24 January 2024 02:20 PM

Consider the equation of the straight line

$y = -2x + 4$

$$x_2 = -2x_1 + 4.$$

$\Rightarrow 2x_1 + x_2 - 4 = 0$

$\begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} -4 \end{bmatrix} = 0$

The straight line can also be represented as a vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

$\boxed{2x_1 + x_2 = 4}$

Solve this equation

$\underbrace{\begin{bmatrix} 2 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \underbrace{4}_{= b}$

Write the vector $\mathbf{x}$ as follows:

$$\mathbf{x} = \boxed{?} \begin{bmatrix} \boxed{?} \\ \boxed{?} \end{bmatrix} + \begin{bmatrix} \boxed{?} \\ \boxed{?} \end{bmatrix}$$

Augmented matrix $\left[\begin{array}{cc|c} 2 & 1 & 4 \end{array}\right]$

Row operations $\longrightarrow$ $\begin{array}{ccc} \text{Pivot} & \text{free} & \\ x_1 & x_2 & \\ \end{array}$ $\left[\begin{array}{cc|c} 1 & 1/2 & 2 \end{array}\right]$

Solve for the pivot variable in terms of the free variable

$\Rightarrow x_1 + 1/2\, x_2 = 2$

$\Rightarrow \begin{bmatrix} x_1 = \\ x_2 = \end{bmatrix} \begin{bmatrix} -1/2\, x_2 + 2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1/2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

$$\Rightarrow \text{Solution vector } x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{x_2}_{\text{any real number}} \begin{bmatrix} -1/2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$x_2 = \overset{m}{-2}x_1 + \overset{c}{4} \iff \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} \overset{1/m}{-1/2} \\ 1 \end{bmatrix} + \begin{bmatrix} \overset{c/m}{2} \\ 0 \end{bmatrix}$$

$$\text{E.g } x_2 = 0 \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \text{E.g } x_2 = 1 \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 1 \end{bmatrix}$$

For a generic line $x_2 = mx_1 + c$

$$\Rightarrow -mx_1 + x_2 = c \Rightarrow \left[\begin{array}{cc|c} -m & 1 & c \end{array}\right]$$

$$\xrightarrow{\text{row ops}} \begin{array}{ccc} x_1 & x_2 & \\ \left[1 \right. & -1/m & \left. \middle| \; c/m \right] \end{array}$$

$$\begin{aligned} x_1 - \frac{1}{m}x_2 &= \frac{c}{m} \\ x_2 &= x_2 \end{aligned} \Rightarrow \begin{aligned} x_1 &= \frac{1}{m}x_2 + \frac{c}{m} \\ x_2 &= x_2 \end{aligned}$$

$$\Rightarrow x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 1/m \\ 1 \end{bmatrix} + \begin{bmatrix} c/m \\ 0 \end{bmatrix}$$

np.arange(-5, 5, 0.01)

$$\begin{bmatrix} -5.0 \\ -4.99 \\ -4.98 \\ \vdots \\ 4.98 \\ 4.99 \\ 5.0 \end{bmatrix} * \text{np.array}([1/m, 1]) + \text{np.array}([c/m, 0])$$

Suppose $\mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $b = -4$. Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$. Solve the equation $\boxed{\mathbf{w}^{\mathrm{T}}\mathbf{x} + b = 0}$ for the

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + (-4) = 0$$

unknown vector $\mathbf{x}$ and fill in the missing entries below:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \left\{ x_2 \begin{bmatrix} ? \\ ? \\ ? \end{bmatrix} + ? \begin{bmatrix} ? \\ ? \\ 1 \end{bmatrix} + \begin{bmatrix} ? \\ ? \\ ? \end{bmatrix} \middle| ?, ? \in \mathbb{R} \right\}.$$

$$w^T \quad x + b$$

$x_2 = -2x_1 + 4 \Rightarrow 2x_1 + x_2 - 4 = 0 \Rightarrow \underbrace{\begin{bmatrix} 2 & 1 \end{bmatrix}}_{\text{known}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 4 = 0$

What is the geometry of $w^T x + b = 0$?

