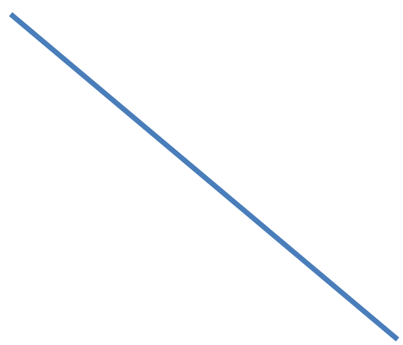
SVM Derivation

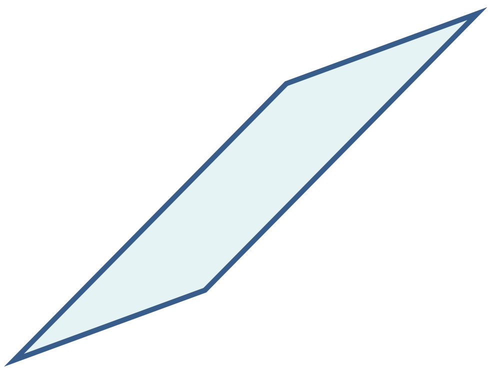
Vector representation

- Represent each sample as a vector
- The decision surface to separate classes is then a hyperplane

A decision surface in $\mathbb{R}^2$



A decision surface in $\mathbb{R}^3$



Vector Operations Review

- Scalar Multiplication, scalar c, vector $\vec{a} = (a_{1,} a_{2,} \dots, a_m)$

$$c\vec{a} = (ca_{1,} ca_{2,} \dots, ca_m)$$

- Addition/Subtraction of two vectors $\vec{a}, \vec{b}$

$$\vec{a} \pm \vec{b} = (a_1 \pm \overrightarrow{b_{1,}} a_2 \pm \overrightarrow{b_{2,}} \dots a_m \pm b_m)$$

- Euclidean length, L2-norm of a vector $\vec{a} = (a_{1,} a_{2,} \dots, a_m)$

$$\| \vec{a} \|_2 = \| \vec{a} \| = \sqrt{a_1^2 + a_2^2 + \cdots + a_m^2}$$

- Dot Product of two vectors $\vec{a}, \vec{b}$

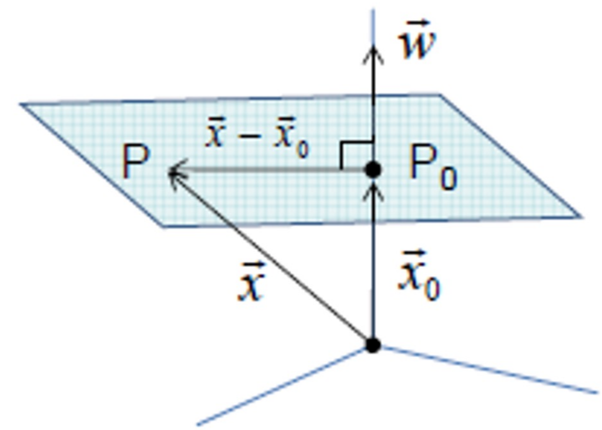
$$\vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2 + \cdots + a_m b_m = \sum_{i=1}^{m} a_i b_i$$

$$Note, \vec{a} \cdot \vec{a} = \| \vec{a} \|_2^2$$

Equation of Hyperplane

- An equation of a hyperplane is defined by a point, P_0, and a vector perpendicular to the plane at that point $\vec{w}$
- Define: $\overrightarrow{x_0} = \overrightarrow{OP_0}, \vec{x} = \overrightarrow{OP}$ for an arbitrary point P
- For P to be on the plane, then the vector $\vec{x} - \overrightarrow{x_0}$ is perpendicular to $\vec{w}$

