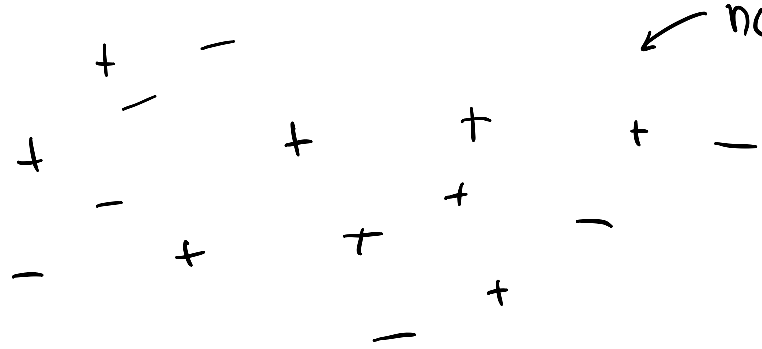
Nonseparable

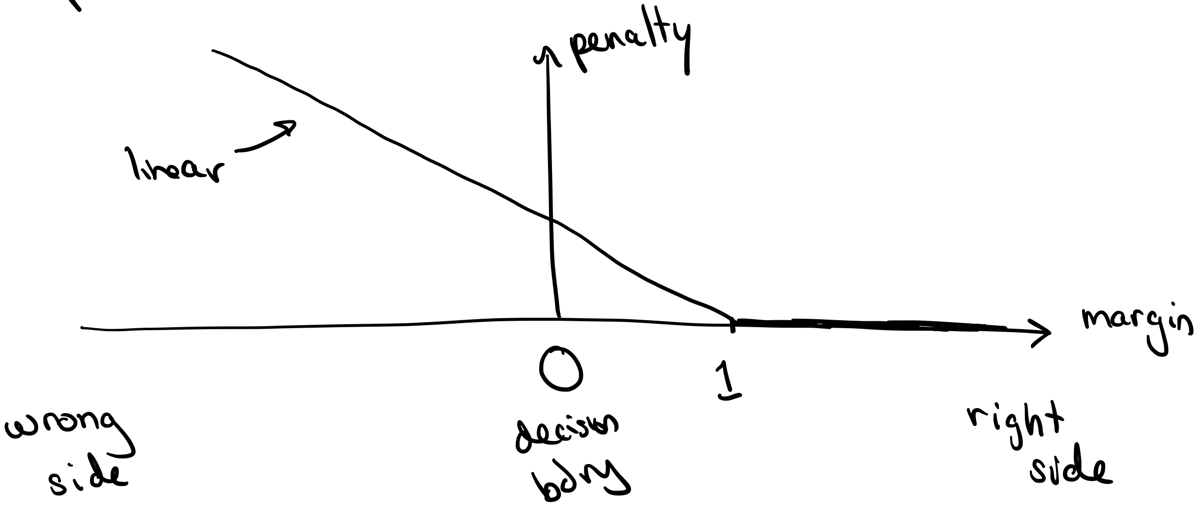
SVM

Non separable case



no separating hyperplane, so feasible solution to regular SVM

The fix: allow SVM to make mistakes but pay a price for each mistake



New primal:

$$\min_{\bar{\lambda}, \lambda_0, \bar{\xi}} \quad \frac{1}{2} \|\bar{\lambda}\|^2 + C \sum_{i=1}^{n} \xi_i$$

$$\text{s.t.} \quad \begin{cases} y_i(\bar{\lambda}^T \bar{x}_i + \lambda_0) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}$$

$$\min_{\bar{\lambda}, \lambda_0, \bar{\xi}} \quad \frac{1}{2}\|\bar{\lambda}\|^2 + C\sum_{i=1}^{n} \xi_i$$

$$\text{s.t.} \quad \begin{cases} y_i(\bar{\lambda}^T \bar{x}_i + \lambda_0) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}$$

$$\begin{cases} \text{if} \quad y_i f(x_i) \geq 1 \quad \text{then } \xi_i = 0, \text{ penalty } 0 \\ \text{otherwise} \quad y_i f(x_i) = 1 - \xi_i, \text{ penalty } \xi_i \end{cases}$$