$$x_1 + 2x_2 + 3x_3 - 4 = 0$$

$$\Rightarrow x_1 + 2x_2 + 3x_3 = 4$$

$\Rightarrow$ Augmented matrix $\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \end{array}\right]$

$$\xrightarrow{\text{Row ops}} \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & \end{array} \\ \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \end{array}\right] \end{array}$$

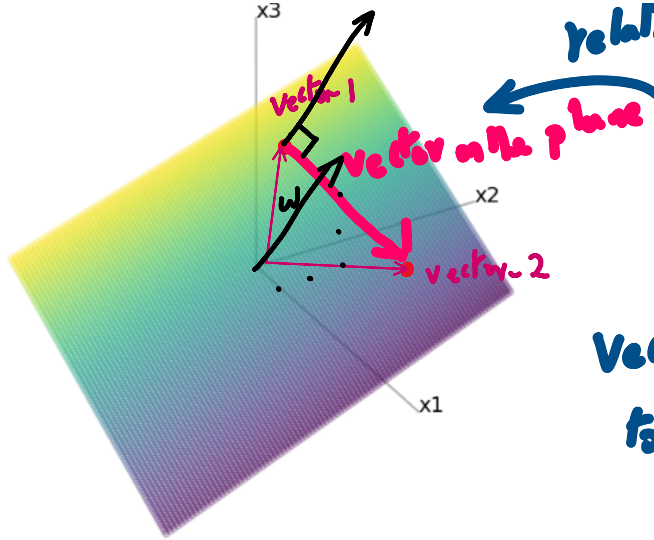
$$x_1 = -2x_2 - 3x_3 + 4$$
$$x_2 = x_2$$
$$x_3 = x_3$$

$$\Rightarrow x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}$$

free variables

w

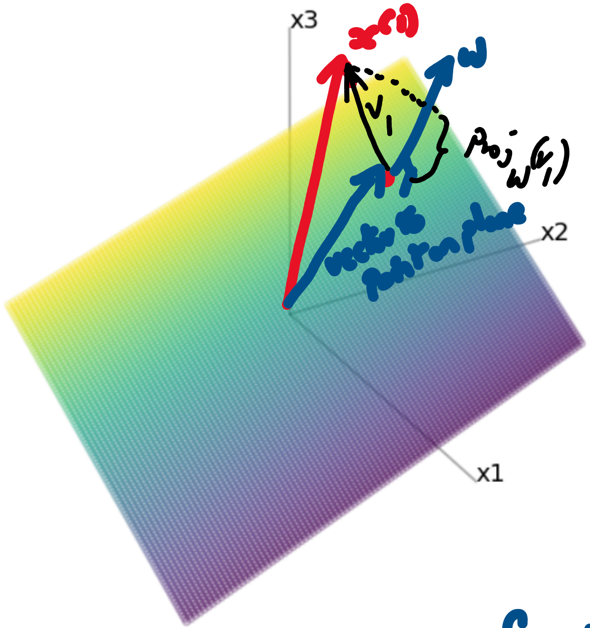
w to the solution



relationship of
of the eqn $w^T x + b = 0$

$$w = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Vector w is normal to the plane



$w^T x + b = 0$ represents a hyperplane with w as normal to the plane

Distance of the sample $x^{(1)}$ from the hyperplane

= Scalar Projection of v_1 onto w

Sample-1 seen from the plane

normal to plane

$$= \frac{v_1^T w}{\|w\|} = \frac{\left(x^{(1)} - \text{vector to point on plane}\right)^T w}{\|w\|}$$

$$= \frac{\left(x^{(1)T} - (\text{vector to point on plane})^T\right) w}{\|w\|}$$

$$= \frac{x^{(1)T} w - \boxed{(\text{vector to point on plane})^T w}}{\|w\|} \qquad w^T x = -b$$