Consider the case of $\mathbb{R}^3$:



$$\vec{w} \cdot (\vec{x} - \overrightarrow{x_0}) = 0$$

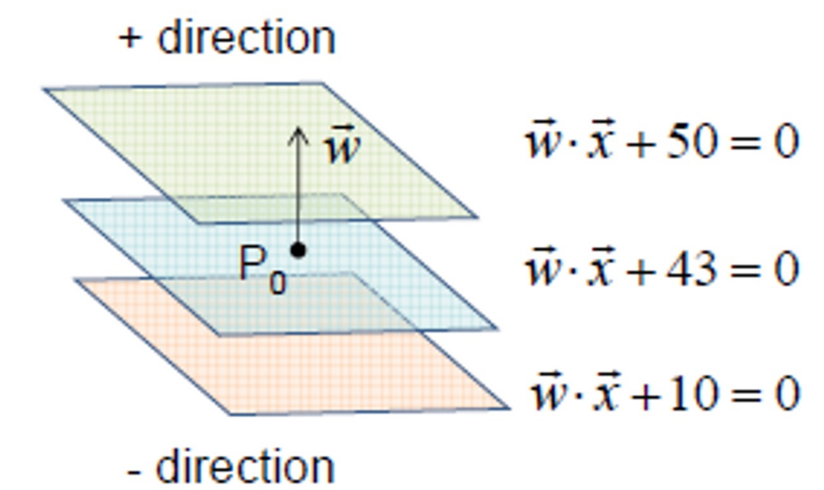
$$\vec{w} \cdot \vec{x} - \vec{w} \cdot \overrightarrow{x_0} = 0 \quad \text{define} \quad b = -\vec{w} \cdot \overrightarrow{x_0}$$

$$\boxed{\vec{w} \cdot \vec{x} + b = 0}$$

Equation of Hyperplane

- Recall, we are interested in maximum margin
- In 2D, find maximum margin as distance between parallel lines from decision boundary
- In general, looking at parallel hyperplanes

- Changing b coefficients get parallel hyperplanes



Distance between Hyperplanes

Distance between two parallel hyperplanes

$$\vec{w} \cdot \vec{x} + b_1 = 0, \qquad \vec{w} \cdot \vec{x} + b_2 = 0$$

$$D = \frac{|\,(b_1 - b_2)\,|}{\| \vec{w} \|}$$

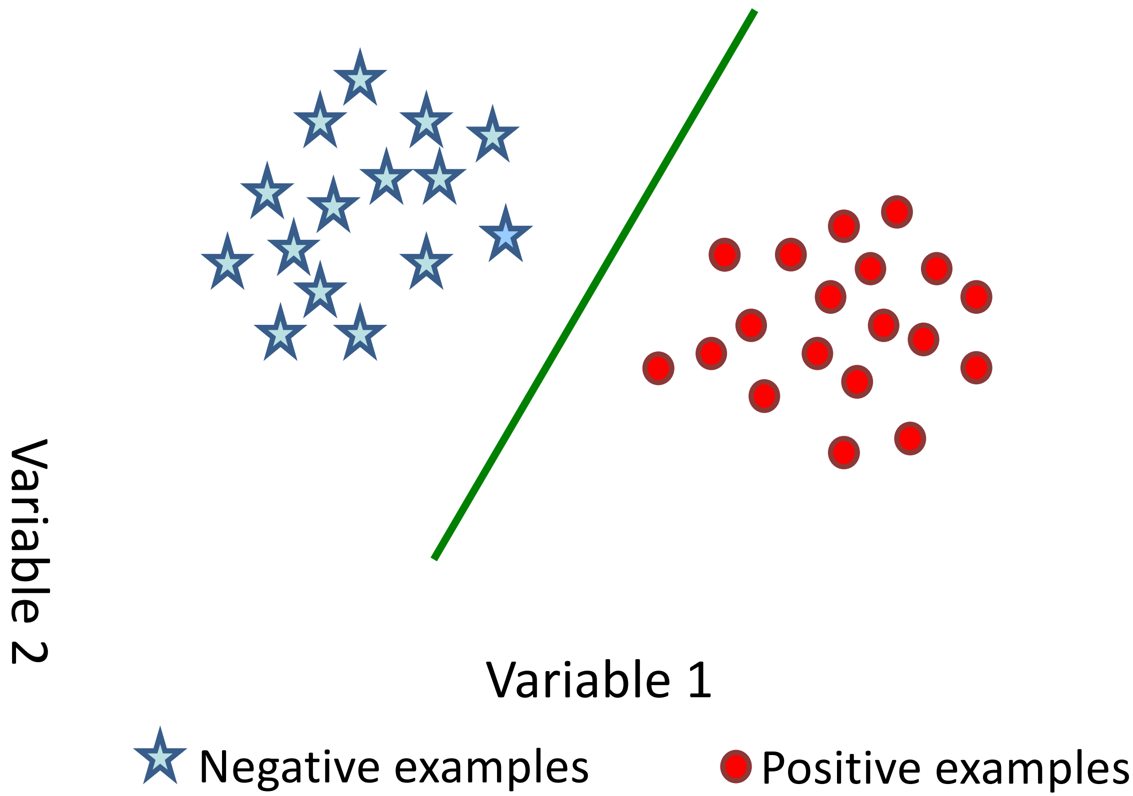
Back to Problem: Find $\vec{w}$ to maximize the margin

