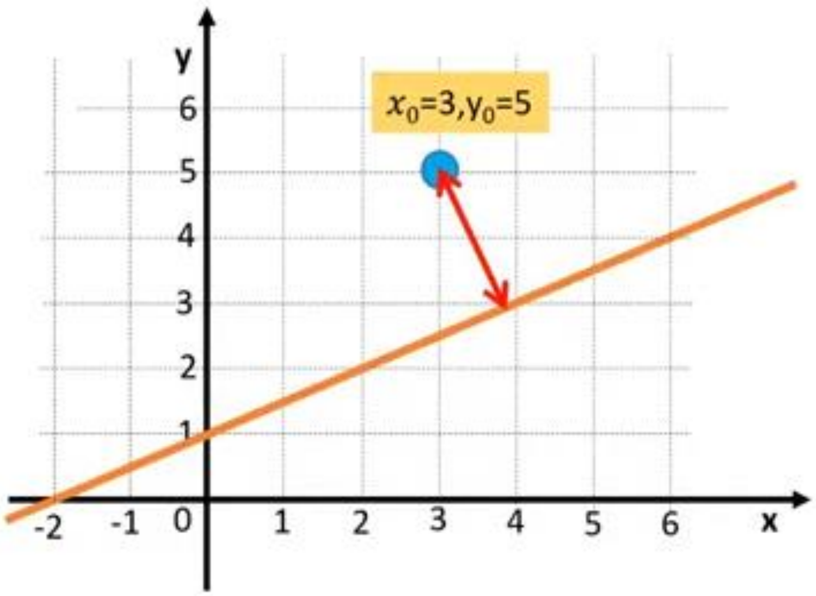
$$-0.5x + 0.5y + 0.5 = 1$$
$$-0.5x + 0.5y + 0.5 = 0$$
$$-0.5x + 0.5y + 0.5 = -1$$

$$b = 0.5$$

$$-0.5x + 0.5y + b = 1$$
$$-0.5x + 0.5y + b = 0$$
$$-0.5x + 0.5y + b = -1$$

The red line in the center between two lines, this is the reason to use 1 and -1 in the above equations.

Note



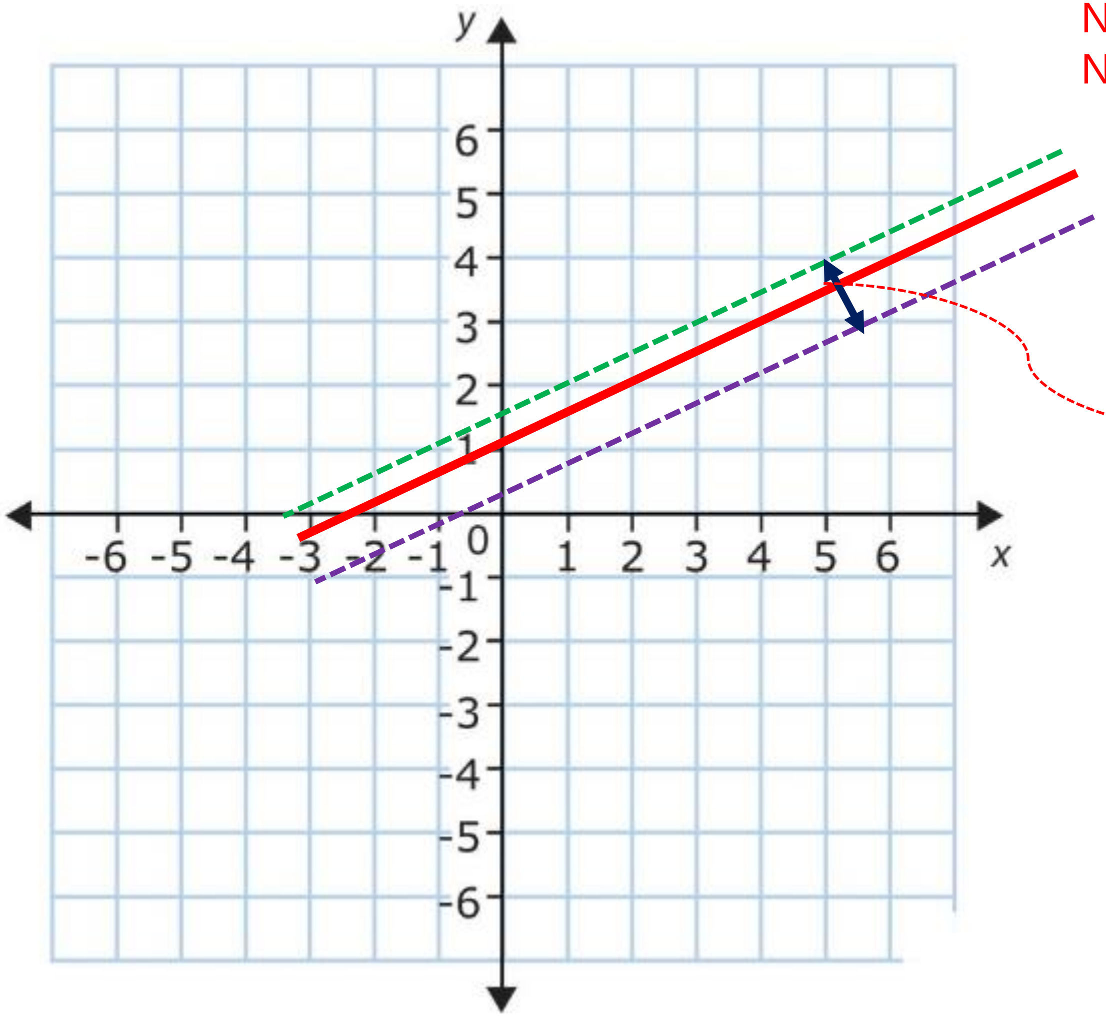
$$d = \frac{|Ax_0 + By_o + C|}{\sqrt{A^2 + B^2}}$$

