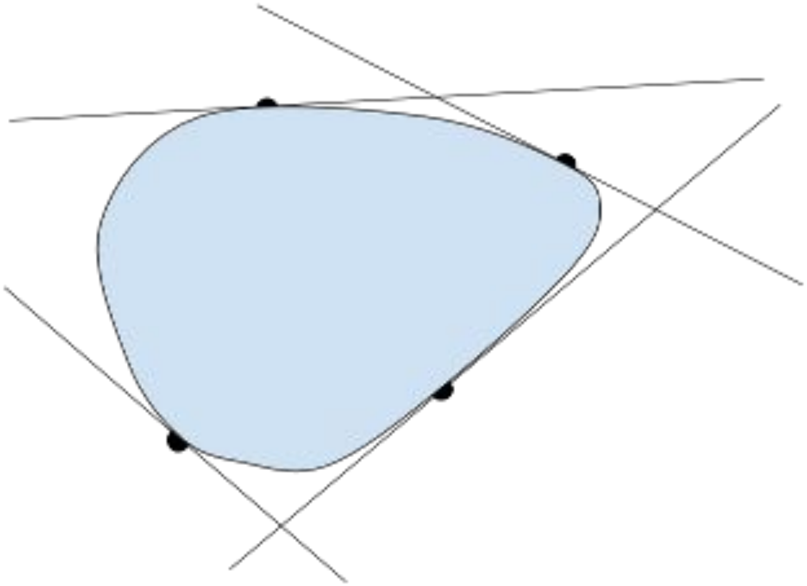
Optimization for ML: Convex Sets

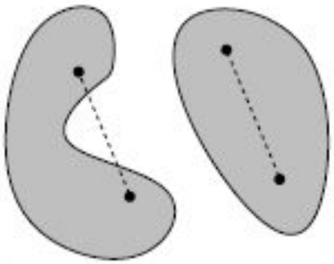


Office: MB 113,AV 106

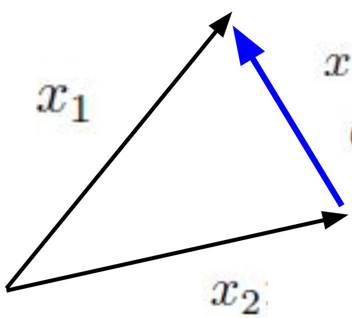
vidyapradananda@gm.rkmvu.ac.in

Definition of Convex set

Def. Set $C \subset \mathbb{R}^n$ called **convex**, if for any $x, y \in C$, the line-segment $\lambda x + (1-\lambda)y$, where $\lambda \in [0,1]$, also lies in C.



- **Convex:** $\lambda_1 x + \lambda_2 y \in C$, where $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$.
- **Linear:** if restrictions on λ_1, λ_2 are dropped
- **Conic:** if restriction $\lambda_1 + \lambda_2 = 1$ is dropped



$x = \theta_1 x_1 + \theta_2 x_2$

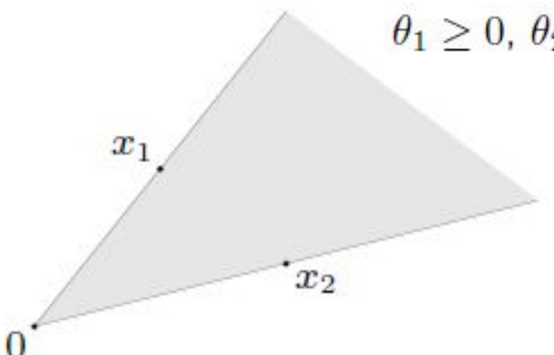
$\theta_1 \geq 0,\ \theta_2 \geq 0$

$\theta_1 + \theta_2 = 1$

convex combination of two vectors lie in the line joining the two vectors

$x = \theta_1 x_1 + \theta_2 x_2$

$\theta_1 \geq 0,\ \theta_2 \geq 0$



convex cone: set that contains all conic combinations of points

Examples of Convex sets

(a) n-dimensional Euclidean space, $\mathbb{R}^n$. Given $x, y \in \mathbb{R}^n$, we must have $\lambda x + (1-\lambda)y \in \mathbb{R}^n$.

