Example. Find a point $\vec{p}$ on the plane $x+y-2z=6$ which lies closest to the origin.

$\longrightarrow$:

Q: Why is this a constrained optimization problem?

$\longrightarrow$: Function we're trying to minimize

$$\tilde{D}(x,y,z) = x^2+y^2+z^2$$

subject to: $x+y-2z=6$.

In general, $f(x,y,z) \longrightarrow$ min/max

subject to the constraint $F(x,y,z)=0$

First, we construct the "Langrangian funtion"

$$L(x,y,z,\lambda) = f(x,y,z) + \lambda F(x,y,z)$$

Then, we optimize the function L as a f'tion of four variables (x,y,z,λ).

Back to our example:

$$\tilde{D}(x,y,z) = x^2+y^2+z^2 \longrightarrow \text{min}$$

subject to $F(x,y,z) = x+y-2z-6=0$

$$\Rightarrow L(x,y,z,\lambda) = x^2+y^2+z^2+\lambda(x+y-2z-6)$$

$$\frac{\partial L}{\partial x} = 2x+\lambda = 0 \Rightarrow x = -\frac{\lambda}{2}$$

$$\frac{\partial L}{\partial y} = 2y+\lambda = 0 \Rightarrow y = -\frac{\lambda}{2}$$

$$\frac{\partial L}{\partial z} = 2z-2\lambda = 0 \Rightarrow z = \lambda$$

$$\frac{\partial L}{\partial \lambda} = x+y-2z-6 = 0 \Rightarrow x+y-2z-6=0$$

$$-\frac{\lambda}{2}-\frac{\lambda}{2}-2\lambda = 6$$

$$\boxed{\lambda=-2}$$

$$\boxed{\vec{p} = (x,y,z) = (1,1,-2)}$$

$\square$

Margins & Separating Hyperplanes.

Linear classifiers can be described geometrically as separating hyperplanes.

Any affine function $x \longmapsto \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ determines a hyperplane in $\mathbb{R}^p$ our predictor space

More precisely, $\{x : \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = 0\}$ is a hyperplane splitting the space $\mathbb{R}^p$ into two "half spaces":

$$\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p > 0$$

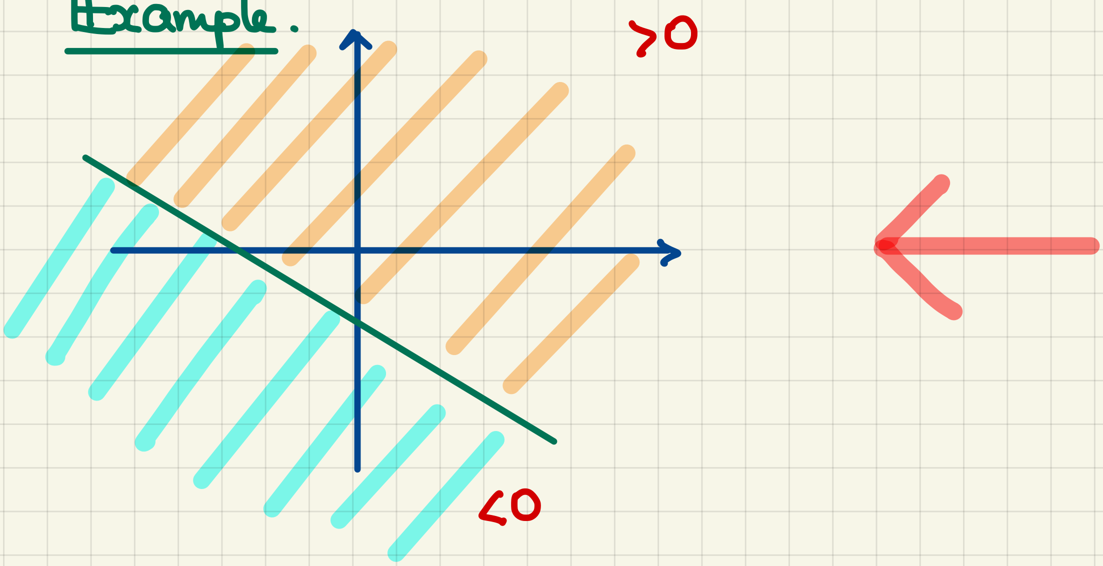
and

$$\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p < 0.$$

The vector $\vec{n} = (\beta_1, \beta_2, \dots, \beta_p)$ is the normal vector of our hyperplane. For a given hyperplane, we can always choose $\vec{n}$ so that $\|\vec{n}\| = 1$

Of course, the coefficient β_0 must also be scaled accordingly.

Example.



Note: • If the hyperplane goes through the origin, then
$\beta_0 = 0$

For any point x in the space, the deviation between it and the hyperplane is equal to

$$\beta \cdot x = \beta_1 x_1 + \cdots + \beta_p x_p$$

• If $\beta_0 \neq 0$, the hyperplane does not go through the origin. The deviation becomes

$$\boxed{\beta_0 + \beta \cdot x}$$

The sign tells us which side of the hyperplane the point x is.

Maximal Margin Classifier.

Suppose that we have a classification problem w/ two classes. We choose to encode these classes as $Y = -1$ and $Y = +1$.

Our criterion for the best among all the separating hyperplanes (if such exist) is to find the one w/ the largest possible margin around the hyperplane.

OPTIMIZATION PROBLEM.

