Today

- SVMs

Support Vector Machines

Data $(\vec{x}_1, y_1)$ $\quad$ $\vec{x}_i$: Feature vector

$(\vec{x}_2, y_2)$ $\quad$ y_i : label $\in \{+1, -1\}$

Training Data:

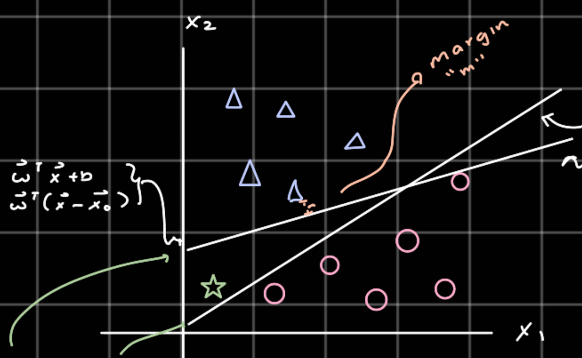
$(\vec{x}_{new}, ?)$

↳ want to find correct label

$$\begin{bmatrix} - & \vec{x}_1^{\,T} & - \\ & \vdots & \\ - & \vec{x}_n^{\,T} & - \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

↳ m features

↳ least squares won't work bc it'll give you a real # (ie, not +1 or -1)



what is the "best" hyperplane? determines where we should classify our new data

seperating hyperplane

$F(\vec{x}) = \vec{w}^T\vec{x} + b$ s.t.

if $\vec{w}^T\vec{x}_i + b > 0$ then $y_i = +1$

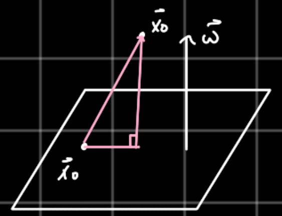
if $\vec{w}^T\vec{x}_i + b < 0$ then $y_i = -1$

$\forall i$:

$y_i(\vec{w}^T\vec{x}_i + b) > 0$

↳ looking for largest margin (orthogonal projection of closest point onto line) btwn points & all possible seperating hyperplanes

Distance btwn a point and a hyperplane



$\vec{w}^T(\vec{x}-\vec{x}_0) = 0$ $\qquad$ $b = -\vec{w}^T\vec{x}_0$

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

Project $(\vec{x}-\vec{x}_0)$ onto $\vec{w}$: $\dfrac{\vec{w}^T}{\|\vec{w}\|_2}(\vec{x}-\vec{x}_0) \Rightarrow \dfrac{|\vec{w}^T\vec{x} + b|}{\|\vec{w}\|_2}$

↳ this is the signed distance → abs. value gives distance away from sep. hyperplane

use this to formulate optimization problem:

$$\max_{\vec{\omega}, \vec{b}, m} \quad m$$

s.t. $y_i(\vec{\omega}^T \vec{x}_i + b) > 0 \quad \forall i$: classify correctly

$$\frac{|\vec{\omega}^T \vec{x}_i + b|}{\|\vec{\omega}\|_2} \geq m \quad \forall i$$: distance of i^{th} point must be at least m

$\omega^T x + b = 0$ ↓ true for $\forall \alpha$

$\alpha \vec{\omega}^T \vec{x} + \alpha b = 0$

b/c if $(m, \vec{\omega}, b)$ is a solution, then

$(m, \alpha\vec{\omega}, \alpha b)$ is also a solution

↳ want to constrain the norm of the normal vector

constrain norm of $\|w\|_2 = \frac{1}{m}$

$$\max_{\vec{\omega}, b} \quad (\|w\|_2^{-1})$$

s.t. $y_i(\vec{\omega}^T \vec{x}_i + b) \geq 1$

Monotone transformation:

$$\min_{\vec{\omega}, b} \quad \|\vec{\omega}\|_2^2$$

s.t. $y_i(\vec{\omega}^T \vec{x}_i + b) \geq 1$

☆ SVM optimization

↳ hard-margin SVM

(ie, we want a clean solution, a distinct hyperplane, but just bc we have an optimization problem, doesn't necessarily mean it's feasible)

Soft-margin SVM

regularizing constant

$$\min_{\vec{\omega}, \vec{b}, \xi_i} \quad \frac{1}{2}\|\vec{\omega}\|_2^2 + C\sum_{i=0}^{n} \xi_i$$

s.t. $y_i(\vec{\omega}^T \vec{x}_i + b) \geq \underbrace{1 - \xi_i}_{\text{slack variable.}}$

