Basics are important like: what is support vector, margin width, etc.

Equation of Hyperplane

*SVM e loss function kibhabe kaj kore

Soft margin er equation explain

<u>Q:</u>

Hard margin: Basic SVM, support vector, margin width, figure

Soft margin: Zeta term, how it reduces misclassification loss functions, role of C hyperparameters

Kernel trick: Concept, figure, name and equations of kernel functions

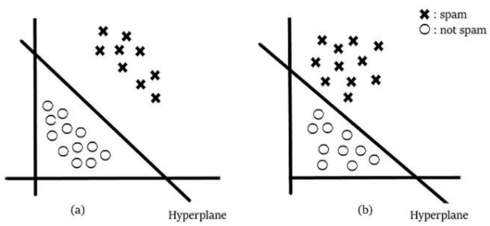
HM vs **SM**

What is SVM:

SVM is a supervised learning algorithm for classification and regression tasks. It is used for small datasets and for complex relationships (e.g. image classification).

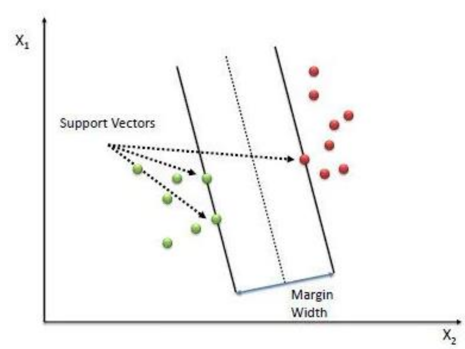
What is Hyperplane:

It is a decision boundary (or a straight line) that separates data points into different classes.
SVM's goal is to create this hyperplane optimally.



Support Vector: The points closest to the hyperplane are called as the support vector points.

Margin: The distance of the vectors from the hyperplane are called margin. Margin should be as large as possible.



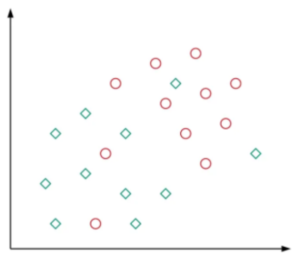
Equation of hyperplane:
General equation of hyperplane is:
$y = w^TX + b$; w is the weight, X is the input feature, b is the bias

SVM is two types:

Linear SVM and **Non-linear** SVM.

Linear SVM means data points are linearly separable that means we can separate the data points with single straight line only and so this is called **Hard Margin** as it strictly separates the data and there is no chance to classify the data wrongly.

Non Linear SVM means data points can't be separated using just a straight line. So, this is solved using two ways, **i) Soft Margin ii) Kernel trick**.



These points are scattered so we can' t separate using straight line.

i) Soft margin:

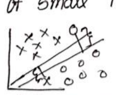
Soft margin allows some wrong classifications or some small mistakes but also keeps the margin as large as possible. It does this by using following the equation $\text{Minimize: } \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i$.

Explaining the equation of Soft Margin:

$$\text{Minimize: } \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i$$

Here, w is the margin width that we want to keep as large as possible.
C is a hyperparameter that has two roles: i) **maximizes the margin**(w) ii) **minimizes the number of mistakes or wrong classifications.** So, small value of C means maximizing **w** instead of minimizing the mistakes and vice versa for large value of C
ξ is the Zeta term that represents how much mistake or wrong classification is done.

· Benefits of Small Mistake em,

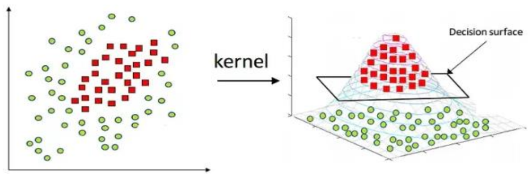


* এই distance গুলোর যোগফল হলো zeta

Benefit of small mistakes:

Allowing small mistakes in SVM helps it focus on maximizing the margin, making the model generalize better to unseen data and handle non-linearly separable datasets. This also reduces overfitting.

ii) Kernel trick:



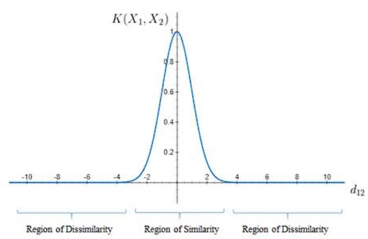
The diagram shows what **Kernel** trick does.
This trick basically **adds a dimension** to the dataset for making it linearly separable.

Here in the diagram, the 2D data is not linearly separable at all. But if we take it to 3D, then it becomes linearly separable. How is it done? Suppose we have (x_1, x_2, x_3,…,x_n) data points, so kernel function adds a dimension and makes it **(**x_1, x_2, x_3,…,x_n, f(x1, x2, x3,…,xn)**)**.

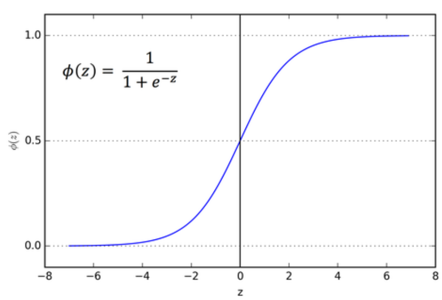
Types of kernel functions:

1. Gaussian kernel radial basis function (RBF):

$$K(X_1, X_2) = exp(-\frac{||X_1 - X_2||^2}{2})$$



2. Sigmoid function:



3. Polynomial kernel:

$$K(X, Y) = (\gamma \cdot X^T Y + r)^d, \gamma > 0$$

4. Linear kernel:
   f(x_1, x_2, x_3,…, x_n) = $w_0+w_1x_1+w_2x_2+w_3x_3…+w_nx_n$

Hard Margin (HM) vs Soft Margin (SM):

Aspect	Hard Margin (HM)	Soft Margin (SM)
Tolerance	No misclassification allowed.	Allows some misclassification (small mistakes).
Use Case	Perfectly separable data.	Noisy or overlapping data.
Flexibility	Rigid, sensitive to outliers.	Flexible, handles noise and outliers well.