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

For any point in the space, the deviation between it $x = (x_1, \dots, x_p)$ and the hyperplane w/ $\beta = (\beta_1, \dots, \beta_p)$ is equal to

$$x \cdot \beta = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

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

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

The sign tells us which side of the hyperplane we're.

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 the largest possible margin around the hyperplane.

OPTIMIZATION PROBLEM.

$\boxed{\text{November } 20^{th}, 2024.}$

The above task can be formulated as

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

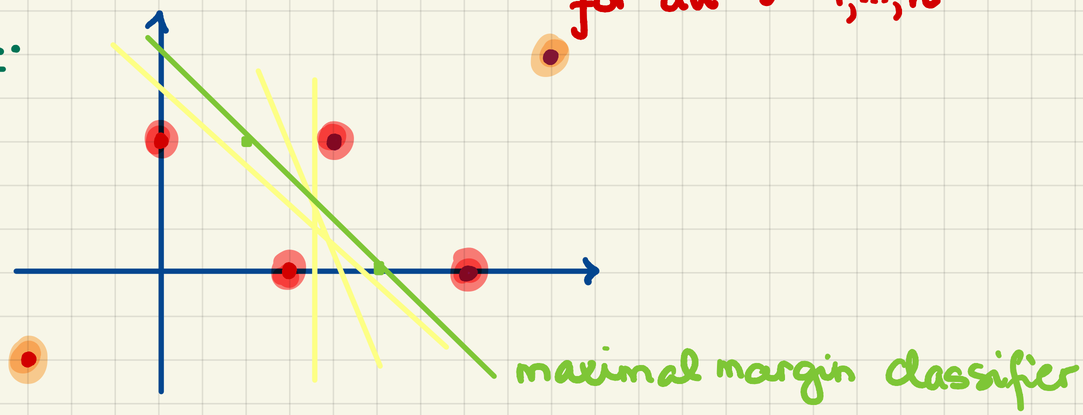
subject to $\displaystyle\sum_{j=1}^{p} \beta_j^2 = 1$

and

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M \quad \leftarrow$$

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

Example.



Reformulation of the Optimization Problem.

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

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

subject to $y_i(\beta_0 + w_1 x_{i1} + w_2 x_{i2} + \dots + w_p x_{ip}) \geq 1$

for all $\boxed{i = 1, \dots, n}$

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

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

subject to $\lambda_i \geq 0$

We differentiate partially above w/ respect to $\beta_0, w_1, \dots, w_p$

We get $\quad w_k - \sum_{i=1}^{n} \lambda_i y_i x_{ik} = 0 \quad$ for all $k = 1, \dots, p$

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

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

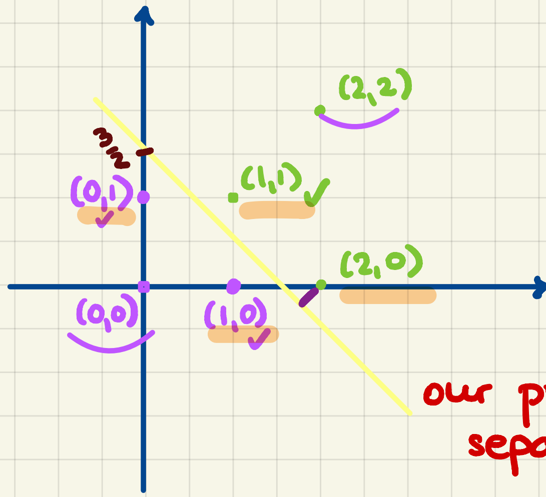
Moreover, by the KKT procedure, we have that

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

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 separating hyperplane candidate

w_1 and w_2 and $\beta_0 = ?$

$$\left.\begin{aligned}(0,1):&\quad \beta_0 + w_1\cdot 0 + w_2\cdot 1 = -1\\ (1,0):&\quad \beta_0 + w_1\cdot 1 + w_2\cdot 0 = -1\end{aligned}\right\}\quad w_1 = w_2 \qquad \beta_0 = -1 - w_1$$

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

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

$$\Downarrow$$

$$w_2 = 2$$

$$\Downarrow$$

$$\beta_0 = -3$$

$\Rightarrow$ <u>Our eq'n for the hyperplane</u>: $-3 + 2x_1 + 2x_2 = 0$

$$2x_1 + 2x_2 = 3$$

$$x_1 + x_2 = \frac{3}{2}$$

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

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