Support Vector Machines(SVM)

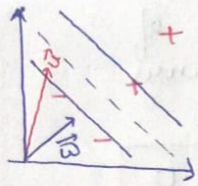
Support Vector Machine (SVM) is a machine learning algorithm mainly used for classification tasks but can also handle regression. The goal of SVM is to find a line (or a hyperplane in higher dimensions) that best separates data into different classes.

- **Data Points and Classes**: Imagine you have two groups of points on a graph, and each group belongs to a different class. The job of the SVM is to find a boundary (called a hyperplane) that separates these two groups as clearly as possible.
- **Maximum Margin**: SVM looks for the hyperplane that has the largest margin between the two classes. The margin is the distance between the hyperplane and the nearest data points from each class. These nearest points are called **support vectors**.
- **Separation**: If the data can be separated perfectly by a straight line (in 2D) or a flat plane (in 3D), it's called linearly separable. If not, SVM can use a technique called **kernel trick** to map the data into a higher dimension where it can be separated more easily.
- **Objective**: The main objective of SVM is to not only classify data but also to do so with the maximum margin, which helps in making the model more generalizable and less prone to overfitting.

Day 5: Support Vector Machines (SVM)

→ It is a supervised Machine learning algorithm that classifies data by finding an optimal line or hyperplane that maximizes the distance between each class in an N-dimensional space

- SVM:



$\vec{w} \cdot \vec{u} \geq c \qquad c = -b$

1

$$\boxed{\vec{w} \cdot \vec{u} + b \geq 0 \text{ Then } +}$$

The Decision Rule

$\vec{w} \cdot \vec{x}_+ + b \geq 1$

$x_+ \rightarrow$ positive sample

$\vec{w} \cdot \vec{x}_- + b \leq -1$

$x_- \rightarrow$ Negative sample

y_i such that $y_i = +1$ for + sample
$= -1$ for − sample

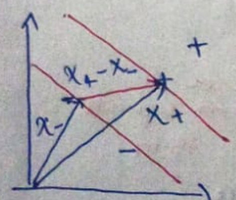
$$\left.\begin{aligned} y_i(x_i \vec{w} + b) &\geq 1 \\ y_i(x_i \cdot \vec{w} + b) &\geq 1 \end{aligned}\right\} \text{ from upper equations}$$

$$y_i(x_i \cdot \vec{w} + b) - 1 \geq 0$$

2

$$\boxed{y_i(\vec{x}_i \cdot \vec{w} + b) - 1 = 0}$$

for x_i in Gutter



3. $\text{width} = x_+ - x_-$ } difference vector

$$\boxed{\text{width} = (x_+ - x_-) \cdot \frac{\vec{w}}{\|w\|} \Rightarrow \frac{2}{\|w\|}}$$

$1 - b \qquad 1 + b$

$$\text{width} = \frac{2}{\|w\|} \rightarrow \text{Maximize}$$

$$\frac{1}{\|w\|} \rightarrow \text{maximize} \rightarrow \|w\| \rightarrow \text{minimize} \rightarrow \boxed{\tfrac{1}{2}\|w\|^2 \rightarrow \text{minimize}}$$

* Langrange Multipliers.

$$L = \frac{1}{2}\|w\|^2 - \sum \alpha_i \left[y_i(\vec{w}\cdot\vec{x}_i + b) - 1 \right]$$

$$\frac{\partial L}{\partial w} = \vec{w} - \sum \alpha_i y_i x_i = 0$$

$$\Rightarrow \boxed{\vec{w} = \sum_i \alpha_i y_i x_i}$$ } linear sum of these vectors

$$\frac{\partial L}{\partial b} = -\sum \alpha_i y_i = 0 \Rightarrow \boxed{\sum \alpha_i y_i = 0}$$

substituting $\vec{w}$

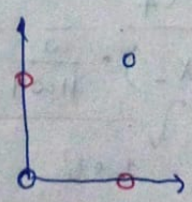
$$L = \frac{1}{2}\left(\sum \vec{\alpha}_i \vec{y}_i \vec{x}_i\right)\left(\sum \vec{\alpha}_j \vec{y}_j \vec{x}_j\right) - \left(\sum \alpha_i y_i x_i\right)\cdot\left(\sum \alpha_j y_j x_j\right) - \underbrace{\sum \alpha_i y_i}_{0} b + \sum \alpha_i$$

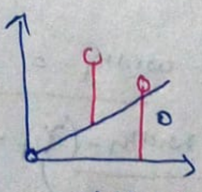
$$\Rightarrow \boxed{\sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \boxed{x_i \cdot x_j}}$$

Here we see it depends upon dot product of vectors

$$\boxed{\sum \vec{\alpha}_i y_i x_i \cdot \vec{u} + b > 0}$$ Then +

Decision rule





Here we can seperate them

different space

$\phi(x_i)\cdot\phi(x_j)$ to Max

$\phi(x_j)\cdot\phi(u)$

Kernel $\Rightarrow (\vec{u}\cdot v + 1)^n$

$n = 2$ for SVM

$K(x_i, x_j) = \phi(x_i)\cdot\phi(\bar{x}_j)$