Support Vector Machines- an overview

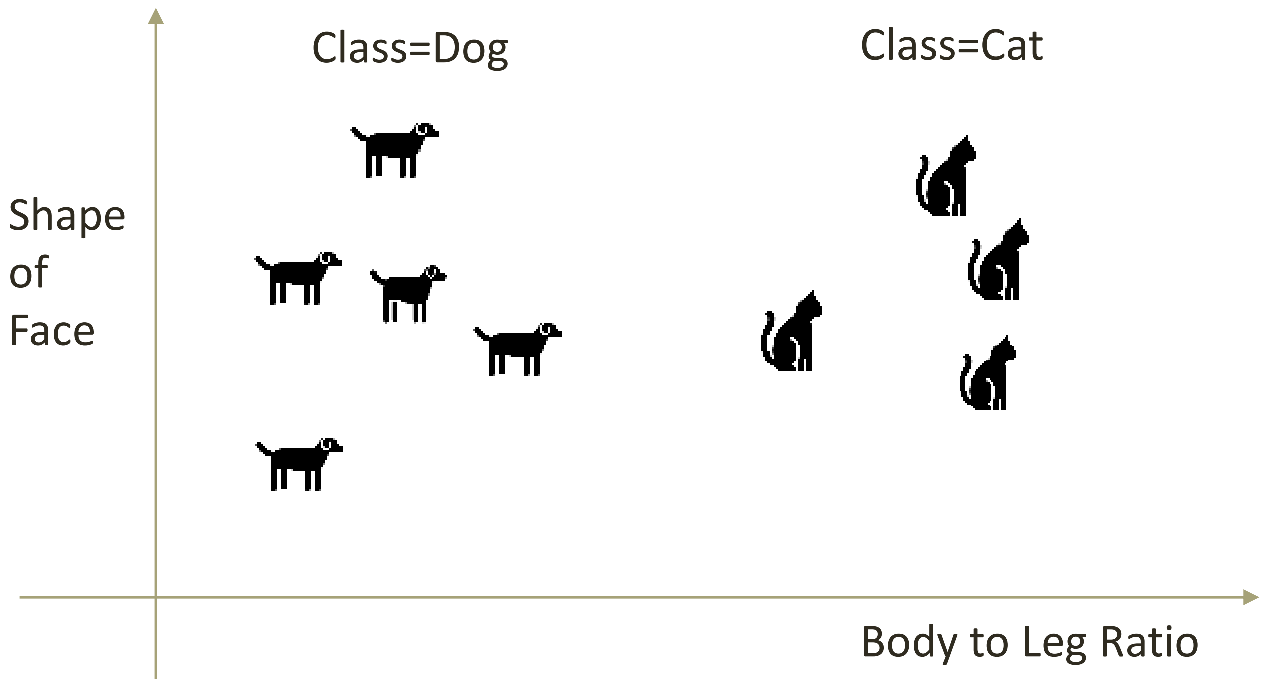
B.Sc. 6th Semester (Data Mining:Paper Code: DSE-4)

Paulami Basu Ray
Assistant Professor
Department of Computer Science & Applications
Prabhat Kumar College, Contai

What is a Support Vector Machine?