We formulate the above task as

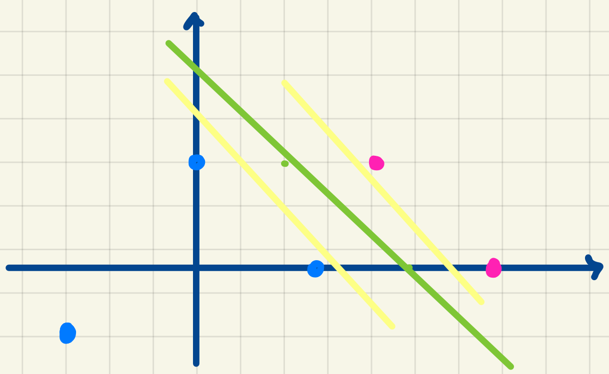
$$\max_{\beta_0, \beta_1, \ldots, \beta_p} M$$

subject to $\sum_{i=1}^{p} \beta_i^2 = 1$

and

$$y_i(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}) \geq M$$

for all $i = 1, \ldots, n$



maximal margin classifier

Reformulation of the Optimization Problem.

Define the vector

$$w = (w_1, \dots, w_p) = \frac{\beta}{M}$$

$$\boxed{\min_{\beta_0, w} \frac{1}{2} \|w\|^2}$$

subject to $y_i(\beta_0 + w_1 x_i + \dots + w_p x_{ip}) \geq 1$

for all $i = 1, \dots, n$

This is a quadratic optimization problem.

We introduce Karush-Kuhn-Tucker (KKT) multipliers

$$\lambda = (\lambda_1, \dots, \lambda_n)$$

Now, we have an optimization problem which is equivalent to

$$\max_{\lambda} \min_{\beta_0, w} \left(\frac{1}{2} \|w\|^2 - \sum_{i=1}^{n} \lambda_i \cdot (y_i(\beta_0 + w_1 x_{i1} + \dots + \overset{w_k x_{ik}}{} + w_p x_{ip}) - 1) \right)$$

subject to $\lambda_i \geq 0$ for all $i = 1..n$

We differentiate partially the above w.r.t. $\beta_0, w_1, \dots, w_p$

We get

$$w_k - \sum_{i=1}^{n} \lambda_i \cdot y_i x_{ik} = 0 \qquad \text{for all } k = 1..p$$

and

$$-\sum_{i=1}^{n} \lambda_i y_i = 0,$$

i.e.,

$$w_k = \sum_{i=1}^{n} \lambda_i y_i x_{ik} \quad \text{and} \quad \sum_{i=1}^{n} \lambda_i y_i = 0.$$

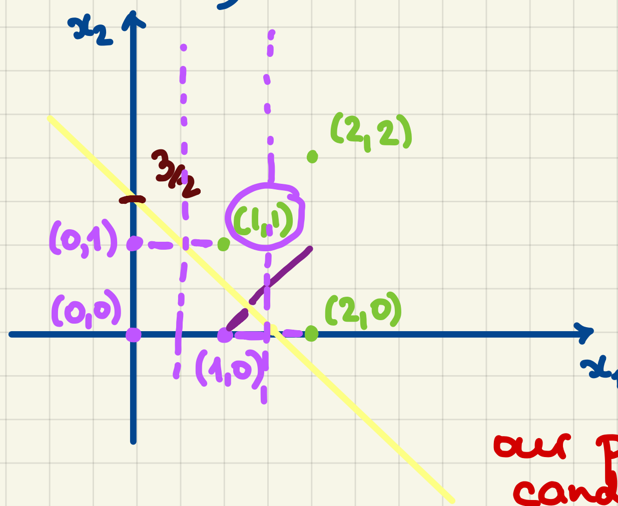
Moreover, by the KKT procedure, we know that

$$\lambda_i > 0 \quad \text{iff} \quad y_i(\beta_0 + w_1 x_{i1} + \dots + w_p x_{ip}) = 1,$$

i.e., the point x_i falls on the margin

Problem. Consider these training data.

	x_1	x_2	y
$i=1$	1	1	+1
$i=2$	2	2	+1
$i=3$	2	0	+1
$i=4$	0	0	−1
$i=5$	1	0	−1
$i=6$	0	1	−1



our principal candidate for the separating hyperplane.

w_1 and w_2 and $\beta_0 = ?$

$(0,1)$: $\beta_0 + w_1 \cdot 0 + w_2 \cdot 1 = -1$

$(1,0)$: $\beta_0 + w_1 \cdot 1 + w_2 \cdot 0 = -1$ $\Big\}$ $w_1 = w_2$ $\quad \beta_0 = -1 - w_1$

$(1,1)$: $\beta_0 + w_1 \cdot 1 + w_2 \cdot 1 = 1 \longrightarrow -1 \cancel{-w_1} \cancel{+w_1} + w_1 = 1 \Rightarrow w_1 = 2$

$(2,0)$: $\beta_0 + w_1 \cdot 2 + w_2 \cdot 0 = 1$

$\Rightarrow w_2 = 2$

$\Rightarrow \beta_0 = -3$

$\Rightarrow$ Our eq'n for the hyperplane:

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

$$2x_1 + 2x_2 = 3$$

$$x_1 + x_2 = \frac{3}{2} \qquad \boxed{x_2 = -x_1 + \frac{3}{2}}$$

$$\|w\|^2 = 2^2 + 2^2 = 8 \Rightarrow \|w\| = 2\sqrt{2} \Rightarrow M = \frac{1}{2\sqrt{2}} = \frac{\sqrt{2}}{4}$$

Task: Convince yourselves that the optimal margin does not increase if we discard (0,1) or (2,0).