Classification problems

$$(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^D \times \{+1, -1, 0\}^N$$

For example tomorrow (hopefully) in Notebook 5.

x_i = picture of a piece of clothing

y_i = { "T-shirt-top", "Trousers", "Coat" "Sneaker", "Sandal", "bag" }

What other examples do you have in mind that come up in classification problems?

Binary Classification $\quad (x_i, y_i) \in \mathbb{R}^D \times \{-1, 1\}$

for example again picture data

First examples: deterministic models

$$f_\theta : \mathbb{R}^D \longrightarrow \{-1, +1, 0\}$$

boundary between categories

is not a cat

is a cat

Support Vector Machines (SVM's)

↳ optimize value of θ to best match data.

$$f_\theta(x) = \text{sgn}\left(\theta_0 + x^T\theta'\right)$$

where $\theta = (\theta_0, \underbrace{\theta_1, \dots, \theta_D}_{\theta'})^T \in \mathbb{R}^{D+1}$

* We are taking the linear model from last time, and using the model to "split" into categories, so

$$\text{Sgn}(x) = \text{sign of } x = \begin{cases} +1 & \text{if } x > \\ -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \end{cases}$$

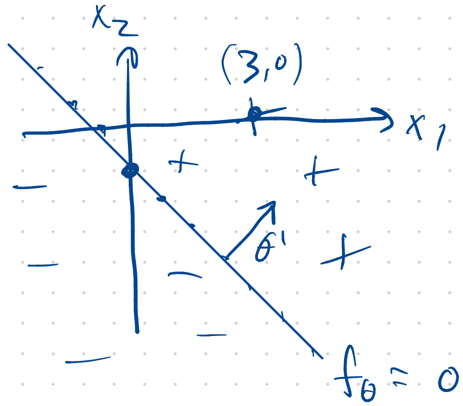
example $f_\theta\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \text{sgn}\left(\theta_0 + \theta_1 x_1 + \theta_2 x_2\right)$

If $\theta_1 = \theta_2 = 1$, $\theta_0 = 2$

$$f_{(2,1,1)}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \text{sgn}\left(2 + x_1 + x_2\right).$$

① Determine boundary of region when $f_\theta = 0$.

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



② Plug in point outside of boundary to determine +/− side

$x_1 = 3, \; x_2 = 0$

$$f_\theta\begin{pmatrix}3\\0\end{pmatrix} = \text{sgn}(2+3+0) = +1$$

$H = \{2 + x_1 + x_2\}$ called a line.

$H = \{2 + x_1 + x_2 + x_3\}$ is a plane

$H = \{2 + x_1 + x_2 \cdots + x_D\}$ is a <u>hyperplane</u>.

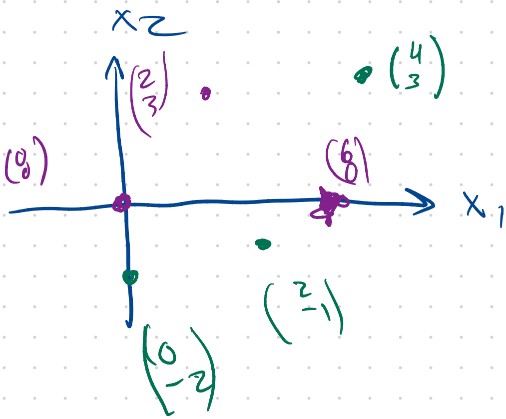
A hyperplane in $\mathbb{R}^D$ always splits $\mathbb{R}^D$ into two regions, so one side is $+1$, and the other -1.

$\boxed{\text{Remark}}$ We can take $\|\theta'\| = 1$ since

$\langle \theta', \theta' \rangle = 1$ gives same hyperplane (slope) as linear multiple.

Exercise · Given the following data points, come up with a hyperplane separating these points.

· Pick an "optimal" hyperplane + justify



· Add the point $\binom{6}{0}$ to the purple dataset. How do you account for this?