Optimization: Quadratic Programming (QP)

- Quadratic programming (QP) is a special optimization problem:
  - The function to optimize ("objective") is quadratic, subject to linear constraints
- The problems are solved by efficient greedy algorithms (for convex problems)
- Example:

Minimize $\underbrace{\frac{1}{2} \| \vec{x} \|_2^2}_{\textit{quadratic objective}}$ subject to $\underbrace{x_1 + x_2 - 1 \geq 0}_{\textit{linear constraints}}$

Case 1: Linearly separable data, “Hard-margin” linear SVM

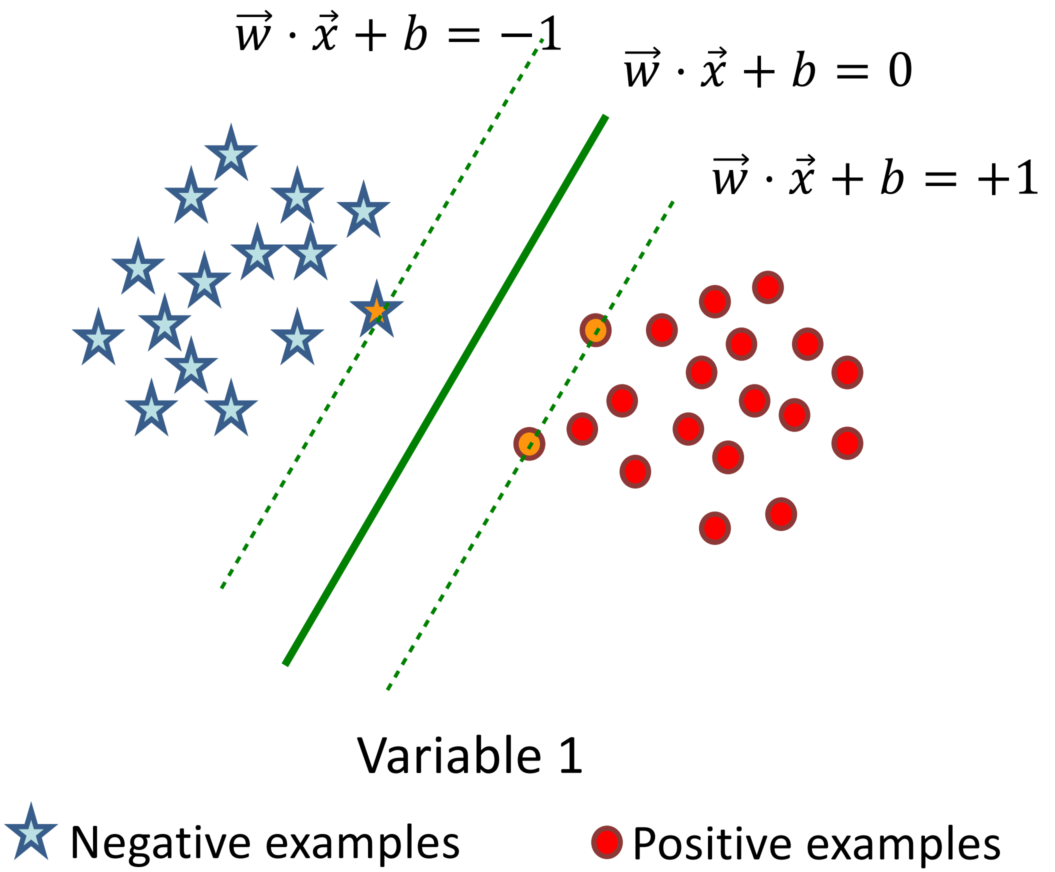


Training Data $\overrightarrow{x_1}, \overrightarrow{x_2}, \ldots \overrightarrow{x_N} \in \mathbb{R}^m$