$\Rightarrow$ Distance of $x^{(1)}$ to the hyperplane $w^T x + b = 0$

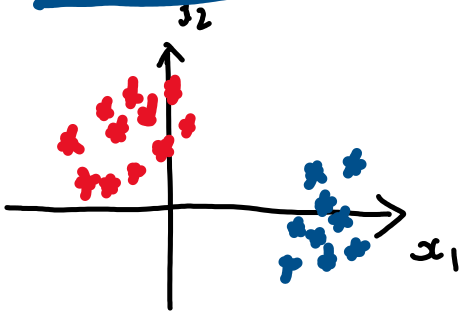
is equal to $\dfrac{w^T x^{(1)} - (-b)}{\|w\|} = \dfrac{w^T x^{(1)} + b}{\|w\|}$

Suppose we had n samples $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$

Distance of sample $x^{(i)}$ (ith sample) to the hyperplane

is equal to $\dfrac{w^T x^{(i)} + b}{\|w\|}$

- Hyperplane-based Classifiers



Find a hyperplane that maximally seperates the red (-1) and blue (+1) samples

Find w and b

We put in a condition which will simplify finding w and b.

$\Rightarrow$ Find w and b s.t. $w^T x^{(i)} + b \geq 1$ for positive samples

and $w^T x^{(i)} + b \leq -1$ for negative samples

If $y^{(i)}$ is the ith sample's label, $y^{(i)} = 1$ or -1

equivalent to $y^{(i)}(w^T x^{(i)} + b) \geq 1$ for all samples

Directed distance of sample $x^{(i)}$ to the hyperplane $w^T x + b = 0$ is

$$\frac{w^T x^{(i)} + b}{\|w\|}$$

100 Samples

$$\left\{ \frac{|w^T x^{(0)} + b|}{\|w\|}, \frac{|w^T x^{(1)} + b|}{\|w\|}, \cdots, \frac{|w^T x^{(98)} + b|}{\|w\|}, \frac{|w^T x^{(99)} + b|}{\|w\|} \right\}$$

Find w and b such that we maximize the minimum among $\frac{|w^T x^{(i)} + b|}{\|w\|}$.

s.t. $y^{(i)}(w^T x^{(i)} + b) \geq 1$

$$\text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geq 1$$

$$\Downarrow$$

Find w and b s.t. maximize $\frac{1}{\|w\|}$

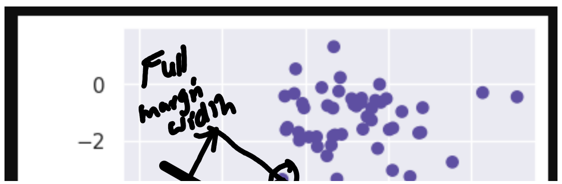
$$\text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geq 1$$

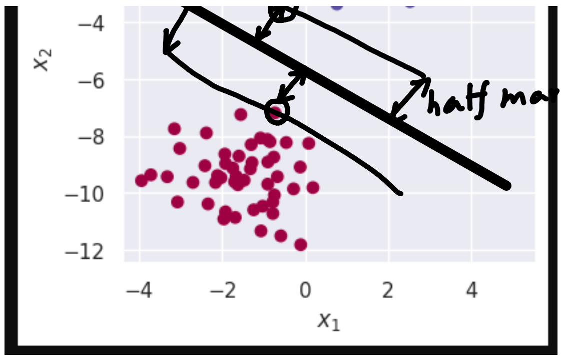
$$\Downarrow$$

Find w and b s.t. minimize $\frac{\|w\|^2}{2}$

$$\text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geq 1$$

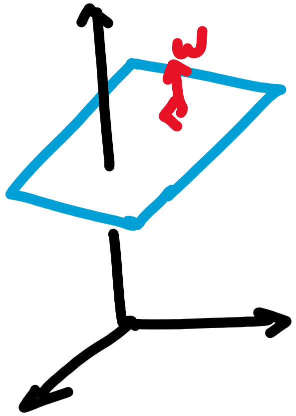
$\Rightarrow$ The hyperplane that comes out of solving this optimization problem is called the hard margin svm classifier.