C trades off between trying to get all points to have margin ≥ 1 and trying to reduce $\|\lambda\|^2$

$$\xi_i = \begin{cases} 0 & \text{if } y_i f(x_i) \geq 1 \\ 1 - y_i f(x_i) & \text{if } y_i f(x_i) \leq 1 \end{cases}$$

$$= \lfloor 1 - y_i f(x_i) \rfloor_+$$

$\lfloor z \rfloor_+$ means take max of z and 0

$$\min_{\bar{\lambda}, \lambda_0} \quad \underbrace{\frac{1}{2}\|\bar{\lambda}\|_2^2}_{} + \underbrace{C\sum_{i=1}^{n} \lfloor 1 - y_i f(x_i) \rfloor_+}_{}$$

SMO - a version of coordinate descent for SVM

↳ sequential minimal optimization

Say we want to take a coordinate descent step in the first direction

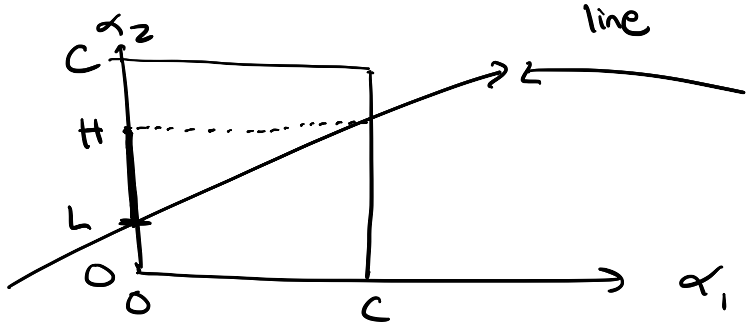
Fix $\alpha_2 \cdots \alpha_n$. Choose α_1 to maximize the dual objective.

Hmm...
$$\sum_{i=1}^{n} \alpha_i y_i = 0 \Rightarrow \alpha_1 y_1 = -\sum_{i=2}^{n} \alpha_i y_i$$

$$\alpha_1 = -\frac{1}{y_1} \sum_{i=2}^{n} \alpha_i y_i$$ it's stuck!

Can't just update α_1, need to update ≥ 2 of them.

Update α_1 & α_2: $\underbrace{\alpha_1 y_1 + \alpha_2 y_2 = -\sum_{i=3}^{n} \alpha_i y_i =: \Upsilon \text{ (fixed)}}_{\text{line}}$



also $0 \leq \alpha_1, \alpha_2 \leq C$

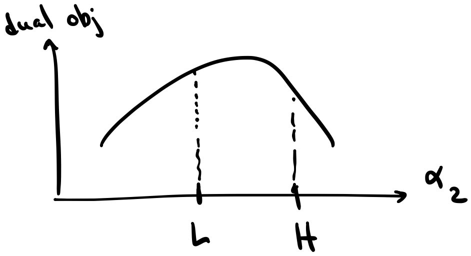
$[L, H]$ is range of α_2 so that $\alpha_1 \in [0, C]$

dual objective restricted to α_1, α_2 becomes:

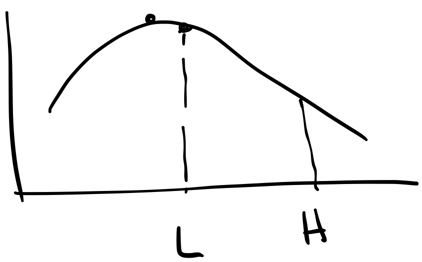
$$\max_{\alpha_2 \in [L,H]} \left[\alpha_1 + \alpha_2 + \text{constants} - \frac{1}{2} \sum_{ik} \alpha_i \alpha_k y_i y_k \vec{x}_i \vec{x}_k \right]$$

quadratic in α_2

get α_1 from the line, given α_2



- Set derivative to 0 to optimize α_2, and if optimal value is outside $[L, H]$, set it to either L or H.



- order of α's to update chosen by heuristics.