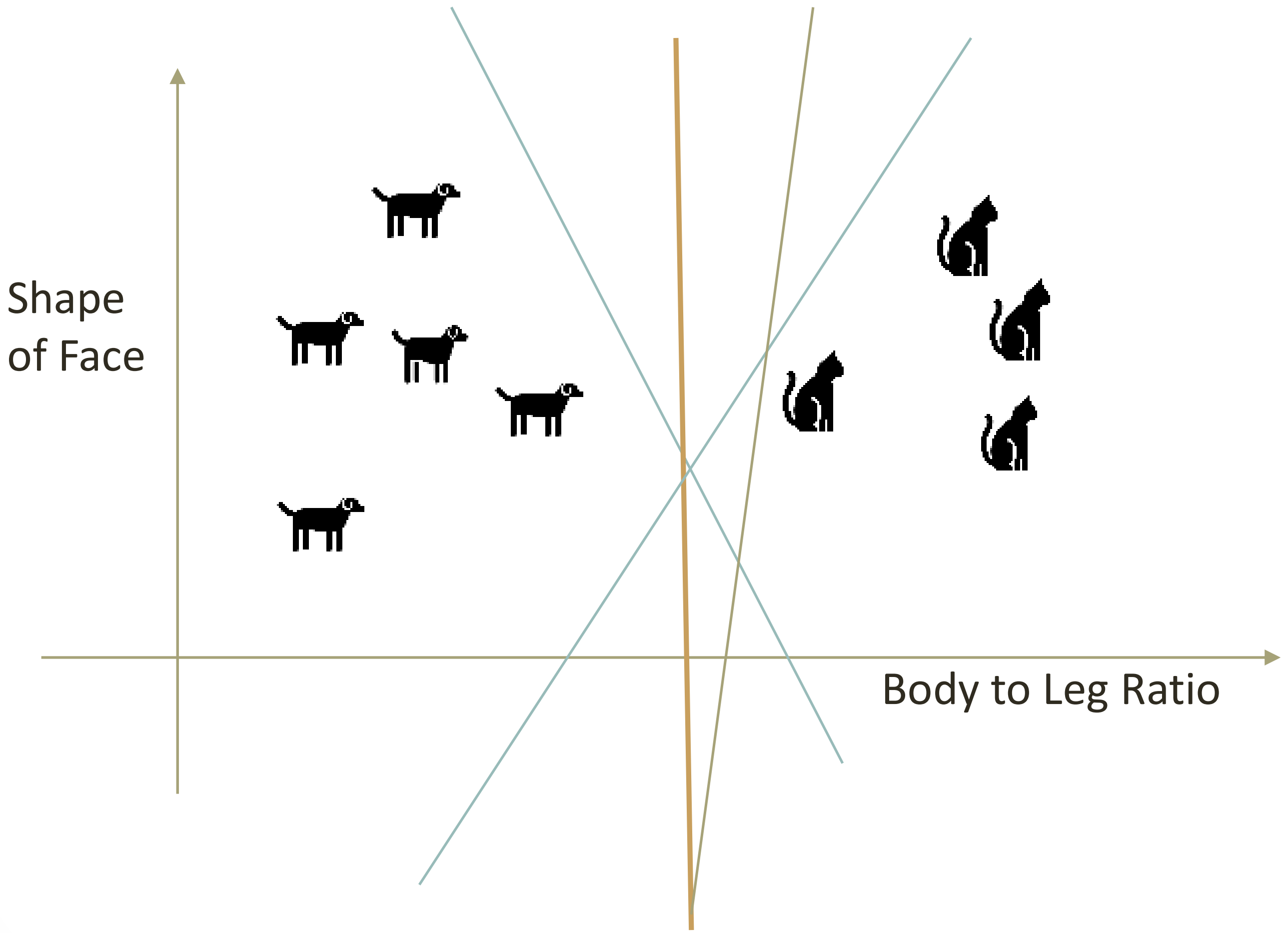
- SVM is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems.
- In the SVM-algorithm we plot each data item as a point in n-dimensional space (where n is the number of features you have) with the value of each feature being the value of a co-ordinate.