(b) Nonnegative orthant, $\mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x_i \geq 0, i = 1, \ldots, n\}$. Let $x, y \in \mathbb{R}^n_+$ be given. Then for any $\lambda \in [0,1]$,

$$(\lambda x + (1-\lambda)y)_i = \lambda x_i + (1-\lambda)y_i \geq 0.$$

(c) Balls defined by an arbitrary norm, $\{x \in \mathbb{R}^n | \|x\| \leq 1\}$ (e.g., the l_2 norm $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ or l_1 norm $\|x\|_1 = \sum_{i=1}^n |x_i|$ balls). To show this set is convex, it suffices to apply the Triangular inequality and the positive homogeneity associated with a norm. Suppose that $\|x\| \leq 1, \|y\| \leq 1$ and $\lambda \in [0,1]$. Then

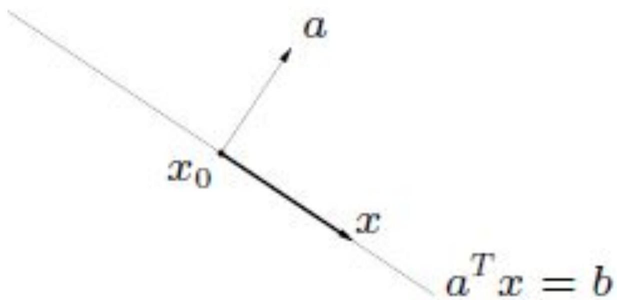
$$\|\lambda x + (1-\lambda)y\| \leq \|\lambda x\| + \|(1-\lambda)y\| = \lambda\|x\| + (1-\lambda)\|y\| \leq 1.$$

(d) Affine subspace, $\{x \in \mathbb{R}^n | Ax = b\}$. Suppose $x, y \in \mathbb{R}^n$, $Ax = b$, and $Ay = b$. Then

$$A(\lambda x + (1-\lambda)y) = \lambda Ax + (1-\lambda)Ay = b.$$

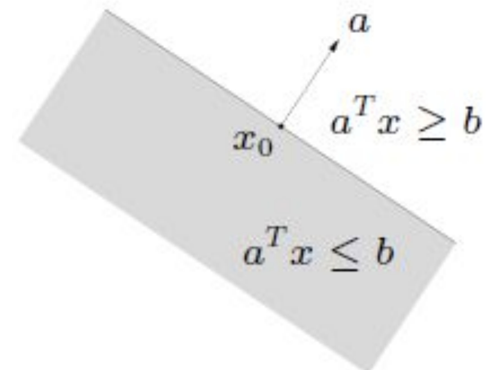
Examples of Convex sets

Hyperplane



$\{x \mid a^T x = b\}$ $(a \neq 0)$

Halfspace

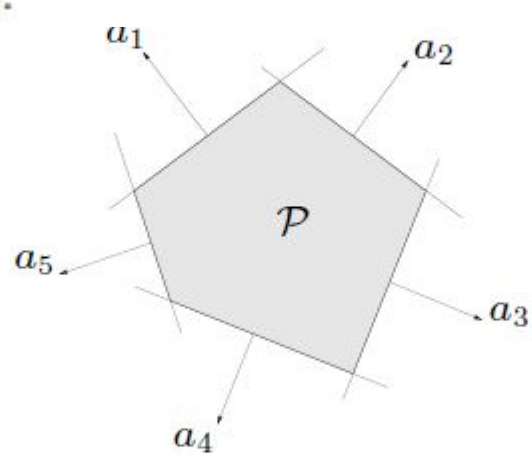


$\{x \mid a^T x \leq b\}$ $(a \neq 0)$

(e) Polyhedron, $\{x \in \mathbb{R}^n | Ax \leq b\}$. For any $x, y \in \mathbb{R}^n$ such that $Ax \leq b$ and $Ay \leq b$, we have

$$A(\lambda x + (1-\lambda)y) = \lambda Ax + (1-\lambda)Ay \leq b$$

for any $\lambda \in [0,1]$.



