# Geometric Derivation of SVMs

Sreyas Mohan and Brett Bernstein

CDS at NYU

20 Feb 2019

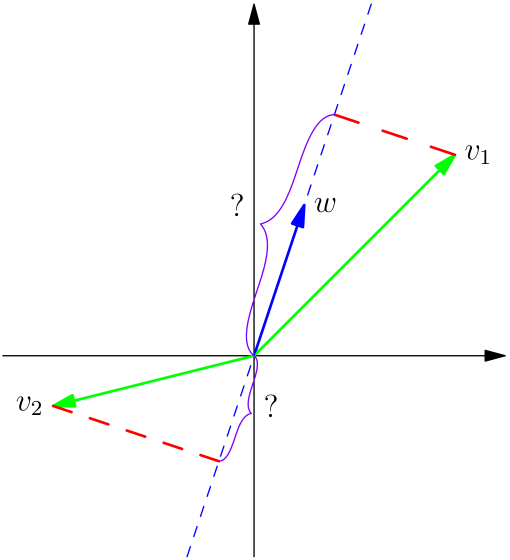
# Intro Question

## Question

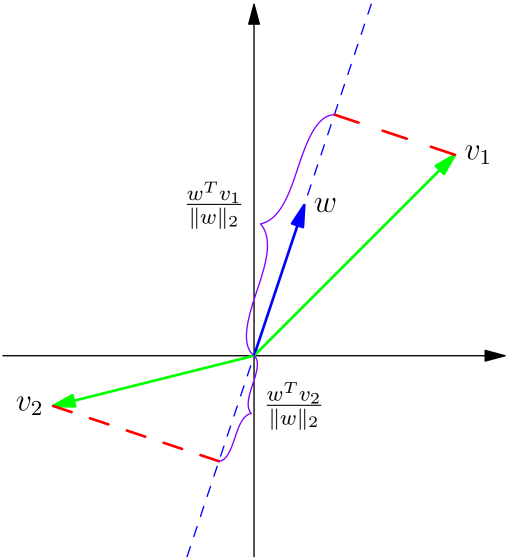
For any $w \in \mathbb{R}^d$ and $a \in \mathbb{R}$, $w^T x - b = 0$ represents the equation of a hyperplane in $\mathbb{R}^d$. A hyperplane divides the space $\mathbb{R}^d$ into two parts. We are given two points $x_1, x_2 \in \mathbb{R}^d$.

1. For $d = 2$, consider $w = (2,3)^T$ and $b = 6$. Does $x_1 = (0,0)^T$ and $x_2 = (4,3)^T$ fall on the same side of the hyper plane?
2. How can we write a computer program to determine this for a generic $w, a, x_i$ and $d$?

# Component of $v_1, v_2$ in the direction $w$



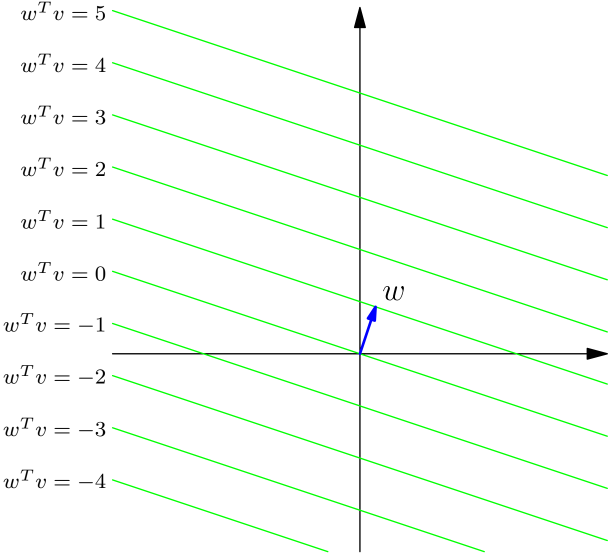
# Component of $v_1, v_2$ in the direction $w$