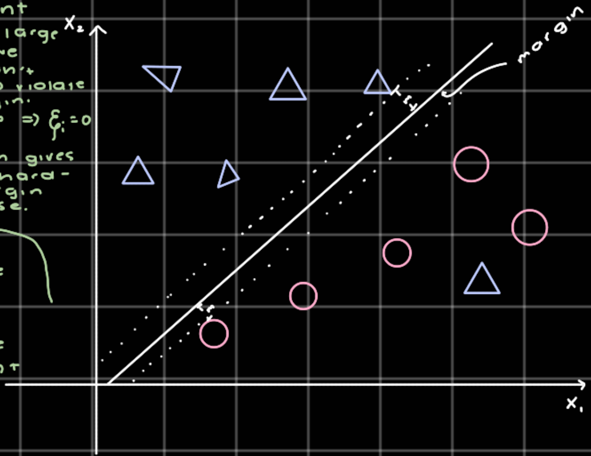
$\xi_i \geq 0$

C very large means we really don't want to violate the margin.

↳ eg $C \to \infty \Rightarrow \xi_i = 0$

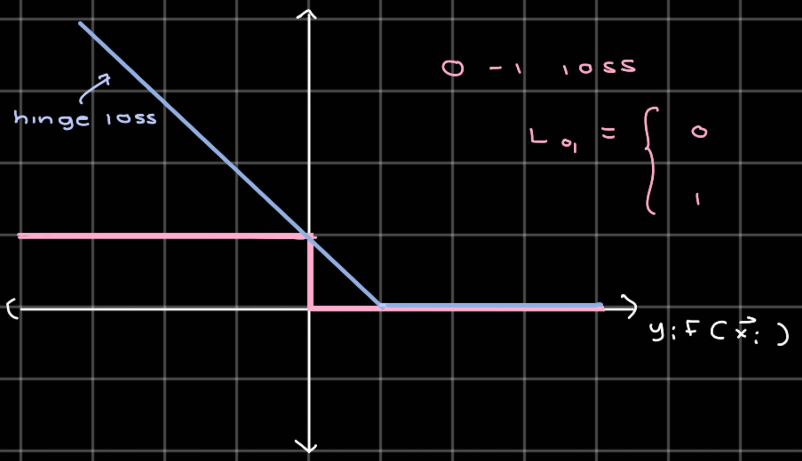
which gives the hard-margin case.

this is the margin we're going to allow; ie that we can violate our constraint by ξ_i



Hinge-loss perspective on classification:

$f(\vec{x}_i) > 0$ if $y_i = +1$
$f(\vec{x}_i) < 0$ if $y_i = -1$ $\Big\}\ y_i f(\vec{x}_i) > 0$



0-1 loss

$$L_{01} = \begin{cases} 0 & y_i f(x_i) > 0 \\ 1 & y_i f(x_i) < 0 \end{cases}$$

$$\text{minimize} \quad \frac{1}{n} \sum_{i=1}^{n} L_{01}(y_i, \vec{\omega}^T \vec{x}_i + b)$$

is not convex :(
↳ hinge loss! (convex!)

$$L_{hinge}(y_i, \omega^T x_i + b) = \max(1 - y_i(\omega^T x_i + b), 0)$$

$$\underset{\vec{\omega}, \vec{b}}{\text{minimize}} \quad \frac{1}{n} \sum L_{hinge}(y_i, \vec{\omega}^T \vec{x}_i + b) + \lambda \|\vec{\omega}\|_2^2$$ ☆ Hinge-loss formulation of the SVM

$$\min \quad \frac{1}{2} \|\vec{\omega}\|_2^2 + C \sum_{i=1}^{n} \xi_i$$
s.t. $\xi_i \geq 1 - y_i(\vec{\omega}^T \vec{x}_i + b)$
$\xi_i \geq 0$

$$= \min \quad \frac{1}{2} \|\vec{\omega}\|_2^2 + C \sum_{i=1}^{n} \xi_i$$
s.t. $\xi_i \geq \max(1 - y_i(\vec{\omega}^T \vec{x}_i + b), 0)$

$$= \min \quad \frac{1}{2} \|\vec{\omega}\|_2^2 + C \sum_{i=1}^{n} \xi_i$$
s.t. $\xi_i \geq L_{hinge}(y_i, f(x_i))$

Claim: At optimality, $\xi_i = \max(1 - y_i(\vec{\omega}^T \vec{x}_i + b), 0)$

$$\min \quad \frac{1}{2} \|\vec{\omega}\|_2^2 + 2C \sum_{i=1}^{n} L_{hinge}(y_i, f(x_i))$$

$$C = \frac{n\lambda}{2}$$