(polyhedron is intersection of finite number of halfspaces and hyperplanes)

Examples of Convex sets

(f) The set of all positive semidefinite matrices $\mathbb{S}^n_+$. $\mathbb{S}^n_+$ consists of all matrices $A \in \mathbb{R}^{n\times n}$ such that $A = A^T$ and $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$. Now consider $A, B \in \mathbb{S}^n_+$ and $\lambda \in [0,1]$. Then we must have

$$[\lambda A + (1-\lambda)B]^T = \lambda A^T + (1-\lambda)B^T = \lambda A + (1-\lambda)B.$$

Moreover, for any $x \in \mathbb{R}^n$,

$$x^T(\lambda A + (1-\lambda)B)x = \lambda x^T Ax + (1-\lambda)x^T Bx \geq 0.$$

(g) Intersections of convex sets. Let $X_i, i = 1,\ldots,k$, be convex sets. Assume that $x, y \in \cap_{i=1}^k X_i$, i.e., $x, y \in X_i$ for all $i = 1,\ldots,k$. Then for any $\lambda \in [0,1]$, we have $\lambda x + (1-\lambda)y \in X_i$ by the convexity of X_i, $i = 1,\ldots,k$, whence $\lambda x + (1-\lambda)y \in \cap_{i=1}^k X_i$.

(h) Weighted sums of convex sets. Let $X_1,\ldots,X_k \subseteq \mathbb{R}^n$ be nonempty convex subsets and $\lambda_1,\ldots,\lambda_k$ be reals. Then the set

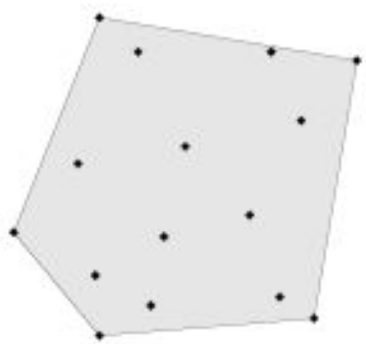
$$\begin{aligned} &\lambda_1 X_1 + \ldots + \lambda_k X_k \\ &\equiv \{x = \lambda_1 x_1 + \ldots + \lambda_k x_k : x_i \in X_i, 1 \leq i \leq k\} \end{aligned}$$

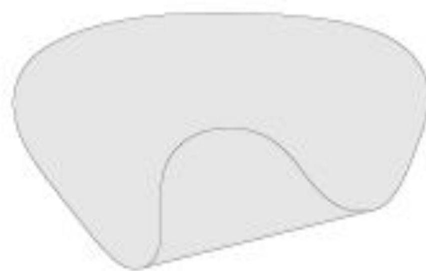
is convex. The proof also follows directly from the definition of convex sets.

Examples of Convex sets

Let $x_1, x_2, \ldots, x_m \in \mathbb{R}^n$. Their **convex hull** is

$$\mathbf{conv}\, C = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_i \in C,\ \theta_i \geq 0,\ i = 1, \ldots, k,\ \theta_1 + \cdots + \theta_k = 1\}.$$





Convex hull is always convex (by definition).
It is the smallest convex set that contains the set C, i.e., If B is any convex set that contains C, then conv C ⊆ B.

Images of Convex sets

1. The image of a convex set under affine mapping is convex

If $C \subset \mathbb{R}^n$ is convex and $\mathcal{A}(x) = \mathbf{A}x+\mathbf{b}$ is an affine mapping from $\mathbb{R}^n$ into $\mathbb{R}^m$ ($\mathbf{A}$ is m x n matrix, $\mathbf{b}$ is m-dimensional vector), then the set

$$\mathcal{A}(C) = \{ y \mid y = \mathbf{A}x + \mathbf{b},\ x \in C\} \text{ is convex in } \mathbb{R}^m$$

2. The inverse image of a convex set under affine mapping is convex

If $C \subset \mathbb{R}^n$ is convex and $\mathcal{A}(y) = \mathbf{A}y+\mathbf{b}$ is an affine mapping from $\mathbb{R}^m$ to $\mathbb{R}^n$ ($\mathbf{A}$ is n x m matrix, $\mathbf{b}$ is n-dimensional vector), then the set

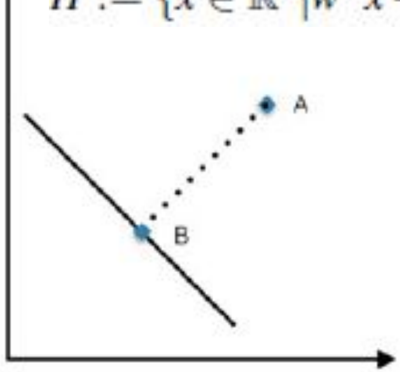
$$\mathcal{A}^{-1}(C) = \{ y \in \mathbb{R}^m : \mathcal{A}(y) \in C\} \text{ is convex in } \mathbb{R}^m$$

