$$\mathcal{L}(w,\alpha,\beta) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m}\alpha_i\left(y^i(w^T x_i + b) - 1\right)$$

$$\theta_d(\alpha) = \min_{w,b} \mathcal{L}(w,\alpha)$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{m}\alpha_i y^i x_i = 0 \Rightarrow w = \sum_{i=1}^{m}\alpha_i y^i x_i \quad (1)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{m}\alpha_i y^i = 0$$

$$\boxed{\begin{array}{l}\max D(\alpha)\\ \text{s.t. } \alpha_i \geq 0\\ \sum \alpha_i y^i = 0\end{array}} \longrightarrow \alpha^*$$

$$\frac{1}{2}\sum_{i=1}^{m}\alpha_i y^i x_i \sum_{j=1}^{m}\alpha_j y^j x_j - \sum_{i=1}^{m}\alpha_i\Big(y^i\big(\underbrace{\big(\sum_{j=1}^{m}\alpha_j y^j x_j\big) x_i + b}\big) - 1\Big)$$

$$= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y^i y^j (x_i \cdot x_j) - \sum_{i=1}^{m}\alpha_i y^i\Big(\sum_{j=1}^{m}\alpha_j y^j x_j\Big)x_i - b\underbrace{\sum_{i=1}^{m}\alpha_i y^i}_{=0} + \sum_{i=1}^{m}\alpha_i$$

$$= \;'' \; - \sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y^i y^j (x_i \cdot x_j) + \sum_{i=1}^{m}\alpha_i$$

$$= -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y^i y^j (x_i \cdot x_j) + \sum_{i=1}^{m}\alpha_i \quad \text{---} \quad D(\alpha)$$

we substitute α^* in (1)

w^*

$\min_{w,b} \frac{1}{2}\|w\|^2$ s.t. $y^i(w^T x_i + b) \geq 1 \quad \forall i = 1 \ldots m$

$\Rightarrow -y^i(w^T x_i + b) + 1 \leq 0 \longrightarrow g_i(w)$

we solve dual problem. $-$ w^*, α^*, β^* satisfy KKT

$g_i(w) = 0$ if $y^i(w^T x_i + b) = 1$; fnⁿ margin $= 1$

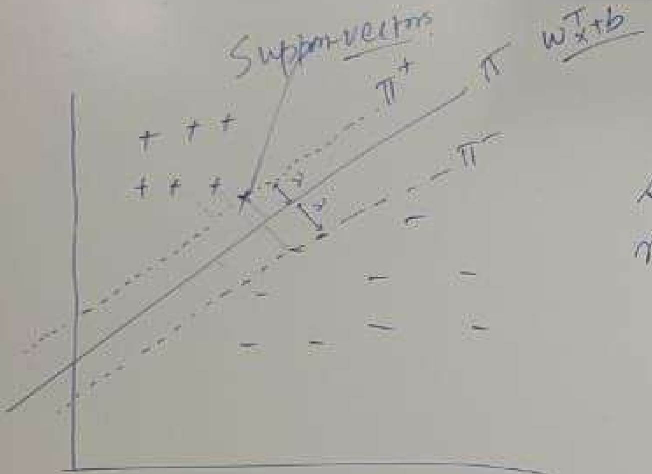
$\alpha_i > 0$

for the remaining data points, $\alpha_i = 0$ as $g_i(w^*) \neq 0$

$\alpha_i^* g_i(w^*) = 0$

Those data points whose fnⁿ margin $= 1$ are called support vectors.

Hard Margin



Max Margin Classifier

aim is to come up with a line/plane such that it will maximize the minimum margin

$$\max_{w,b} \gamma$$

s.t. Geometric margin$_i \geq \gamma$ $\quad i = 1 \ldots m$

$$\Rightarrow \max_{w,b} \frac{\hat{\gamma}}{\|w\|} \quad \text{s.t.} \quad y^i(w^T x_i + b) \geq \hat{\gamma}$$

$$\Rightarrow \max_{w,b} \frac{1}{\|w\|} \quad \text{s.t.} \quad y^i(w^T x_i + b) \geq 1$$

x

$$\gamma = \frac{w^T x + b}{\|w\|}$$

$$\hat{\gamma} = w^T x + b$$

$$\gamma = \frac{\hat{\gamma}}{\|w\|}$$

+1

−1

- Data points are linearly separable
- Geometric Margin γ
- Functional Margin $\hat{\gamma}$