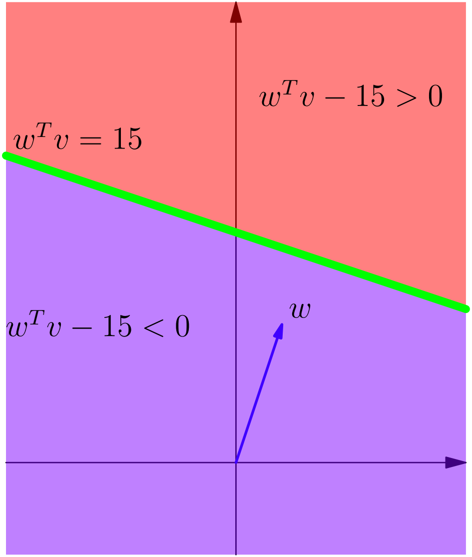
# $w^T x = b$

1. $S = \{x \in \mathbb{R}^d \mid w^T x = b\}$. What does this look like?
2. Note that $w^T x = b \iff \frac{w^T x}{\|w\|_2} = \frac{b}{\|w\|_2}$

# Level Surfaces of $f(v) = w^T v$ with $\|w\|_2 = 1$



# Sides of the Hyperplane $w^T v = 15$



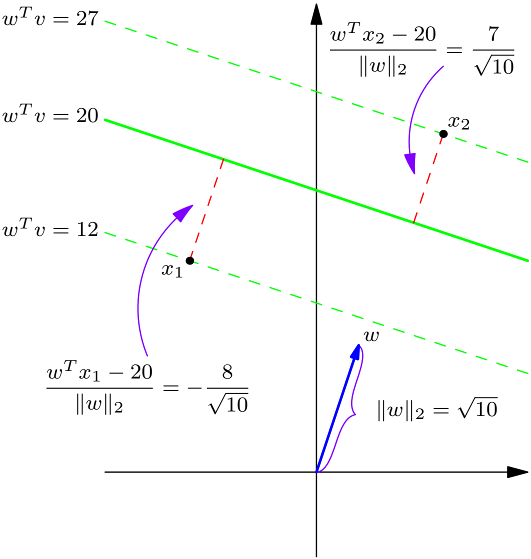
# Signed Distance from $x$ to Hyperplane $w^T v = b$

- "Signed" distance? Isn't distance always non-negative?
- If we have a vector $x \in \mathbb{R}^d$ and a hyperplane $H = \{v \mid w^T v = b\}$ we can measure the distance from $x$ to $H$ by

$$d(x, H) = \left| \frac{w^T x - b}{\|w\|_2} \right|.$$

- Without the absolute values we get the *signed distance*: a positive distance if $w^T x > b$ and a negative distance if $w^T x < b$.

# Signed Distance from $x_1, x_2$ to Hyperplane $w^T v = 20$



# Question

## Question

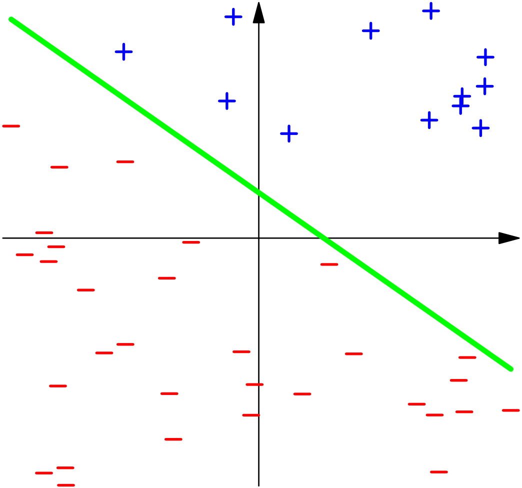
You have been given a data set $(x_i, y_i)$ for $i = 1, \ldots, n$ where $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$. You are also given a hyper plane parameterized by $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$. Give an algorithm to check if $w^T x = b$ separates the data points correctly or in other words, if $y = +1$ and $y = -1$ fall on two different sides of the hyperplane.

