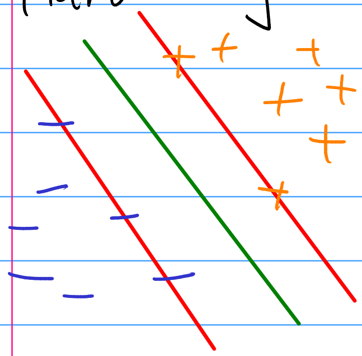
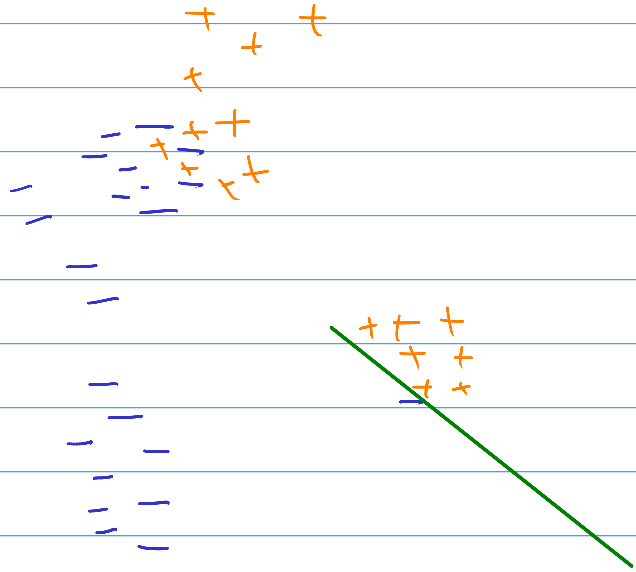
Soft-Margin Support Vector Machine

Hard-margin Drawbacks





Hard to Soft Margin

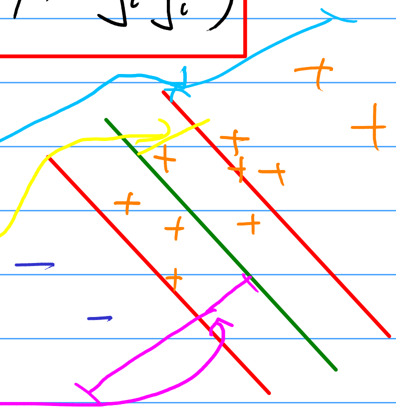
Hard: $\min_{w,b} \frac{1}{2}\|w\|_2^2$ s.t. $y_i(\underbrace{\langle w, x_i\rangle + b}_{\hat{y}_i}) \geq 1 \Leftrightarrow 1 - y_i\hat{y}_i \leq 0 \quad \forall i.$

$$\boxed{\text{Soft: } \min_{w,b} \frac{1}{2}\|w\|_2^2 + C\sum_i \max\left(0, 1 - y_i\hat{y}_i\right)}$$

$1 - y_i\hat{y}_i \leq 0$, on right side of margin

$0 \leq y_i\hat{y}_i \leq 1$, within margin

$y_i\hat{y} < 0$, incorrect



$C = 0 \Rightarrow$ Ignore data

$C = \infty$, only care about data (Hard-margin SVM)

Duality Primal

$$\min_{w,b} \frac{1}{2}\|w\|_2^2 + C\sum_i \max(0, 1-y_i\hat{y}_i)$$

$$= \min_{w,b,\gamma} \frac{1}{2}\|w\|_2^2 + C\sum \gamma_i$$

s.t. $\forall i \max(0, 1-y_i\hat{y}_i) \le \gamma_i$

$$= \min_{w,b,\gamma} \frac{1}{2}\|w\|_2^2 + C\sum \gamma_i$$

s.t. $\forall i$, $0 \le \gamma_i$

$1 - y_i\hat{y}_i \le \gamma_i$

$$= \min_{w,b,\gamma} \max_{\alpha,\beta\ge 0} \frac{1}{2}\|w\|_2^2 + \sum_i C\gamma_i + \alpha_i(1-y_i\hat{y}_i - \gamma_i) - \beta_i\gamma_i$$

$$= \max_{\alpha,\beta\ge 0} \min_{w,b,\gamma} \frac{1}{2}\|w\|_2^2 + \sum_i C\gamma_i + \alpha_i(1-y_i\hat{y}_i - \gamma_i) - \beta_i\gamma_i$$

$$\frac{\partial}{\partial w} = w - \sum \alpha_i y_i x_i = 0, \quad \frac{\partial}{\partial b} = -\sum \alpha_i y_i = 0$$