Projections onto Convex sets

Definition: Let X $\subseteq \mathbb{R}^n$ be a closed convex set, for any **y** $\in \mathbb{R}^n$ we define the closest point to **y** in **X** as

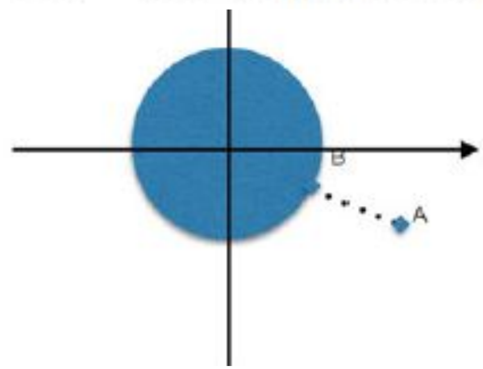
$$\mathrm{Proj}_X(y) = \operatorname*{argmin}_{x \in X} \|y - x\|_2^2.$$

$H := \{x \in \mathbb{R}^n \mid w^T x + b = 0\}$



$$\mathrm{Proj}_H(y) = y - \frac{(w^T y + b)w}{\|w\|_2^2}.$$

$B := \{x \in \mathbb{R}^n \mid \|x\|_2 \leq 1\}$



$$\mathrm{Proj}_B(y) = \frac{y}{\|y\|_2}.$$

Projections onto Convex sets

Definition: Let X $\subseteq \mathbb{R}^n$ be a closed convex set, for any **y** $\in \mathbb{R}^n$ (**y** $\notin$ **X**) we define the closest point to **y** in **X** as

$$\mathrm{Proj}_X(y) = \operatorname*{argmin}_{x \in X} \|y - x\|_2^2.$$

Proposition 1: The projection point is **unique**

Proof. Let a and b be the two closet points in X to the given point y, so that $\|y-a\|_2 = \|y-b\|_2 = d$. Since X is convex, the point $z = (a+b)/2 \in X$. Therefore $\|y-z\|_2 \geq d$. We now have

$$\underbrace{\|(y-a)+(y-b)\|_2^2}_{=\|2(y-z)\|_2^2 \geq 4d^2} + \underbrace{\|(y-a)-(y-b)\|_2^2}_{=\|a-b\|^2} = \underbrace{2\|y-a\|_2^2 + 2\|y-b\|_2^2}_{4d^2},$$

whence $\|a-b\|_2 = 0$. Thus, the closest to y point in X is unique. ∎

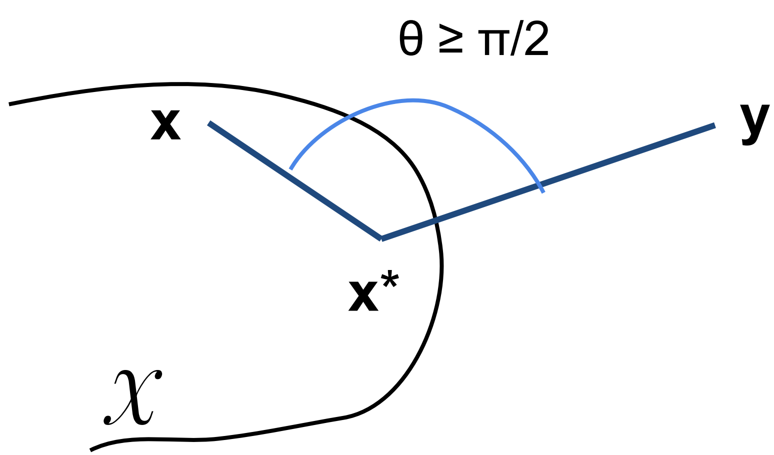
Projections onto Convex sets

Definition: Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a closed convex set, for any **y** $\in \mathbb{R}^n$ (**y** $\notin \mathcal{X}$) we define the closest point x* in $\mathcal{X}$ to **y** as