Short distance between a point and a line.

SVM

$$Ax + By + C = 0 \qquad W^T x + b = 0$$

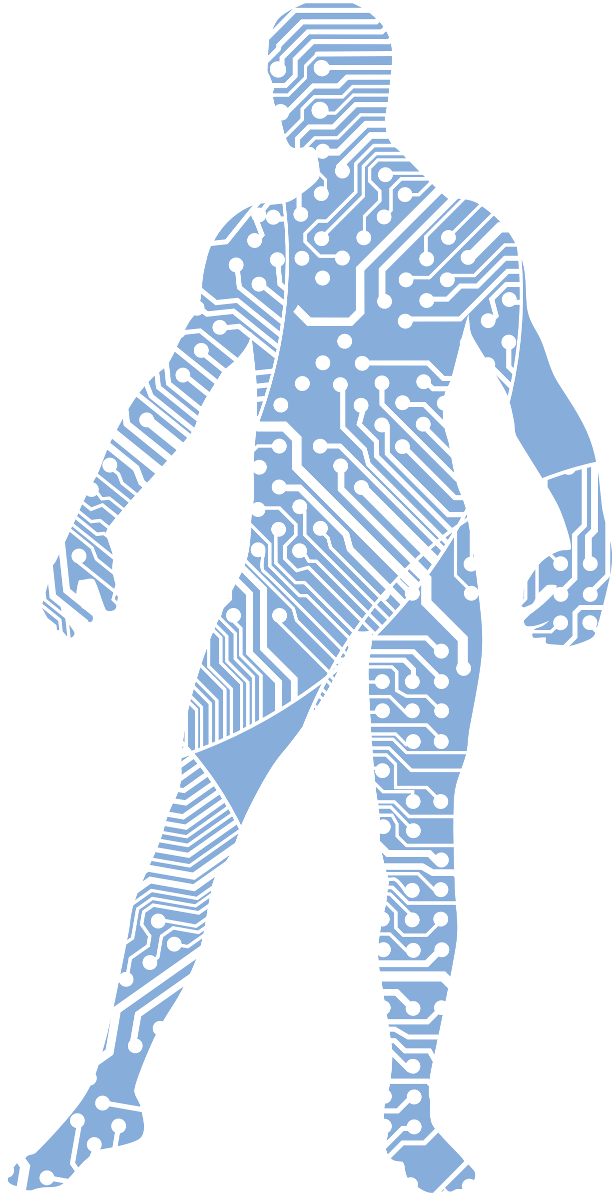
Notes: A and B are weights
Notes: x and y are features