Feature Space



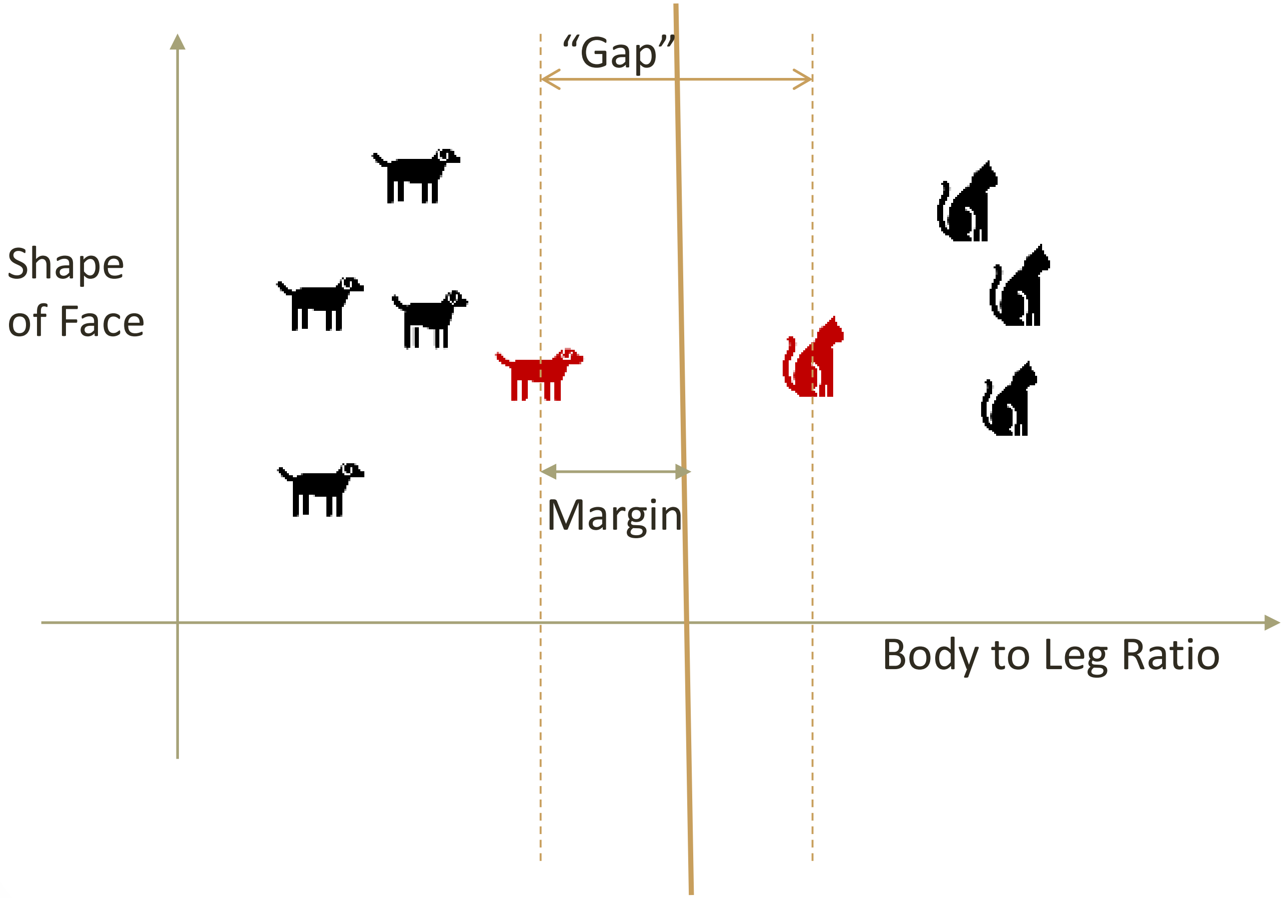
Here, the dimension of the feature space is two and it is a 2-Class Classification problem. Now, given the features of a new animal I should be able to classify it into either a Dog or a Cat.

Linear Discriminant Analysis



My aim is to draw a perfect class boundary among infinite possibilities of class boundaries.