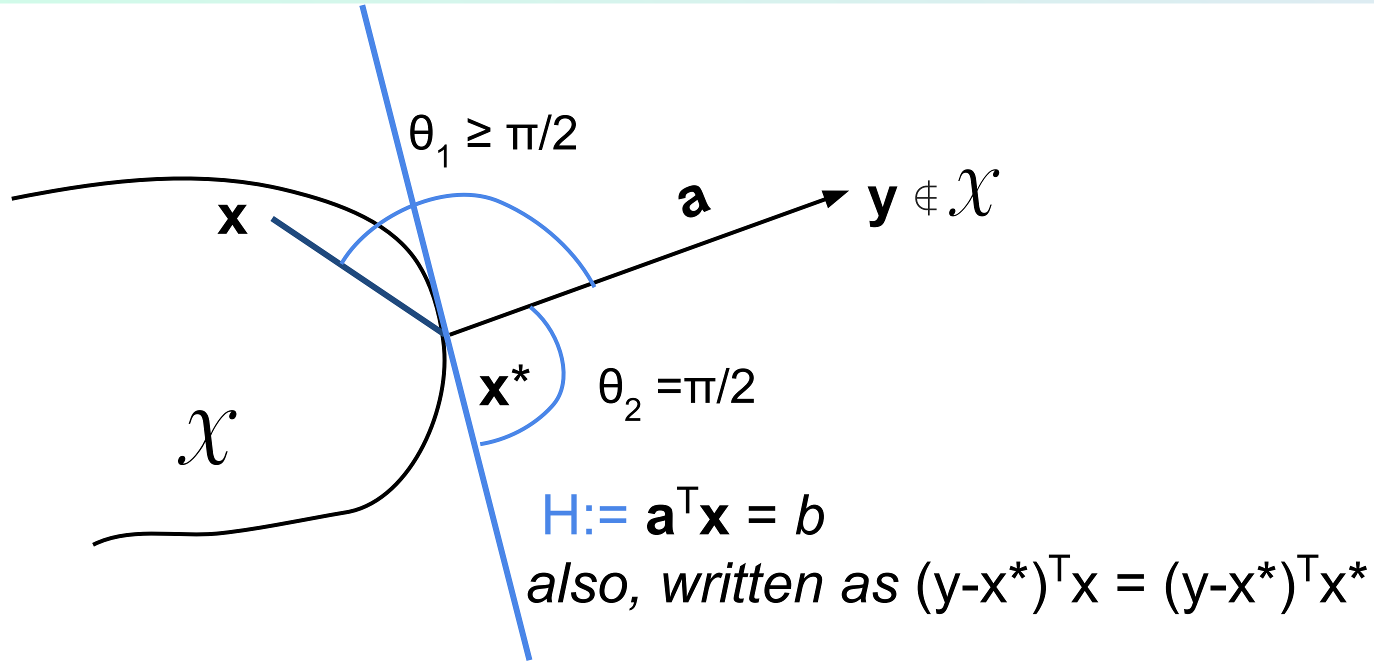
$$x^* = \mathrm{Proj}_X(y) = \operatorname*{argmin}_{x \in X} \|y - x\|_2^2.$$