half margin width $= \frac{1}{\|w\|}$

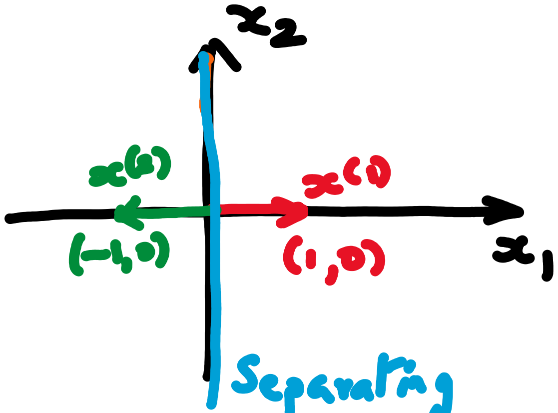
$$\boxed{w^T x + b = 0}$$



A very simple dataset with 2 samples

$$x^{(1)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad y^{(1)} = 1$$

$$x^{(2)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad y^{(2)} = -1$$



Separating hyperplane

Hard margin SVM: Find $\boxed{w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \text{ and } b}$

What is the maximal margin hyperplane

s.t. we $\begin{cases} \text{minimize} \quad \frac{\|w\|^2}{2} \\ \text{s.t. constraints} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1 \end{cases}$

s.t. the constraints $y\,(w x + b) \geq 1$

$\Rightarrow$ apply it for the dataset we have:

$$\begin{cases} \text{minimize } \left(\dfrac{\sqrt{w_1^2+w_2^2}}{2}\right)^2 \\ \text{s.t. the constraints } \quad 1\left([w_1\ w_2]\begin{bmatrix}1\\0\end{bmatrix}+b\right) \geq 1 \\ \qquad\qquad\qquad\qquad\quad -1\left([w_1\ w_2]\begin{bmatrix}-1\\0\end{bmatrix}+b\right) \geq 1 \end{cases}$$

$$\Rightarrow \begin{cases} \boxed{\text{minimize } (w_1^2+w_2^2)/2} \\ \text{s.t. } \boxed{\begin{matrix} w_1 + b \geq 1 \\ w_1 - b \geq 1 \end{matrix}} \end{cases} \Rightarrow$$

What are the optimal values of w_1, w_2, and b?

The maximal margin hyperplane is

$$\begin{cases} x_1 = 0 \\ \underset{\substack{\downarrow \\ 1}}{w_1} x_1 + \underset{\substack{\downarrow \\ 0}}{w_2} x_2 + \underset{\substack{\downarrow \\ 0}}{b} = 0 \end{cases}$$

$\Rightarrow$ I would

like this is

→ In general, a constrained minimization problem ...
solved using convex optimization algorithms

Primal Problem

minimize $\frac{w_1^2 + w_2^2}{2}$

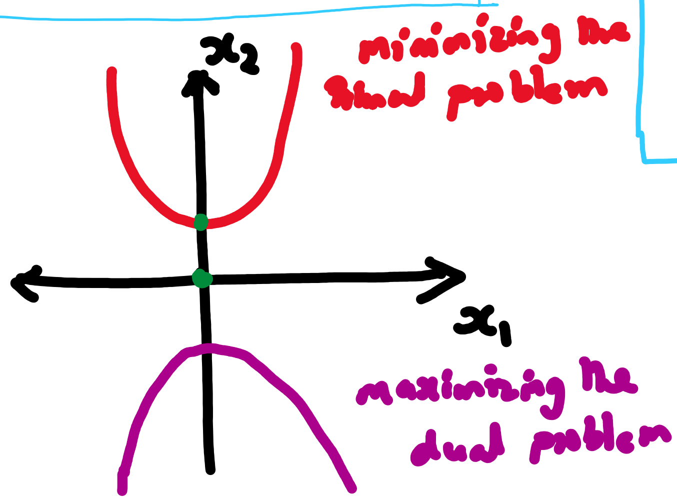
s.t. $w_1 + b \geq 1$
$w_1 - b \geq 1$

Kernel trick — dot product to measure similarity between samples

Dual problem

maximize $[\alpha_1\ \alpha_2]\begin{bmatrix}1\\1\end{bmatrix} - \frac{1}{2}[\alpha_1\ \alpha_2]\begin{bmatrix} y^{(1)}y^{(1)}x^{(1)T}x^{(1)} & y^{(1)}y^{(2)}\underline{x^{(1)T}x^{(2)}} \\ y^{(2)}y^{(1)}x^{(2)T}x^{(1)} & y^{(2)}y^{(2)}x^{(2)T}x^{(2)} \end{bmatrix}\begin{bmatrix}\alpha_1\\\alpha_2\end{bmatrix}$