$y_1, y_2, \ldots, y_N \in \{-1, +1\}$

- Want to find “best” classifier (hyperplane) to separate classes
  - Infinite such hyperplanes exist
- SVMs find hyperplane that maximizes gap between data samples on the boundaries

Linear SVM Classifier



The gap is distance between parallel hyperplanes:

$\vec{w} \cdot \vec{x} + b = -1$ and

$\vec{w} \cdot \vec{x} + b = +1$

Equivalently,

$$\vec{w} \cdot \vec{x} + (b + 1) = 0$$

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

We know:

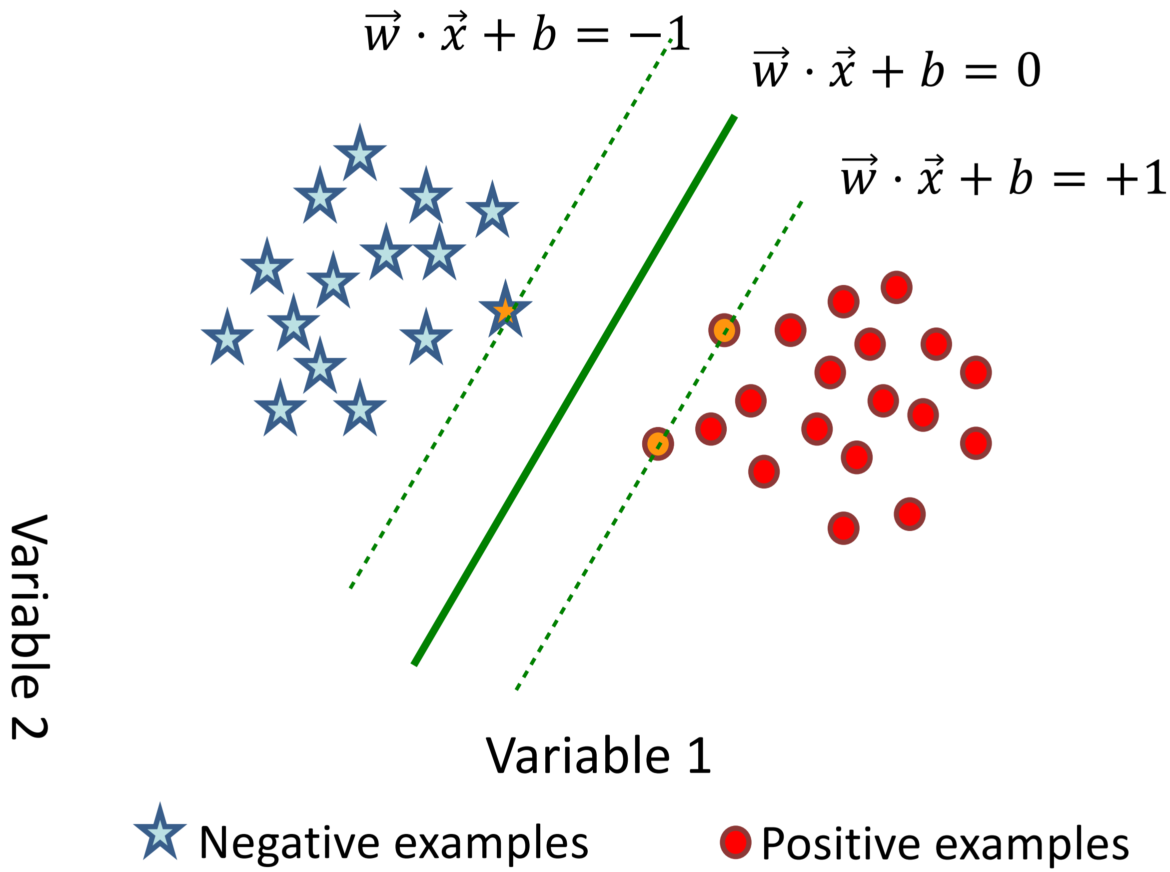
$$D = |\,(b_1 - b_2)\,| / \| \vec{w} \|$$