# Linearly Separable

## Definition

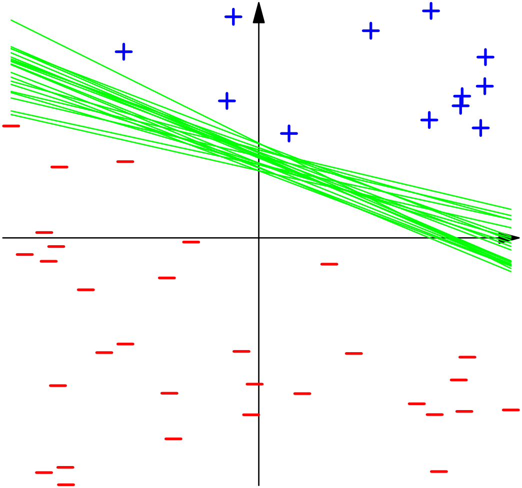
We say $(x_i, y_i)$ for $i = 1, \ldots, n$ are *linearly separable* if there is a $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$ such that $y_i(w^T x_i - b) > 0$ for all $i$. The set $\{v \in \mathbb{R}^d \mid w^T v - b = 0\}$ is called a *separating hyperplane*.

# Linearly Separable Data



How many separating hyperplanes?

# Many Separating Hyperplanes Exist



How do we pick one?

# Geometric Margin

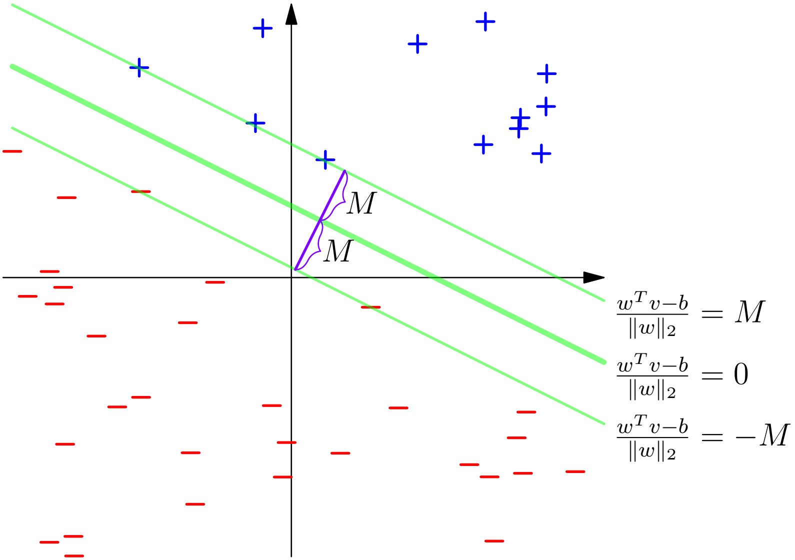
## Definition

Let $H$ be a hyperplane that separates the data $(x_i, y_i)$ for $i = 1, \ldots, n$. The geometric margin of this hyperplane is

$$\min_i d(x_i, H),$$

the distance from the hyperplane to the closest data point.

# Maximum Margin Separating Hyperplane



# Maximizing margin

We want to:

$$\text{maximize} \min_i d(x_i, H)$$

Remember:

$$d(x_i, H) = \left| \frac{w^T x_i - b}{\|w\|_2} \right| = \frac{y_i(w^T x_i - b)}{\|w\|_2}.$$

So:

$$\text{maximize}_{w,b} \min_i \frac{y_i(w^T x_i - b)}{\|w\|_2}.$$

Note, if $M = \min_i \frac{y_i(w^T x_i - b)}{\|w\|_2}$, then $\frac{y_i(w^T x_i - b)}{\|w\|_2} \geq M$ for all $i$

# Maximizing margin

We can rewrite this in a more standard form:

$$\begin{array}{ll} \text{maximize}_{w,b,M} & M \\ \text{subject to} & \frac{y_i(w^T x_i - b)}{\|w\|_2} \geq M \quad \text{for all } i. \end{array}$$