s.t. $\alpha_1, \alpha_2 \geq 0$
$\alpha_1 y^{(1)} + \alpha_2 y^{(2)} = 0$



$\Downarrow$

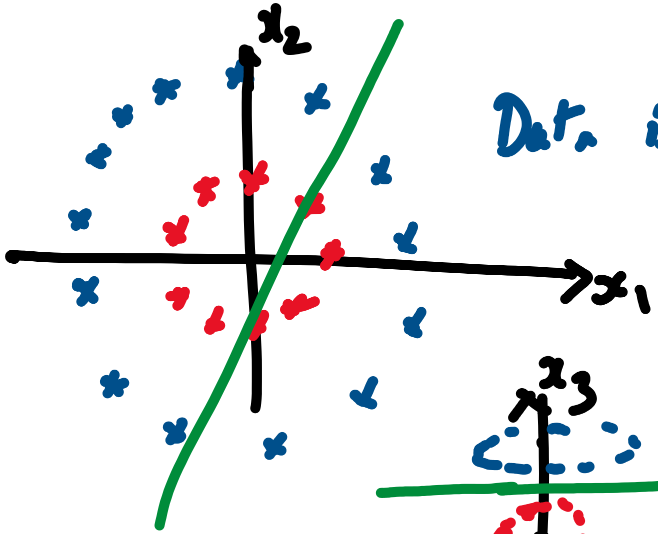
maximize $\alpha^T \mathbf{1} - \frac{1}{2}\alpha^T G \alpha$ $\qquad G_{ij} = y^{(i)}y^{(j)}x^{(i)T}x^{(j)}$

s.t. $\sum_{i=1}^{2} \alpha_i y^{(i)} = 0$

- Solve the dual problem for $\alpha = \begin{bmatrix}\alpha_1\\\alpha_2\end{bmatrix}$ and then calculate the solution for the primal problem as $w = \alpha_1 y^{(1)} x^{(1)}$

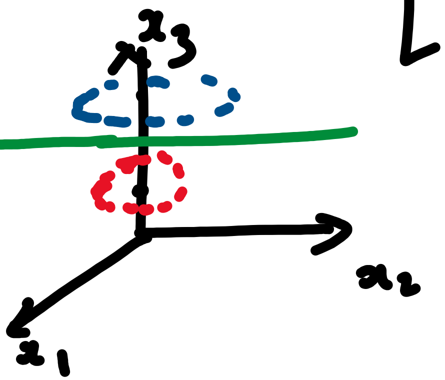
$+\alpha_2 y^{(2)} x^{(2)}$

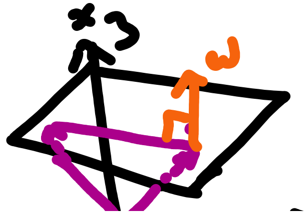
- In general for n samples, if we solve the dual problem to get $w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$, it will turn out that most of the α_i's are zeros (support vectors)



Data is not linearly separable

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow \begin{bmatrix} x_1 \\ x_2 \\ \sqrt{x_1^2 + x_2^2} \end{bmatrix}$$



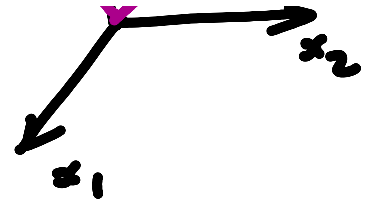


1. Suppose $\mathbf{w} = \begin{bmatrix}1\\2\end{bmatrix}$ and $b = -4$. Let $\mathbf{x} = \begin{bmatrix}x_1\\x_2\end{bmatrix}$. Solve the equation $\mathbf{w}^{\mathrm{T}}\mathbf{x} + b = 0$ for the

