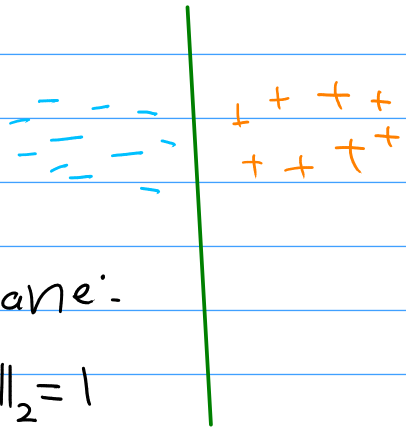
Hard-margin Support Vector Machine (SVM)

Setup

$(X, y) \in \mathbb{R}^d \times \{\pm 1\}$, dataset is lin. sep.



Recall Perceptron

(w', b')

Dist from X_i to hyperplane:

$\gamma_i = y_i(\langle w', X_i \rangle + b')$, $\|w'\|_2 = 1$

Margin: $\min_i \gamma_i$

SVMs: Perceptrons, but maximize the margin

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

$$\text{s.t. } \|w'\|_2 = 1$$

$$y_i(\langle w', x_i \rangle + b') \geq \gamma$$

Substitute $w' = \gamma w$, $b' = \gamma b$

$$\max_{\gamma w, \gamma b} \gamma \qquad\qquad \max_{\gamma w, \gamma b} \frac{1}{\|w\|_2}$$

$$\text{s.t. } \|w\|_2 = \frac{1}{\gamma} \quad \Longrightarrow \quad \text{s.t. } y_i(\langle w, X_i \rangle + b) \geq 1$$

$$y_i(\langle w, x_i \rangle + b) \geq 1$$

$$\Rightarrow \boxed{\min_{w,b} \frac{1}{2}\|w\|_2^2, \text{ s.t. } y_i \underbrace{(\langle w, X_i \rangle + b)}_{\hat{y}_i = \langle w, X_i \rangle + b} \geq 1}$$

$\text{sign}(\hat{y}_i)$

C.f. Perceptron: $\min_{w,b} 0$, s.t. $y_i \hat{y}_i \geq 1$

Regularization

Optimize?

Lagrangian Duality
-Alternate view

$$\boxed{\min_{w,b} \frac{1}{2}\|w\|_2^2 \ , \ \text{s.t. } y_i(\langle w, x_i\rangle + b) \geq 1}$$ Primal

$$= \min_{w,b} \max_{\alpha \geq 0} \frac{1}{2}\|w\|_2^2 - \sum_i \alpha_i \left(y_i(\langle w, x_i\rangle + b) - 1\right)$$

Primal vars $\rightarrow w, b$

$\alpha \geq 0$, $\alpha \in \mathbb{R}^n$

dual var.

strong duality

$$= \max_{\alpha \geq 0} \min_{w,b} \frac{1}{2}\|w\|_2^2 - \sum_i \alpha_i \left(y_i(\langle w, x_i\rangle + b) - 1\right)$$

$$\frac{\partial}{\partial b} = -\sum_i \alpha_i y_i = 0$$

$$\frac{\partial}{\partial w} = w - \sum_i \alpha_i y_i x_i = 0 \Rightarrow w = \sum_i \alpha_i y_i x_i$$

$$= \max_{\alpha \geq 0} \frac{1}{2}\left\|\sum \alpha_i y_i x_i\right\|_2^2 + \sum_i \alpha_i - \cancel{\sum_i \alpha_i y_i b} - \sum_i \alpha_i y_i \langle w, x_i\rangle$$

$$-\sum_i \left(\sum_j \alpha_j y_j x_j\right)^T x_i \alpha_i y_i$$

$$= -\left(\sum_j \alpha_j y_j x_j\right)^T\left(\sum_i \alpha_i y_i x_i\right)$$

$$= -\left\|\sum_i \alpha_i y_i x_i\right\|_2^2$$

$$= \max_{\alpha \geq 0} -\frac{1}{2}\left\|\sum_i \alpha_i y_i x_i\right\|_2^2 + \sum_i \alpha_i$$

$$\text{s.t. } \sum \alpha_i y_i = 0$$

$$= \boxed{\begin{array}{l} \min_{\alpha \geq 0} \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \langle x_i, x_j\rangle - \sum_i \alpha_i \\ \text{s.t. } \sum_i \alpha_i y_i = 0 \end{array}}$$

Dual

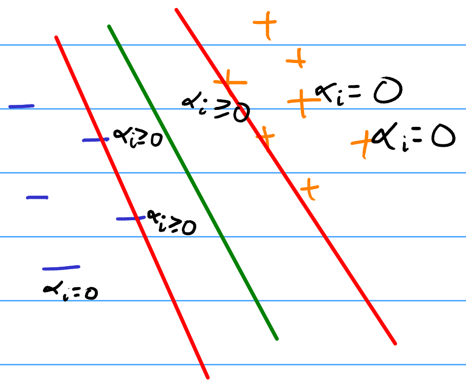
KKT Conditions

Necessary conds for optimality

- primal feasibility
- dual feasibility
- stationarity.
- Complementary slackness

$$\forall i,\ \alpha_i\left(y_i(\underbrace{\langle w, x_i\rangle + b}_{\hat{y}_i}) - 1\right) = 0$$

either: $\alpha_i = 0$, OR $y_i \hat{y}_i - 1 = 0$ (or both $=0$)



$$w = \sum_{i:\alpha_i > 0} \alpha_i y_i x_i$$

If $\alpha_i > 0$, x_i is a support vector