$$\frac{\partial}{\partial \gamma_i} = C - \alpha_i - \beta_i = 0 \Rightarrow C = \alpha_i + \beta_i$$

$$\frac{1}{2}\|w\|_2^2 = \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle$$

$$\sum -\alpha_i y_i \hat{y}_i = -\sum \alpha_i y_i(\langle w, x_i\rangle + b) = -\left(\sum \alpha_i y_i \langle w, x_i\rangle + \sum \alpha_i y_i b\right)$$

$$= -\sum_i\sum_j \alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle$$

$$\sum_i C\gamma_i + \alpha_i - \alpha_i\gamma_i - \beta_i\gamma_i = \sum (\alpha_i+\beta_i)\gamma_i + \alpha_i - \alpha_i\gamma_i - \beta_i\gamma_i$$

$$= \sum \alpha_i$$

Diffs v.s. hard margin.

$$\min_{\alpha, \beta \geq 0} \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle - \sum_i \alpha_i$$

s.t. $\sum \alpha_i y_i = 0$

$\forall i$, $C = \alpha_i + \beta_i \Rightarrow \beta_i = C - \alpha_i \geq 0 \Leftrightarrow C \geq \alpha_i$.

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

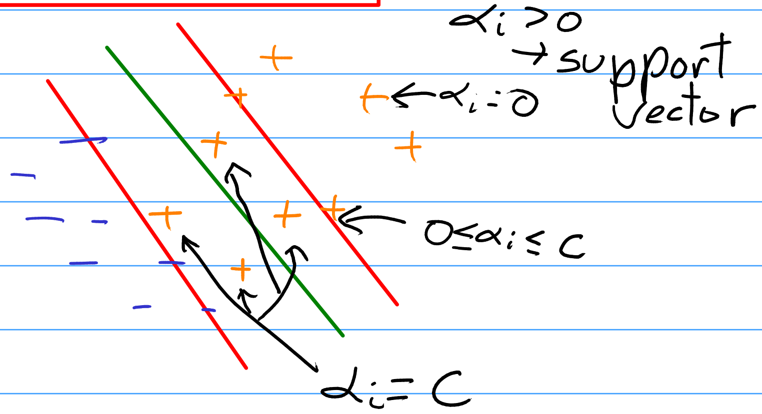
Dual soft SVM

$w = \sum \alpha_i y_i x_i$

$\forall i$ $\boxed{\alpha_i(1 - y_i \hat{y}_i - \gamma_i) = 0}$

$\beta_i \gamma_i = 0$

$\hookrightarrow \boxed{(C - \alpha_i)\gamma_i = 0}$



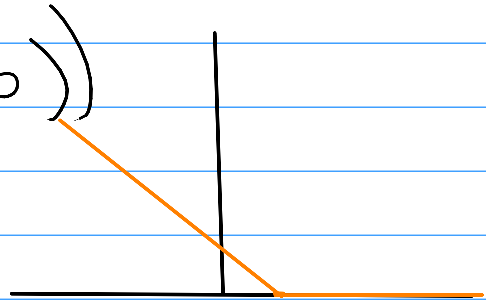
Suppose $\alpha_i > 0$, $\Rightarrow 1 - y_i \hat{y}_i - \gamma_i = 0$

Say $\gamma_i = 0$, $\Rightarrow 0 < \alpha_i \leq C$

Say $\gamma_i > 0 \Rightarrow \alpha_i = C$

Optimization

$$\ell_{w,b}(x,y) = \max(0, 1 - y(\langle w, x \rangle + b))$$

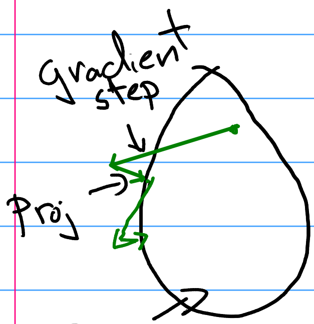


$$L = \frac{C}{n}\sum \ell(x_i, y_i) + \frac{1}{2}\|w\|_2^2$$

Gradient descent

$$\frac{\partial L}{\partial w} = w + \frac{C}{n}\sum \delta_i \qquad \delta_i = \begin{cases} 0 & \text{if } 1 - y_i \hat{y}_i \le 0 \\ -y_i x_i & \text{if } 1 - y_i \hat{y}_i \ge 0 \end{cases}$$

Projected gradient descent



$$C \ge \alpha \ge 0$$

$$\sum \alpha_i y_i = 0$$