$\lfloor 3 \rfloor$ $\lfloor x_3 \rfloor$

$x_1 + 2x_2 + 3x_3 - 4 = 0$

$= 4$



unknown vector **x** and fill in the missing entries below:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \left\{ x_2 \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} \middle| \, ?, ? \in \mathbb{R} \right\}.$$

$$\begin{array}{cccc} x_1 & x_2 & x_3 & \\ \end{array}$$
$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \end{array}\right]$$

$\boxed{x_1 = 4 - 2x_2 - 3x_3}$ ✓

$\boxed{x_2 = x_2}$ ✓

$\boxed{x_3 = x_3}$ ✓

w is a <u>normal</u> to the hyperplane

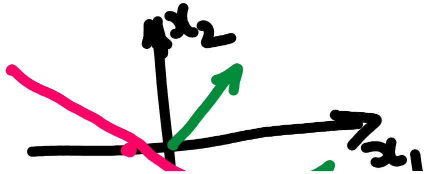
<u>E.g.</u> $\boxed{\overset{w_1}{4} x_1 + \overset{w_2}{3} x_2 + \overset{b}{1} = 0} \Rightarrow \boxed{x_2 = -\frac{4}{3} x_1 - \frac{1}{3}}$

The equation of the same line

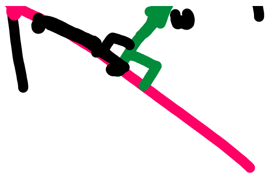
Equation of a line (hyperplane)

$w^T x + b = 0 \Rightarrow \begin{bmatrix} 4 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 1 = 0$

$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ is a normal to the hyperplane



Directed distance of a sample $x^{(i)}$ from the hyperplane $w^T x + b = 0$



is $\dfrac{w^T x^{(i)} + b}{\|w\|}$

$$\text{Margin} = \frac{|w^T x^{(i)} + b|}{\|w\|} \Rightarrow \text{hard-margin SVM}$$

- Samples $x^{(1)}, x^{(2)}, \dots x^{(n)}$ with labels $y^{(1)}, y^{(2)}, \dots, y^{(n)}$ (-1 or 1)
- Goal in hard-margin SVM is to maximize the worst margin among all samples subject to the constraints that $y^{(i)}(w^T x^{(i)} + b) \geq 1$

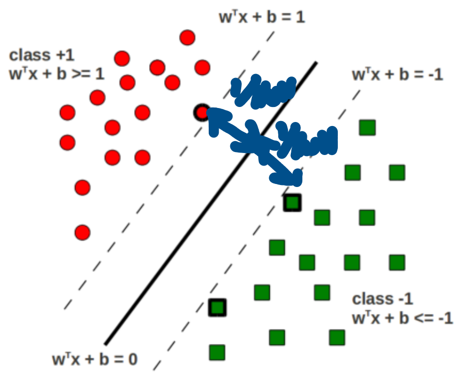
maximize the minimum of $\dfrac{|w^T x^{(i)} + b|}{\|w\|}$

s.t. $y^{(i)}(w^T x^{(i)} + b) \geq 1$

$$\boxed{\begin{array}{l} w^T x^{(i)} + b \geq 1 \text{ for } y^{(i)} = 1 \\ w^T x^{(i)} + b \leq -1 \text{ for } y^{(i)} = -1 \end{array}}$$

$$\boxed{\begin{array}{l} \text{maximize } \dfrac{1}{\|w\|} \\ \text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geq 1 \end{array}} \Longleftrightarrow \boxed{\begin{array}{l} \min \dfrac{1}{2}\|w\|^2 \\ \text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geq 1 \end{array}}$$

Regularization loss in DL



Is this dataset linearly separable? yes

Hard margin SVM will perfectly separate the classes if the dataset is linearly separable

Full margin width $= \frac{2}{\|w\|}$

Soft-margin SVM

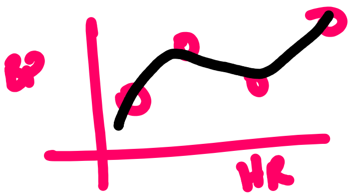
as wide a margin as possible

as much slack as possible

$$\min \boxed{\frac{1}{2}\|w\|^2} + C \boxed{\sum_{i=1}^{n} \xi^{(i)}}$$

s.t. $y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi^{(i)}$ → slack for the ith sample