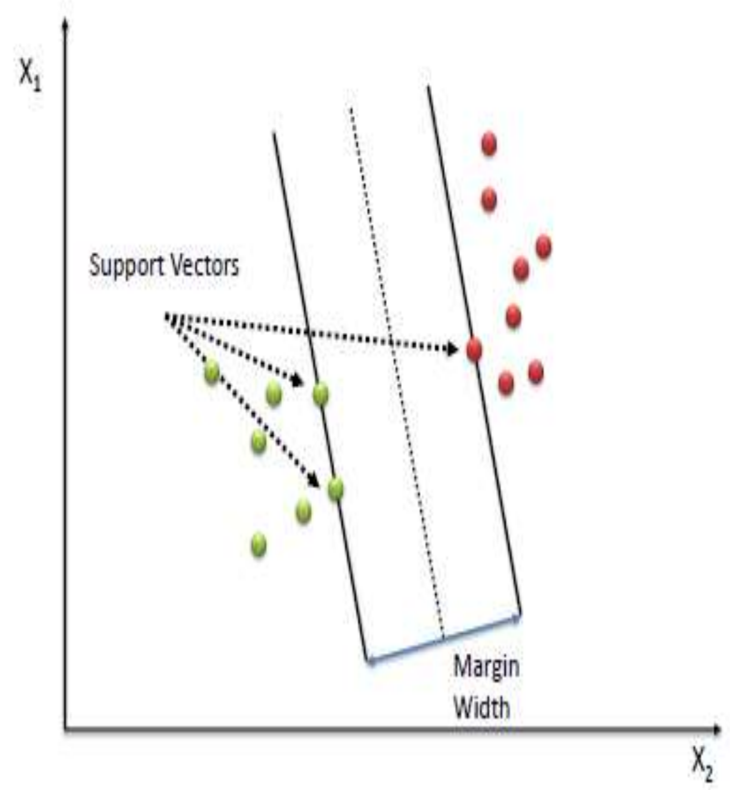
Linear Discriminant Analysis



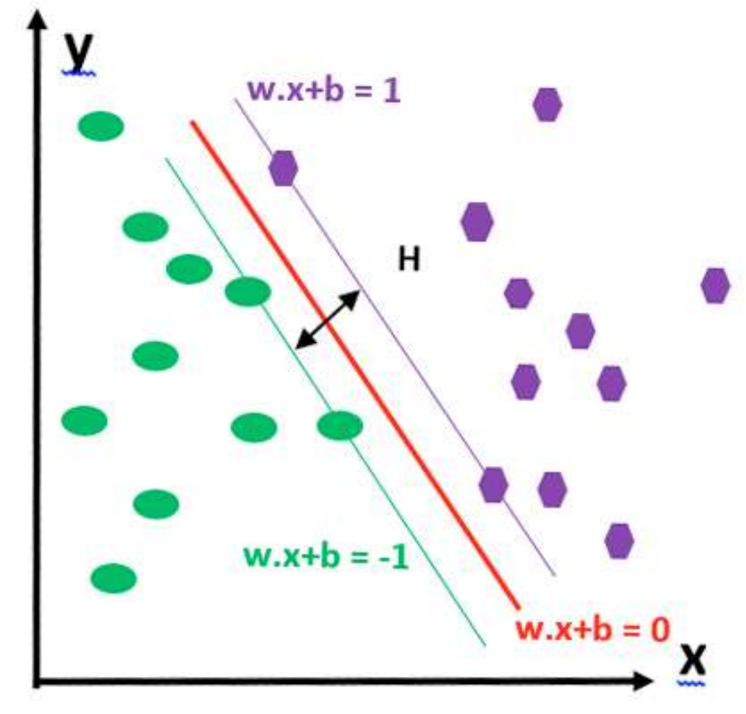
I want the nearest cat and the nearest dog to be well separated. This is the idea of a perfect separator.
SVM is a generalization of a Maximal/Optimal Margin Classifier(MMC/OMC)

Hyperplane, Support Vector

- A **hyperplane** in an n-dimensional Euclidean space is a flat, n-1 dimensional subset of that space that divides the space into two disconnected parts.
- In 2D the hyperplane is a line(Linear Discriminator), in 3D it is plane, when more than 3D it is referred to as a hyperplane.
- **Support Vectors** are the data points that lie closest to the decision surface/hyperplane. They are the data points most difficult to classify.



Hyperplane – Equations..



- Equation of the hyperplane:

  $w_1x_1 + w_2x_2 + \ldots\ldots + w_nx_n + b = 0$,

where W is the weight vector and x_1, x_2,…. represents the features.

