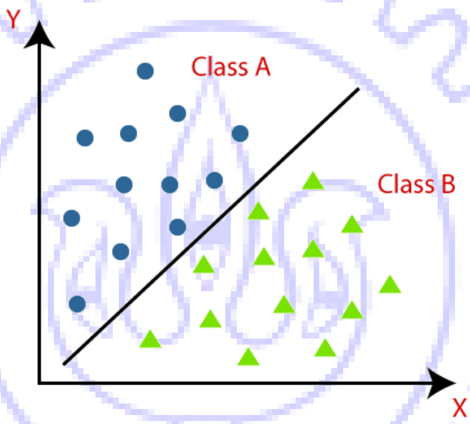
Live Interaction #3:

27th January 2024

E-masters Next Generation Wireless Technologies

EE902 Advanced ML Techniques for Wireless Technology

- **Support Vector Machines:**



- **Linear classifier**:

$$\bar{\mathbf{a}}^T\bar{\mathbf{x}} = b$$
$$\bar{\mathbf{a}}^T\bar{\mathbf{x}} \geq b$$
$$\bar{\mathbf{a}}^T\bar{\mathbf{x}} < b$$

- In 2 dimensions

$$a_1x_1 + a_2x_2 = b$$
$$\mathcal{C}_0: a_1x_1 + a_2x_2 \geq b$$
$$\mathcal{C}_1: a_1x_1 + a_2x_2 < b$$

- Two sets of points

$$\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \ldots, \bar{\mathbf{x}}_M : \mathcal{C}_0$$

$$\bar{\mathbf{x}}_{M+1}, \bar{\mathbf{x}}_{M+2}, \ldots, \bar{\mathbf{x}}_{M+L} : \mathcal{C}_1$$

- How to determine the linear classifier?

$$\mathcal{C}_0 : \bar{\mathbf{a}}^T \bar{\mathbf{x}}_i \geq b, i = 1,2,..,M$$

$$\mathcal{C}_1 : \bar{\mathbf{a}}^T \bar{\mathbf{x}}_i \leq b, i = M+1, M+2,..,M+L$$

- This leads to the **trivial solution**!

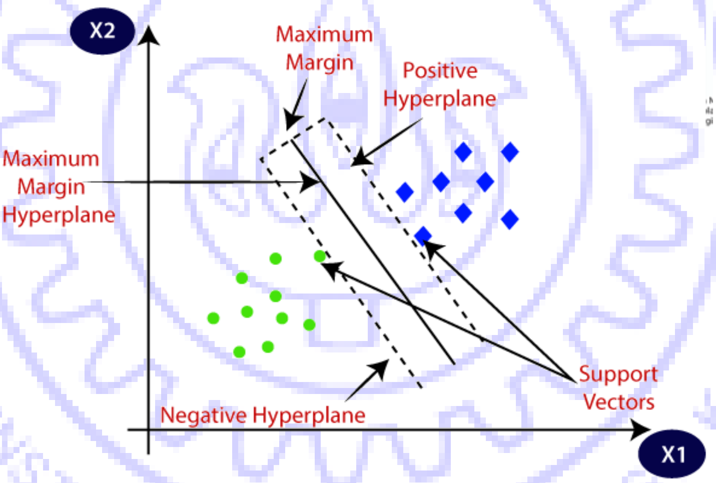
$$\bar{\mathbf{a}} = \mathbf{0}, b = 0$$

- How to modify this to avoid the trivial solution?

$$\mathcal{C}_0 : \bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \geq 1, i = 1,2,..,M$$

$$\mathcal{C}_1 : \bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \leq -1, i = M+1, M+2,..,M+L$$

- Two **parallel hyperplanes** – **Positive hyperplane**, **negative hyperplane**.



- The distance between positive and negative hyperplane is called the **margin**.
- **Aim of the classifier**: **Maximize the margin**!

$$\bar{\mathbf{a}}^T \bar{\mathbf{x}} = c_1$$

$$\bar{\mathbf{a}}^T \bar{\mathbf{x}} = c_2$$

- What is the **distance** between the parallel hyperplanes?

$$\frac{|c_1 - c_2|}{\|\bar{\mathbf{a}}\|}$$

$$\|\bar{\mathbf{a}}\|_2 = \sqrt{|a_1|^2 + |a_2|^2 + \cdots}$$

$$\|\bar{\mathbf{a}}\|_1 = |a_1| + |a_2| + \cdots$$

$$\underbrace{\|\bar{\mathbf{a}}\|_p = \left(|a_1|^p + |a_2|^p + \cdots\right)^{\frac{1}{p}}}_{l_p \text{ norm}}$$

- Coming to our problem

$$\mathcal{C}_0: \bar{\mathbf{a}}^T \mathbf{x} = 1 - b = c_1$$

$$\mathcal{C}_1: \bar{\mathbf{a}}^T \bar{\mathbf{x}} = -1 - b = c_2$$

- Distance between positive and negative hyperplanes is

$$\frac{2}{\|\bar{\mathbf{a}}\|}$$

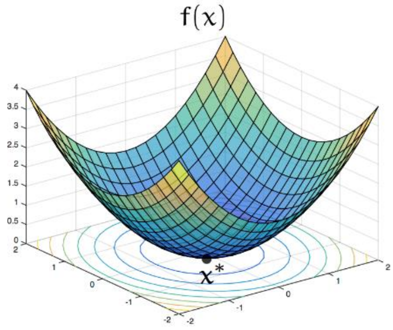
- **Maximize the margin**!!!
- **Good fences make good neighbours**!!!
- **Good margins make good classifiers**!!!
- Optimization problem