$\xi^{(i)} \geq 0$



C is giving us a compromise between a wide margin (low training error) and good generalizability (low variance)

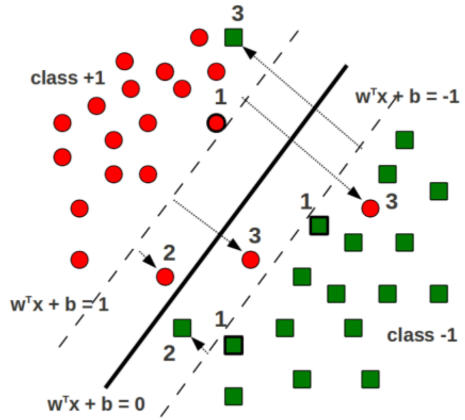
low bias

The soft-margin SVM solution has three types of support vectors

$\xi = 0$, 1 = right on the margin

$0 < \xi < 1$ 2 = on the correct side but within the margin

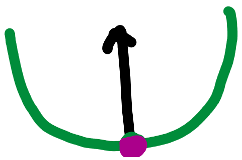
$\xi \geq 1$ 3 = on the incorrect side



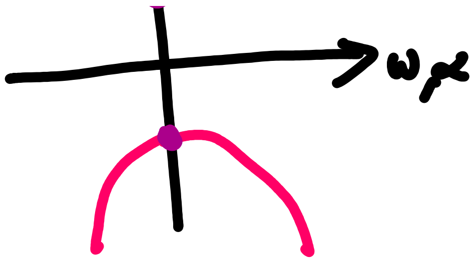
$$y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi^{(i)}$$
$$\xi^{(i)} \geq 0$$

- Both hard margin and soft margin SVM can be solved directly using constrained convex-optimization techniques.
- Both hard margin and soft margin SVM can also be solved using a loss-based approach as in DL.

Both hard and soft margins have the so-called dual formulation



$\boxed{\text{Find } w \text{ s.t. } \min w^2}$ $\Leftrightarrow$ $\boxed{\text{Find } \alpha \text{ s.t. } \max (\Sigma)^2 + \alpha(b-w)}$



s.t. $w \geq b$

Primal Problem

s.t. $\alpha \geq 0$

Dual Problem

$$\boxed{\begin{aligned}&\min \frac{\|w\|^2}{2}\\ &\text{s.t. } y^{(n)}(w^T x^{(n)} + b) \geq 1\end{aligned}} \Leftrightarrow \boxed{\begin{aligned}&\max\ \alpha^T 1 - \frac{1}{2}\alpha^T G \alpha\\ &\text{s.t. } \sum_i \alpha_i y^{(i)} = 0\end{aligned}}$$

n constraints = n samples → α = n-vector

G = n×n-matrix

$[G]_{i,j} = x^{(i)} \cdot x^{(j)}$

a measure of similarity between samples i and j

Eg. n=10,000 samples

Primal ⇒ 10,000 constraints

Dual ⇒ $\alpha_1 y^{(1)} + \alpha_2 y^{(2)} + \dots + \alpha_{10000} y^{(10000)} = 0$, 1 constraint

α = 10000-vector, G = 10000×10000-matrix

→ not preferred when we have a large number of samples and is typically preferred when there are a large number of features

$[G]_{ij} = x^{(i)} \cdot x^{(j)}$ = how similar are samples i and j

$$= \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \end{bmatrix} \cdot \begin{bmatrix} x_1^{(j)} \\ x_2^{(j)} \end{bmatrix}$$

Kernel trick $\Rightarrow$ $[G]_{ij} = x^{(i)}_{new} \cdot x^{(j)}_{new}$

$$= \phi(x^{(i)}) \cdot \phi(x^{(j)})$$

$$= x^{(i)}_{new} \cdot x^{(j)}_{new}$$