Proposition 2: The unique projection point x* satisfies

$$(y-x^*)^T(x-x^*) \leq 0, \text{ for all } x \in \mathcal{X}$$



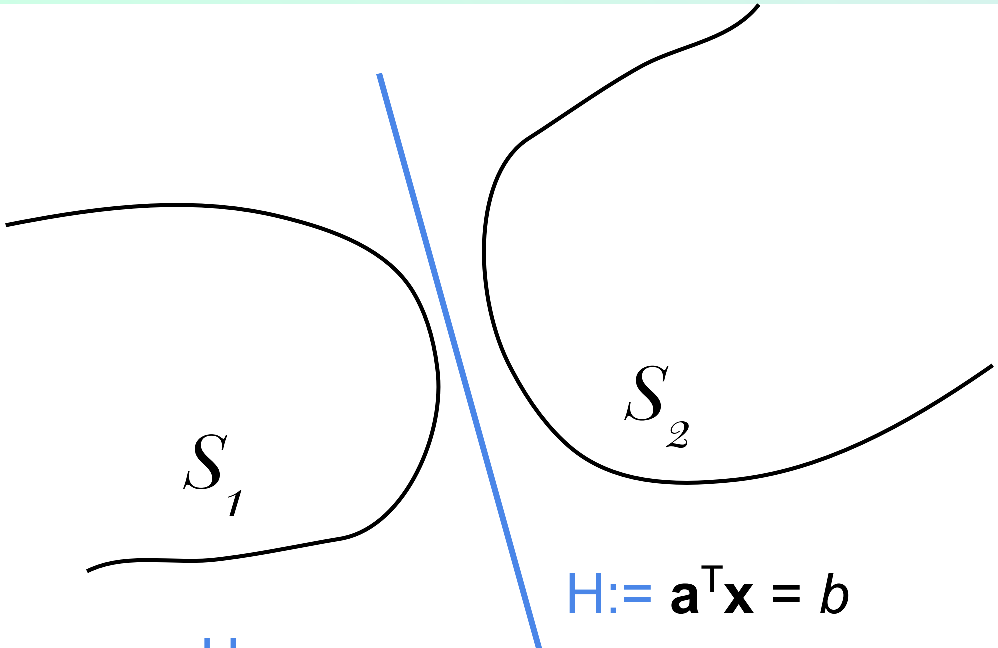
Supporting Hyperplane



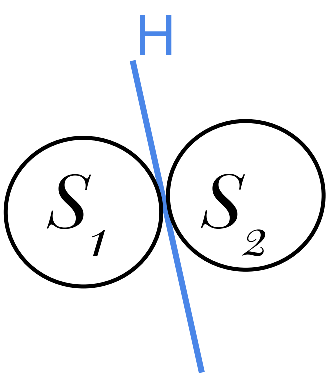
Proposition: Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a set, $\mathcal{X} \neq \emptyset$ (null set), and consider any boundary point $\mathbf{x}^*$. A hyperplane H:= $\mathbf{a}^T\mathbf{x} = b$ is a supporting hyperplane at the point $\mathbf{x}^*$ if $\mathbf{a}^T(\mathbf{x} - \mathbf{x}^*) \leq 0$, for all $x \in \mathcal{X}$

The supporting hyperplane $\mathbf{a}^T(\mathbf{x} - \mathbf{x}^*) = b$, also written as, $(\mathbf{y} - \mathbf{x}^*)^T\mathbf{x} = (\mathbf{y} - \mathbf{x}^*)^T\mathbf{x}^*$ is the tangent plane of the set $\mathcal{X}$ at the point $\mathbf{x}^*$

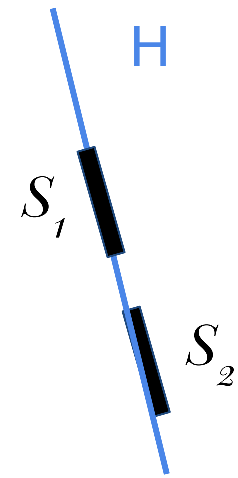
Separating Hyperplane



H= $\mathbf{a}^T\mathbf{x} = b$ is a separating hyperplane of the sets $\mathcal{S}_1$ and $\mathcal{S}_2$ if $\mathbf{a}^T\mathbf{x} \le b$ for x $\in \mathcal{S}_1$ and $\mathbf{a}^T\mathbf{x} \ge b$ for x $\in \mathcal{S}_2$ or vice versa

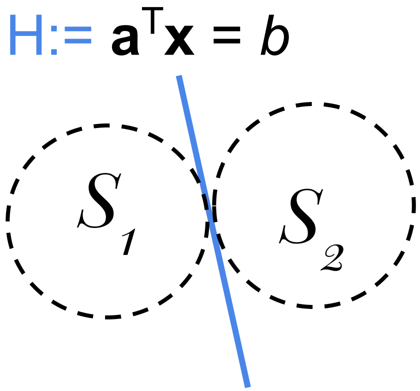


H is a proper separation since, $\mathcal{S}_1 \cup \mathcal{S}_2 \nsubseteq$ H



H is a not proper separation since, $\mathcal{S}_1 \cup \mathcal{S}_2 \subseteq$ H