$$d = \frac{|C_2 - C_1|}{\sqrt{A^2 + B^2}}$$

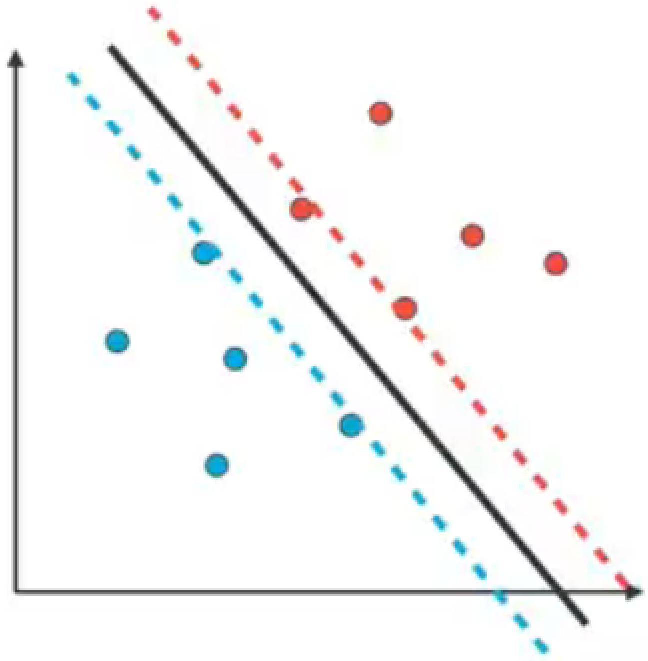
$$d = \frac{2}{\sqrt{A^2 + B^2}}$$

$$d = \frac{2}{\|w\|}$$



SVM tries to find the optimal weights so the distance between two blue lines is maximized
Or minimize the denominator because this will also maximize the distance between the lines.
The area between the two lines is called margin, SVM tries to find a hyper-plane that maximize the width of this margin

SVM algorithm



Step 1: Start with a line, and two equidistant parallel lines to it.

Step 2: Pick a large number. 1000 (number of repetitions, or epochs)

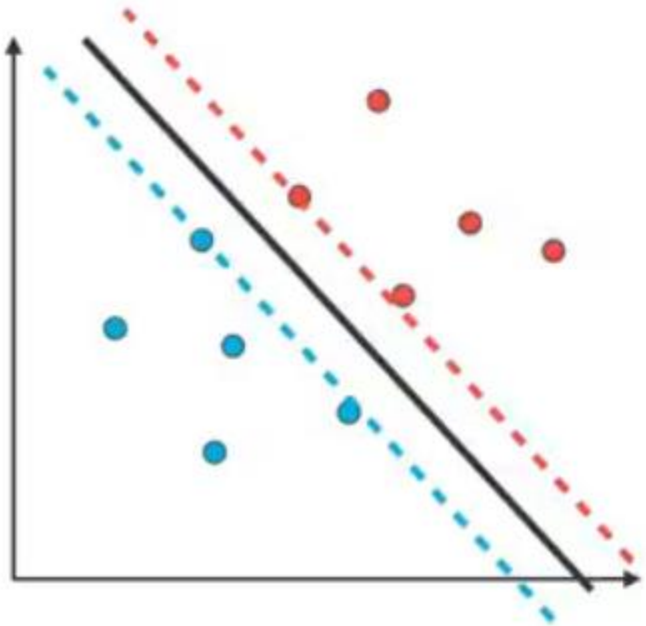
Step 3: Pick a number close to 1. (the expanding factor) 0.99

Step 4: (repeat 1000 times)

- Pick random point
- If point is correctly classified:
  - Do nothing
- If point is incorrectly classified
  - Move line towards point
- Separate the lines using the expanding factor

Step 5: Enjoy your lines that separate the data!

SVM algorithm



Step 1: Start with a random line of equation ax + by + c = 0.
Draw parallel lines with equations:

- ax + by + c = 1, and
- ax + by + c = -1

Step 2: Pick a large number. **1000** (number of repetitions, or epochs)

Step 3: Pick a learning rate. **0.01**

Step 4: Pick an expanding rate. 0.99

Step 5: (repeat **1000** times)

- Pick random point **(p,q)**
- If point is correctly classified
    - Do nothing
- If point is blue, and ap+bq+c > 0
    - Subtract 0.01**p** to a
    - Subtract 0.01**q** to b
    - Subtract 0.01 to c
- If point is, red and ap+bq+c < 0
    - Add 0.01**p** to a
    - Add 0.01**q** to b
    - Add 0.01 to c

- **Multiply a, b, c, by** 0.99