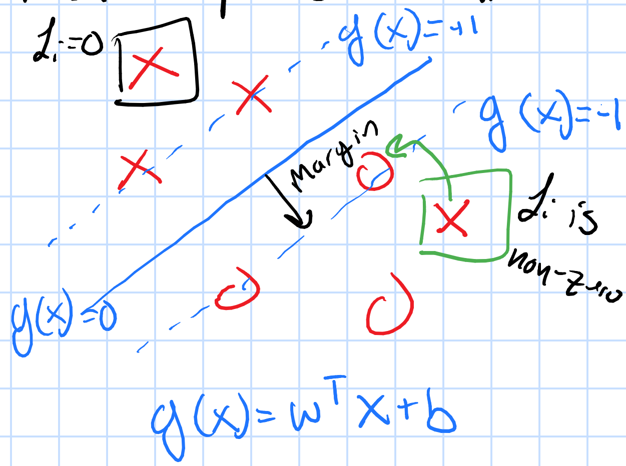
Recap

Our goal was to maximize the margin for a linear classifier (w/ parameters w & b). This optimization problem became

Primal

$$\boxed{\begin{aligned}&\arg\min \tfrac{1}{2}\|w\|_2^2\\ &\text{s.t. } y_i(w^T x_i + b) \geq 1\end{aligned}}$$

Form the Lagrangian



* Assumes all data are perfectly separable

$$\boxed{\alpha_i \geq 0}$$

$$L(w, b, \alpha) = \frac{1}{2}\boxed{\|w\|_2^2} - \sum_{i=1}^{n} \alpha_i \left[y_i (w^T x_i + b) - 1\right]$$

① $\dfrac{\partial L}{\partial w} = w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0 \implies \boxed{w = \sum_{i=1}^{n} \alpha_i y_i x_i}$ *

② $\dfrac{\partial L}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i = 0 \implies \boxed{\sum_{i=1}^{n} \alpha_i y_i = 0}$ $\boxed{\text{Another constraint}}$

$$L = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{n} \alpha_i y_i \boxed{w^T} x_i - b \cancel{\sum_{i=1}^{n} \alpha_i y_i}^{\,0} + \sum_{i=1}^{n} \alpha_i$$

$$= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{n} \alpha_i$$

$$L(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\boxed{\begin{aligned}&\max_{\alpha} L(\alpha)\\ &\text{s.t. } \sum_{i=1}^{n} \alpha_i y_i = 0,\ \alpha_i \geq 0 \quad \forall i\end{aligned}}$$

Dual Form

$$\arg\max_{\alpha} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \right\}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$\alpha_i \geq 0 \quad \forall i$$

This is a quadratic program.

side

$$\arg\min \; \alpha^T H \alpha + f^T \alpha$$

$$\text{s.t.} \quad A\alpha = b \qquad \mathbb{1}^T \alpha = 0$$

$$C\alpha \leq d$$

We assumed the data are perfectly seperable, which is generally not true. To address this we need slack variables.

$\swarrow$ zeta

$$\arg\min_{w,b} \; \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^{n} \zeta_i$$

$\zeta_i \geq 0$

$C \geq 0$ [user defined]

$$\text{s.t.} \quad y_i (w^T x_i + b) \geq 1 - \zeta_i$$

$$\boxed{\zeta_i \geq 0}$$

$$\boxed{-\zeta_i \leq 0}$$

$$L = \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \mu_i \xi_i - \sum_{i=1}^{n} \alpha_i \left[y_i(w^T x_i + b) - 1 + \xi_i \right]$$

Lagrange multiplier α $\mu \geq 0$

(1) $$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \mu_i = 0$$ $$\boxed{C = \alpha_i + \mu_i}$$

(2) $$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0$$ $$\boxed{w = \sum_{i=1}^{n} \alpha_i y_i x_i}$$

(3) $$\frac{\partial L}{\partial b} = 0$$ $$\boxed{\sum_{i=1}^{n} \alpha_i y_i = 0}$$

* KKT Conditions

$$\boxed{\xi_i \mu_i = 0}$$

$$L = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j + C\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i \xi_i - \sum_{i=1}^{n} \mu_i \xi_i$$

$$\boxed{C\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \xi_i(\alpha_i + \mu_i) = 0}$$

$$\boxed{\begin{aligned} &\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum\sum \alpha_i \alpha_j y_i y_j x_i^T x_j \\ &\text{s.t.} \quad \sum_{i=1}^{n} \alpha_i y_i = 0 \\ &\qquad 0 \leq \alpha_i \leq C \end{aligned}}$$