Soft margin SVM:

last time (hard margin) $y_i(x_i^T w + \alpha) \geq 1$

adapting for soft margin

↳ $y_i(x_i^T w + \alpha) \geq 1 - \varepsilon_i \quad : \quad \varepsilon_i \geq 0 \quad \forall i$

$$\boxed{\begin{gathered} \min_{w,\alpha,\varepsilon} \|w\|_2^2 + C\sum_{i=1}^{n} \varepsilon_i \\ y_i(x_i^T w + \alpha) \geq 1 - \varepsilon_i \\ \varepsilon_i \geq 0 \end{gathered}}$$

Small C => less weight on slack
=> more emphasis on the margin "width"
=> may underfit

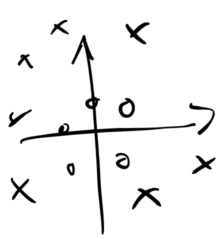
big C => keeps ε_i small
=> more emphasis on enforcing the margin
=> may overfit

How to select C? validation

Features:

Ex. parabolic lifting map

$\Phi : \mathbb{R}^d \Rightarrow \mathbb{R}^{d+1} \quad \Phi(x) = \begin{bmatrix} x \\ \|x\|^2 \end{bmatrix} \leftarrow |x|^2$



$\Phi(x_1) \cdots \Phi(x_n)$ is linearly seperable iff

$x_1 \cdots x_n$ is seperable by a hypersphere

Hypersphere in $\mathbb{R}^d$ w/ center c radius ρ

$$\|x-c\|_2^2 < \rho^2$$

$$\|x\|^2 + \|c\|^2 - 2c^T x < \rho$$

$$\begin{bmatrix} -2c^T & 1 \end{bmatrix} \begin{bmatrix} x \\ \|x\|^2 \end{bmatrix} < \rho^2 - \|c\|^2$$

come back to this tbh

ex 2. axis-aligned ellipsoid/hyperparaboloid

$$Ax_1^2 + Bx_2^2 + Cx_3^2 + Dx_1 + Ex_2 + Fx_3 + \alpha = 0$$

$$\Phi : \mathbb{R}^d \to \mathbb{R}^{2d}$$

$$\Phi(x) = \begin{bmatrix} x_1^2 & x_2^2 & \cdots & x_d^2 & x_1 & x_2 & \cdots & x_d \end{bmatrix}^T$$

$$\Phi(x)^T \underbrace{\begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \end{bmatrix}}_{w} + \alpha = 0 \quad \Rightarrow \quad \Phi(x)^T w + \alpha = 0$$

ex 3. $\|A\|$ general

$$Ax_1^2 + Bx_2^2 + Cx_2^2 + Dx_1x_2 + Ex_2x_3 + Fx_1x_3$$
$$+ Gx_1 + Hx_2 + Ix_3 + \alpha = 0$$

$$\Phi = \begin{bmatrix} \underbrace{x_1^2 \; x_2^2 \cdots x_d^2}_{d} & \underbrace{x_1x_2 \; x_1x_3 \cdots x_{d-1}x_d}_{\binom{d}{2} \sim d^2} & \underbrace{x_1 \cdots x_d}_{d} \end{bmatrix}$$

$$\Phi(x): \mathbb{R}^d \to \mathbb{R}^{(d^2+3d)/2} \qquad O(d^2)$$

"Quadrics"

ex 4. $\underbrace{\text{deg-}p\text{ polynomials}}$

$\Phi(x): \mathbb{R}^d \to \mathbb{R}^{O(d^p)}$

$\Phi(x) = [x_1^3 \;\; x_2^3 \;\; x_1^2 x_2 \;\; x_1 x_2^2 \;\; x_1 x_2 \;\; x_1 \;\; x_2]^T$

kernal trick!