- In case of a 2D feature space the equation reduces to:

$$w_1x_1 + w_2x_2 + b = 0$$

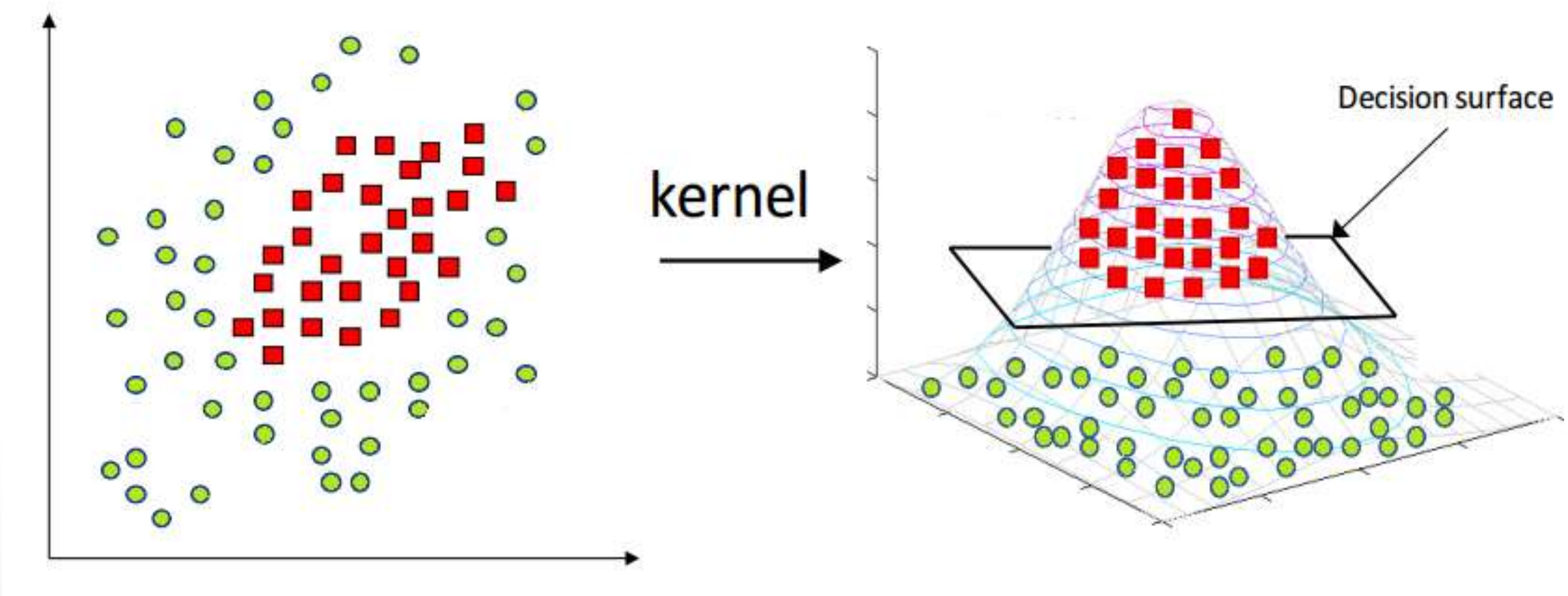
$$w_1x_1 + w_2x_2 + b \geq +1 \text{ when } y = +1$$

$$w_1x_1 + w_2x_2 + b \leq -1 \text{ when } y = -1$$

Optimal Margin Classifier

- Distance between a point (x_1, x_2) and a line $w_1x + w_2y + b = 0$
- is given by $|w_1x_1 + w_2x_2 + b| / \sqrt{(w_1^2 + w_2^2)}$
- ${|w_1x_1 + w_2x_2 + b|}/_{||w||}$
- ${}^{1}/_{||w||}$ [We have introduced scaling such that functional margin is 1]
- Coverting the maximization problem to a minimization problem:
- $min_{w,b} \frac{1}{2} ||w||^2$
- *such that :* $y^{(i)} (w^T x^{(i)} + b) \geq 1$ *, i=1,……,m*
- *It is an optimization problem with a convex quadratic objective function and only linear constraints.*

Non-Linearly Separable data



The "Kernel Trick"

- For all x and x ′ in the input space X ,certain functions k (x , x ′) can be expressed as an inner product in another space V .
- The function k : X × X → R is often referred to as a kernel or a kernel function