Lemma 3.22 For $y \in \{-1, 1\}$,

$$f_\theta(x) = y \iff y\,(\theta_0 + x^T\theta') > 0.$$

$\boxed{\text{Pf}}$ 5 min discussion

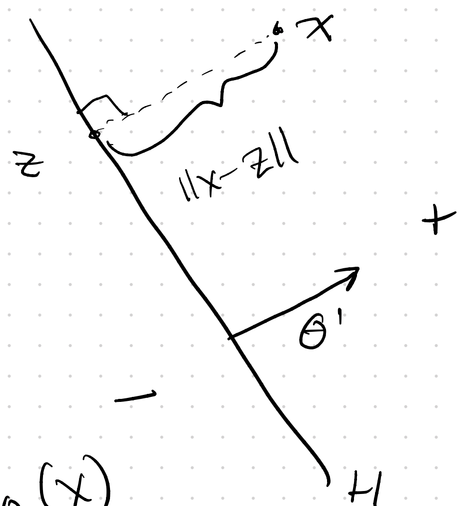
$\boxed{\text{Lemma 3.23}}$ Consider the hyperplane for $x \in \mathbb{R}^D$, $\theta \in \mathbb{R}^{D+1}$ with $\|\theta'\| = 1$.

$$H = \{ z \in \mathbb{R}^D \mid z^T \theta' + \theta_0 = 0 \}$$

The <u>Euclidean distance</u> from x to H is

$$y\left(x^T\theta' + \theta_0\right) \quad \text{where } y = f_\theta(x).$$

$\boxed{\text{Pf}}$



Let z be point minimizing distance to x

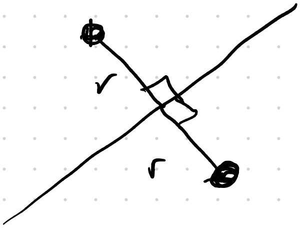
So $x = z \pm \|x - z\| \theta'$

So $\overbrace{\boxed{x^T\theta' + \theta_0}}^{f_\theta(x)} = \langle \theta', z \pm \|x-z\|\theta' \rangle + \theta_0$

$$= \left(\langle \theta', z \rangle + \theta_0 \right) \pm \|x - z\| \langle \theta', \theta' \rangle$$

$$= \pm \|x - z\|$$

Suppose our hyperplane separates data



To choose an optimal hyperplane, we can maximize distance to data points

That is we want to find

$$\max_{\substack{\theta \in \mathbb{R}^{D+1} \\ \langle \theta', \theta' \rangle = 1}} \left\{ r \geq 0 \;\middle|\; y_k \left(x_k^T \theta' + \theta_0 \right) \geq r, \quad k = 1, \dots, n \right\}$$

Normalize by r: $y_k \left(x_k^T \left(\frac{\theta'}{r} \right) + \left(\frac{\theta_0}{r} \right) \right) \geq 1$

$$\tilde{\theta} = \frac{\theta}{r}$$

then $\|\tilde{\theta}'\| = \frac{1}{r}$

maximizing $r \Rightarrow$ minimizing $\|\theta\|$

$\boxed{\text{Definition}}$ The Hard Margin SVM is given by optimization problem

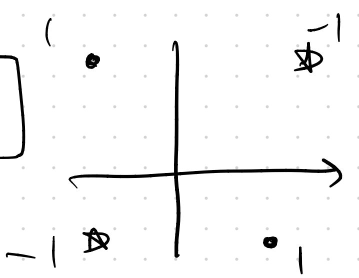
$$\min_{\theta \in \mathbb{R}^{D+1}} \left\{ \|\theta'\|^2 \;\middle|\; y_k(x_k^T \theta' + \theta_0) \geq 1 \text{ for } k = 1, \ldots, n \right\}$$

What happens when we have noisy data + outliers? add slack variable

$\boxed{\text{Definition}}$ The Soft margin SVM is given by the optimization problem

$$\min_{\theta, \xi} \left\{ \|\theta'\|^2 + C \sum_{k=1}^{n} \xi_k \;\middle|\; \begin{array}{l} y_k(x_k^T \theta' + \theta_0) \geq 1 - \xi_k \\ \text{for } \xi_k \geq 0,\ k = 1, \ldots, n \end{array} \right\}$$

$\boxed{\text{exercise}}$



Show hard margin SVM has no solution.