Therefore: $\boxed{D = 2 / \| \vec{w} \|}$

To maximize gap,
Minimize $\| \vec{w} \|$ equivalently $\boxed{\frac{1}{2} \| \vec{w} \|^2}$

Quadratic objective

Linear SVM Classifier



Add constraints, so that the samples are correctly classified

$$\vec{w} \cdot \vec{x} + b \leq -1 \;\; if \;\; y_i = -1$$

$$\vec{w} \cdot \vec{x} + b \geq +1 \;\; if \;\; y_i = +1$$

Equivalently,

$$\boxed{y_i(\vec{w} \cdot \vec{x} + b) \geq 1}$$

Summary:

Minimize $\frac{1}{2} \| \vec{w} \|^2$ subject to $y_i(\vec{w} \cdot \vec{x} + b) \geq 1$ for *i*=1,...,N

Classifier is: $f(\vec{x}) = sign(\vec{w} \cdot \vec{x} + b)$

SVM optimization problem: Primal formulation

Minimize $\frac{1}{2}\sum_{j=1}^{m} w_i^2$ subject to $y_i(\vec{w} \cdot \vec{x_i} + b) - 1 \geq 0$ for $i = 1,\ldots,N$

Objective function Constraints

- Primal formulation of linear SVM
- A convex quadratic programming optimization problem with *m* variables

(Derivation of dual formulation)

Minimize $\boxed{\frac{1}{2}\sum_{i=1}^{n} w_i^2}$ subject to $\boxed{y_i(\vec{w}\cdot\vec{x}_i + b) - 1 \geq 0}$ for $i = 1,\ldots,N$

Objective function

Constraints

Apply the method of Lagrange multipliers.

Define Lagrangian $\Lambda_P(\vec{w}, b, \vec{\alpha}) = \frac{1}{2}\sum_{i=1}^{n} w_i^2 - \sum_{i=1}^{N} \alpha_i \left(y_i(\vec{w}\cdot\vec{x}_i + b) - 1\right)$

a vector with *n* elements

a vector with *N* elements

We need to minimize this Lagrangian with respect to $\vec{w}, b$ and simultaneously require that the derivative with respect to $\vec{\alpha}$ vanishes , all subject to the constraints that $\alpha_i \geq 0.$

(Derivation of dual formulation)

If we set the derivatives with respect to $\vec{w}, b$ to 0, we obtain:

$$\frac{\partial \Lambda_P(\vec{w}, b, \vec{\alpha})}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\frac{\partial \Lambda_P(\vec{w}, b, \vec{\alpha})}{\partial \vec{w}} = 0 \Rightarrow \vec{w} = \sum_{i=1}^{N} \alpha_i y_i \vec{x}_i$$

We substitute the above into the equation for $\Lambda_P(\vec{w}, b, \vec{\alpha})$ and obtain "*dual formulation of linear SVMs*":

$$\Lambda_D(\vec{\alpha}) = \sum_{i=1}^{N} \alpha_i - \tfrac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j$$

We seek to maximize the above Lagrangian with respect to $\vec{\alpha}$, subject to the constraints that $\alpha_i \geq 0$ and $\sum_{i=1}^{N} \alpha_i y_i = 0$.

10

SVM optimization problem: Dual formulation

Recast problem to "dual form"

Also, a convex quadratic optimization problem with *N* variables, where *N* is the number of samples

Maximize $$\boxed{\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \vec{x_i} \cdot \vec{x_j}}$$ Objective function

subject to $$\boxed{\alpha_i \geq 0 \text{ and } \sum_{i=1}^{N} \alpha_i y_i = 0}$$ Constraints

Then the w-vector is defined in terms of α_i: $$\vec{w} = \sum_{i=1}^{N} \alpha_i y_i \vec{x_i}$$

And the solution becomes: $$f(\vec{x}) = sign\left(\sum_{i=1}^{N} \alpha_i y_i \vec{x_i} \cdot \vec{x_j} + b\right)$$