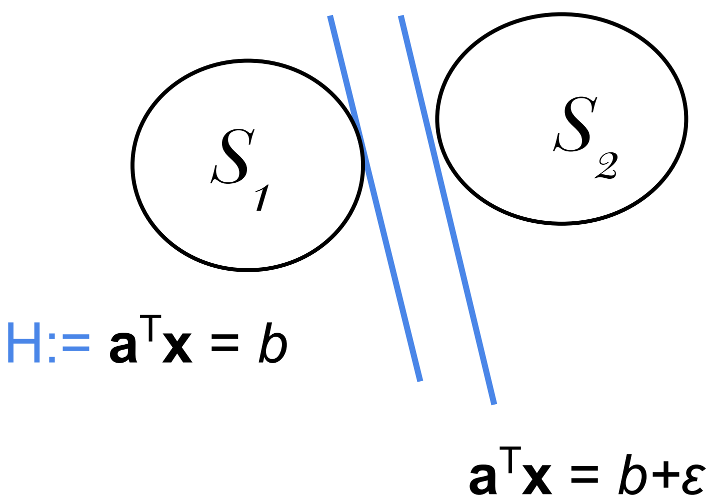
Strict and Strong Separation



H strictly separates since,

$\mathbf{a}^T\mathbf{x} < b$ for $\mathbf{x} \in S_1$ and

$\mathbf{a}^T\mathbf{x} > b$ for $\mathbf{x} \in S_2$



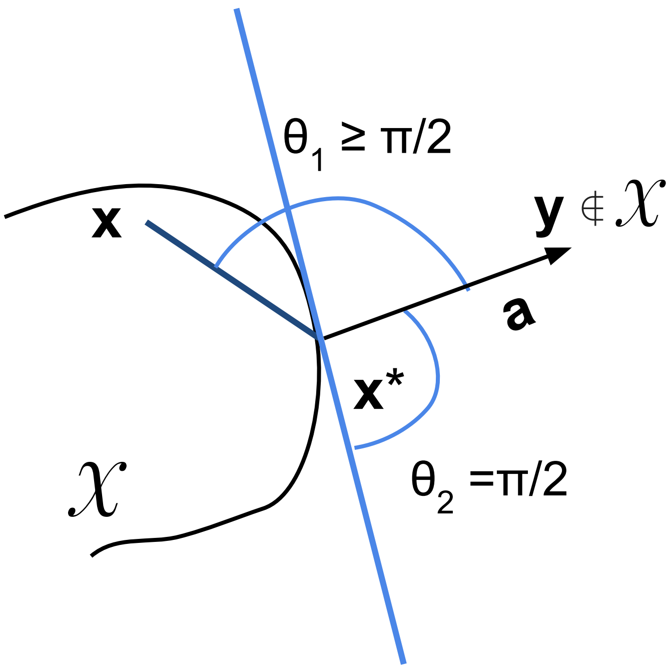
H strongly separates since,

$\mathbf{a}^T\mathbf{x} \leq b$ for $\mathbf{x} \in S_1$ and

$\mathbf{a}^T\mathbf{x} \geq b+\varepsilon$ for $\mathbf{x} \in S_2$,for some $\varepsilon > 0$

Strongly separating Hyperplane

Proposition: Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a closed convex set, $\mathcal{X} \neq \emptyset$ (null set), and consider any point $\mathbf{y} \notin \mathcal{X}$. Then there exists a hyperplane that strongly separates $\mathcal{X}$ and $\mathbf{y}$



H:= $\mathbf{a}^T\mathbf{x} = b$

also, written as

$(y-x^*)^T x = (y-x^*)^T x^*$

Proof: Let the projection from the given point $\mathbf{y} \notin \mathcal{X}$ to the set $\mathcal{X}$ be the point $\mathbf{x}^*$ which is unique and satisfies

$$(\mathbf{y} - \mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) \leq 0, \ \forall \ \mathbf{x} \in \mathcal{X}$$

Let $\mathbf{a} = \mathbf{y} - \mathbf{x}^*$ and $\mathbf{a}^T\mathbf{x}^* = b$, then we have

$$\mathbf{a}^T(\mathbf{x} - \mathbf{x}^*) \leq 0 \Rightarrow \mathbf{a}^T\mathbf{x} \leq \ b, \ \forall \ \mathbf{x} \in \mathcal{X}$$