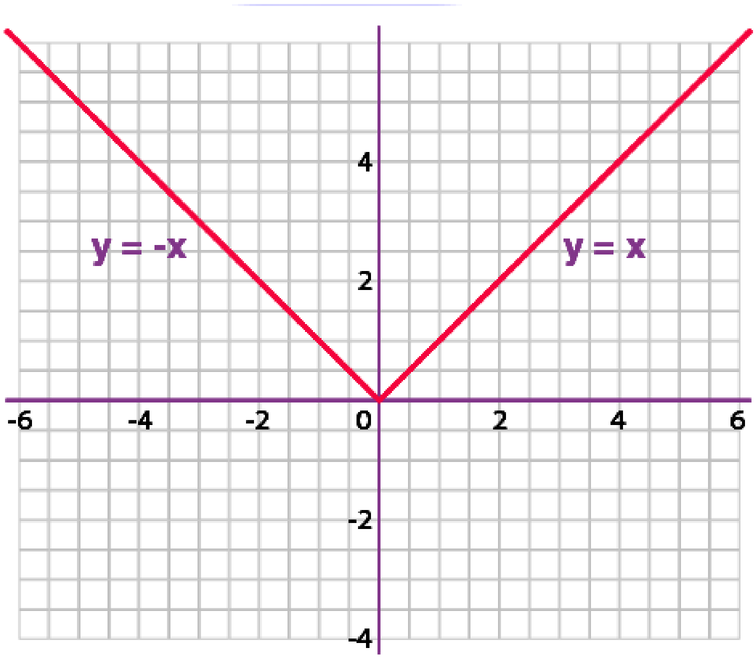
$$\min \|\bar{\mathbf{a}}\|$$

$$\mathcal{C}_0: \bar{\mathbf{a}}^T \bar{\mathbf{x}}_i \geq 1 - b, i = 1,2,..,M$$

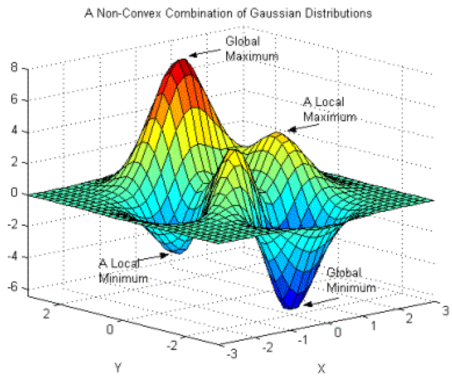
$$\mathcal{C}_1: \bar{\mathbf{a}}^T \bar{\mathbf{x}}_i \leq -1 - b, i = M+1, M+2,..,M+L$$

- Objective function, constraints

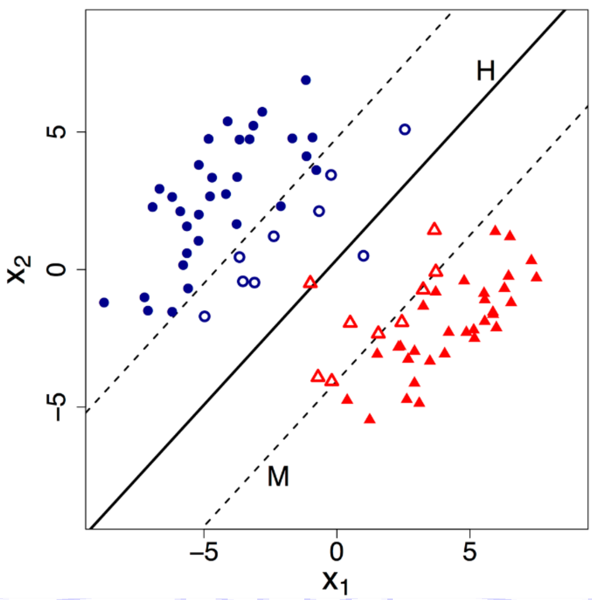




- This is known as a **convex optimization problem**.
- Non-convex



- Why do we call it **support vector machine**?
- **Soft classifier:**



- Linear separation is **NOT possible**.
- In such a scenario we **tolerate classification error**.
- We have to introduce a **relaxation**.

$$\underbrace{\min \sum_i u_i + \sum_j v_j}_{\text{Total slack}}$$

$$\mathcal{C}_0: \bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \geq 1 - u_i, i = 1,2,..,M$$

$$\mathcal{C}_1: \bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \leq -1 + v_i, i = M+1, M+2,..,M+L$$

$$u_i \geq 0$$

$$v_i \geq 0$$

- u_i, v_i: **Slack variables**
- **Part II - Kernel SVM**.
- **Next week: Naïve Bayes**!!
- **Assignment #1, 2 Discussion: 27th January Saturday 4:30 PM onward**.
- **Quiz #1: 28th January Sunday 2:00-3:00 PM.**
- **Assignment #3 deadline: Feb 2nd Friday 11:59 PM.**
- **Live interaction #4: Feb 3rd Saturday 10:30 AM – 11:30 AM**.