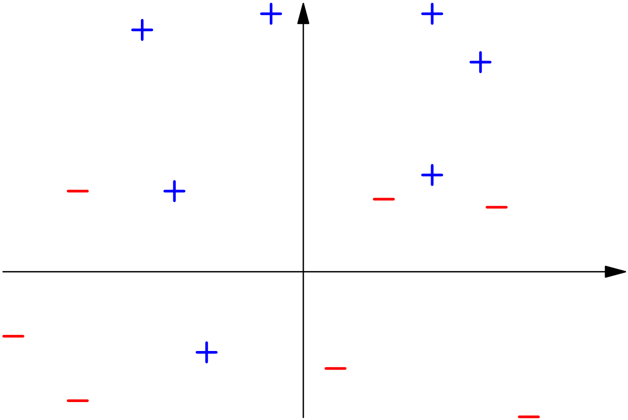
fix $\|w\|_2 = 1/M$ to obtain

$$\begin{array}{ll} \text{maximize}_{w,b} & 1/\|w\|_2 \\ \text{subject to} & y_i(w^T x_i - b) \geq 1 \quad \text{for all } i. \end{array}$$

To find the optimal $w, a$ we can instead solve the minimization problem

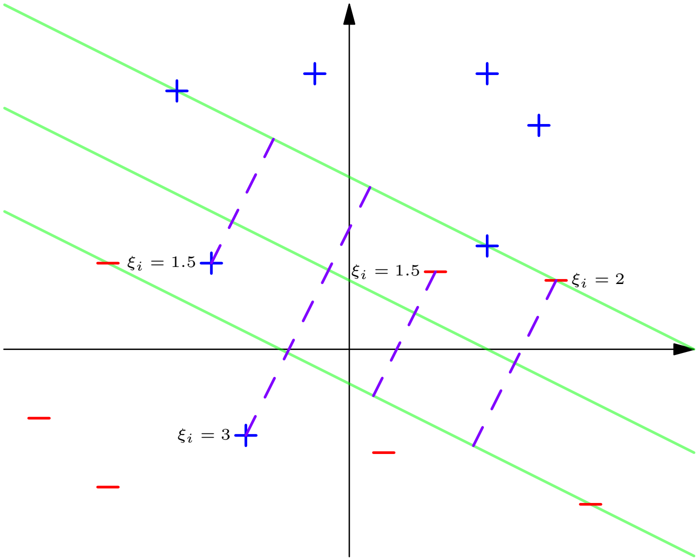
$$\begin{array}{ll} \text{minimize}_{w,b} & \|w\|_2^2 \\ \text{subject to} & y_i(w^T x_i - b) \geq 1 \quad \text{for all } i. \end{array}$$

# Linearly Non-Separable




What should we do if the data isn't linearly separable?

# Soft Margin SVM (unlabeled points have $\xi_i = 0$)



# Soft Margin SVM (introduce slack)

## Questions

$$
\begin{array}{ll}
\text{minimize}_{w,b,\xi} & \|w\|_2^2 + \frac{C}{n}\sum_{i=1}^n \xi_i \\
\text{subject to} & y_i(w^T x_i - b) \geq 1 - \xi_i \quad \text{for all } i \\
& \xi_i \geq 0 \quad \text{for all } i.
\end{array}
$$

1. What does $\xi_i > 0$ mean?
2. What does $C$ control?
3. Note that $\frac{y_i(w^T x_i + a)}{\|w\|_2} \geq \frac{1-\xi_i}{\|w\|_2}$. What does $\xi_i = 1$ and $\xi_i = 3$ mean?

## Question

Explain geometrically what the following optimization problem computes:

$$\begin{array}{ll} \text{minimize}_{w,a,\xi} & \frac{1}{n}\sum_{i=1}^{n}\xi_i \\ \text{subject to} & y_i(w^T x_i + a) \geq 1 - \xi_i \quad \text{for all } i \\ & \|w\|_2^2 \leq r^2 \\ & \xi_i \geq 0 \quad \text{for all } i. \end{array}$$

# Optimize Over Cases Where Margin Is At Least $1/r$