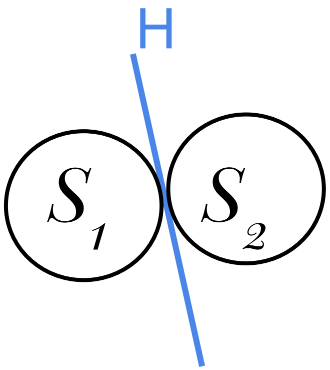
To show strong separation we need to show

$$\mathbf{a}^T\mathbf{y} \geq b + \varepsilon, \text{ for some } \varepsilon > 0$$

Note that $\mathbf{a}^T\mathbf{y} - b \ = \mathbf{a}^T\mathbf{y} - \mathbf{a}^T\mathbf{x}^*$

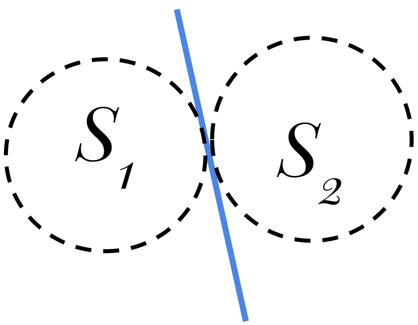
$$= \mathbf{a}^T(\mathbf{y} - \mathbf{x}^*)$$
$$= (\mathbf{y} - \mathbf{x}^*)^T(\mathbf{y} - \mathbf{x}^*)$$
$$= \|\mathbf{y} - \mathbf{x}^*\|^2 \ \geq 0 > \varepsilon, \text{for some } \varepsilon$$

What conditions are needed for separation?



H is a separating hyperplane of the sets S_1 and S_2 if $\mathbf{a}^T\mathbf{x} \leq b$ for x $\in S_1$ and $\mathbf{a}^T\mathbf{x} \geq b$ for x $\in S_2$ or vice versa

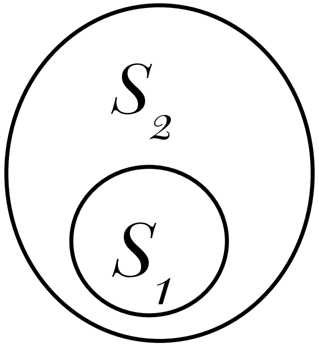
H:= $\mathbf{a}^T\mathbf{x} = b$



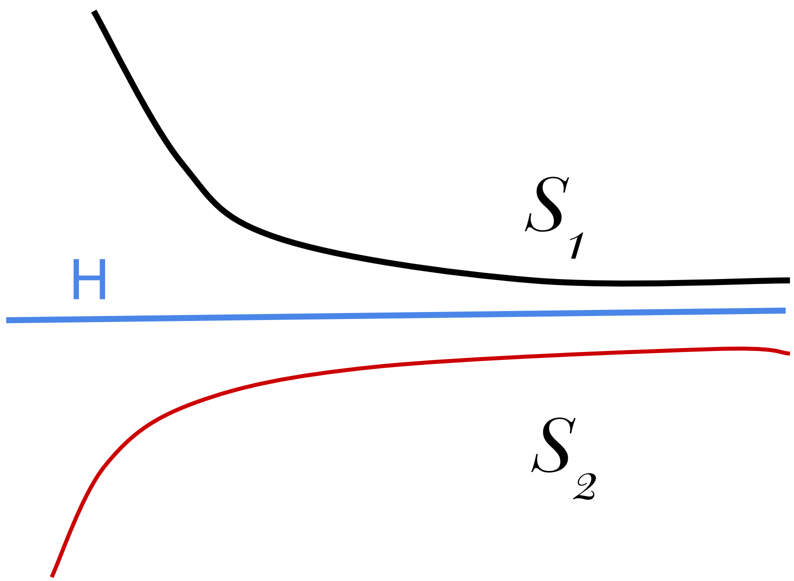
H strictly separates since, $\mathbf{a}^T\mathbf{x} < b$ for $\mathbf{x} \in S_1$ and $\mathbf{a}^T\mathbf{x} > b$ for $\mathbf{x} \in S_2$

In both cases $\mathbf{int}(S_1) \cap \mathbf{int}(S_2) = \varnothing$

What conditions are needed for strong separation?



The boundaries do not intersect,i.e., $\partial S_1 \cap \partial S_2 = \varnothing$, but there is no separating hyperplane



It is sufficient that the closures have no intersection: $\mathbf{cl}(S_1) \cap \mathbf{cl}(S_2) = \varnothing$