Support Vector Machines

Mon Feb 22

$$y = \text{sign}(w^T x + w_0)$$

$$p(y=+1) = \frac{1}{1+\exp(w^T x)}$$

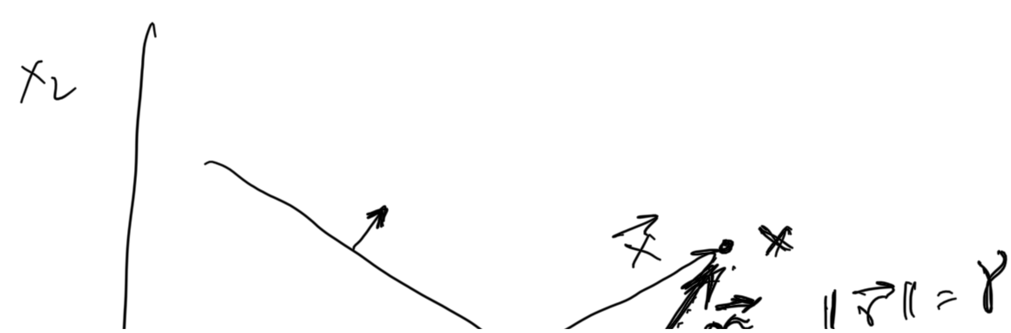
If $p(\hat{y}=+1) \geq 0.5$ $\quad \hat{y} = +1$

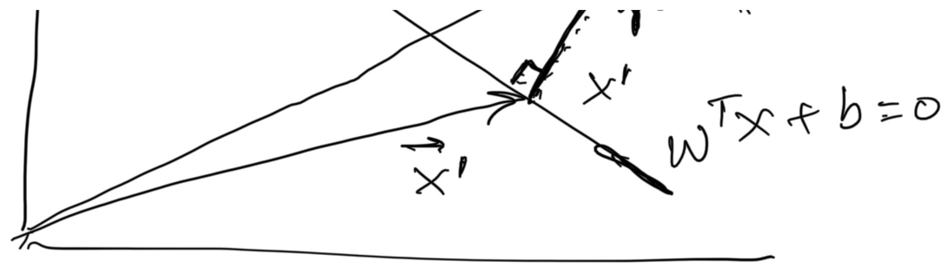
For a new data point:

$p(y=+1) = 0.95$

For another data point:

$p(y=+1) = 0.55$





$$\boxed{\vec{x} = \vec{x}' + \vec{r}}$$

$$\vec{x}' = \vec{x} - \vec{r}$$

$$\hat{w} = \frac{\vec{w}}{\|w\|}$$ → unit vector orthogonal to the line

$$\|w\| = \sqrt{w_1^2 + w^2 + \cdots} = \sqrt{w^T w}$$

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

$$\vec{x}' = \vec{x} - \frac{\gamma \vec{w}}{\|w\|}$$

Since $\vec{x}'$ lies on the decision line

$$w^T x' + b = 0$$

$$w^T \left(\vec{x} - \frac{\gamma \vec{w}}{\|w\|} \right) + b = 0$$

$$w^T x - \gamma \frac{w^T w}{\|w\|} + b = 0$$

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

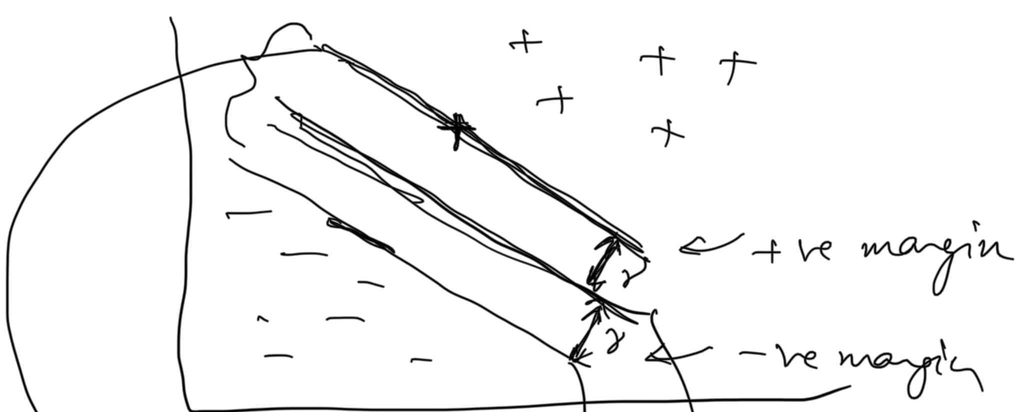
$$\gamma = \frac{w^T x + b}{\|w\|} \quad \left[\text{since } w^T w = \|w\|^2 \right]$$

for a −ve example

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

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

This is margin of an example.



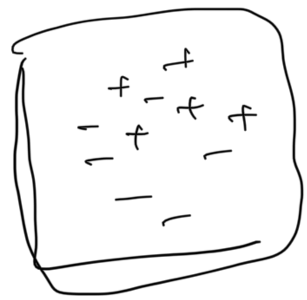
$$w^T x + b = +1$$

$w^T x + b = 0$

$w^T x + b = -1$

Distance between two margins:

will be $\frac{2}{\|w\|}$



---

i^{th} Training example: x_i, y_i

If w, b has to correctly classify this instance

$$y_i(w^T x_i + b) \geq 0$$

If w, b have to be on the correct side of the margin:

$$y_i(w^T x_i + b) \geq 1$$

Opt with equality constraints.

$$f(x,y) = x^2 + 2y^2 - 2$$

$$h(x,y) = \boxed{x + y - 1 = 0}$$

$$L(x, y, \beta) = x^2 + 2y^2 - 2 + \beta(x + y - 1) = 0$$

$$\frac{\partial L}{\partial x} = 2x + \beta = 0$$

$$\frac{\partial L}{\partial y} = 4y + \beta = 0$$

$$\frac{\partial L}{\partial \beta} = x + y - 1 = 0$$

$$4y - 2x = 0$$

$$x = 2y$$

$$x = 1 - y$$

$$1 - y = 2y$$

$$\boxed{y = \frac{1}{3} \qquad x = \frac{2}{3}}$$

Wednesday Feb 24

$$\nabla f(w_1, w_2) = \lambda \nabla g(w_1, w_2)$$

$$g(w_1, w_2) = 0$$

$$L = f(w_1, w_2) + \lambda g(w_1, w_2)$$

$$\nabla_{(w_1, w_2)} L = \nabla f(w_1, w_2) + \lambda \nabla g(w_1, w_2) = 0$$

$$\cancel{\nabla L} \quad g(w_1, w_2) = 0$$

$$f(x,y) = x^3 + y^2$$

$$g(x): \quad x^2 - 1 \le 0$$

$$L(x, y, \alpha) = f(x,y) + \alpha g(x,y)$$

$$= x^3 + y^2 + \alpha(x^2 - 1)$$

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\|\mathbf{w}\|^2 = w_1^2 + w_2^2$$

$$\min_{w_1, w_2} f \quad \frac{w_1^2 + w_2^2}{2}$$

$x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $y_1 = -1$

$$\text{s.t.} \quad 1 - (-1)(w_1 + w_2 + b) \le 0$$

$$\equiv 1 + w_1 + w_2 + b \le 0$$

$x_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ $y_2 = +1$

$$1 - (+1)(2w_1 + 2w_2 + b) \le 0$$

$$1 - 2w_1 - 2w_2 - b \le 0$$

$$L = \boxed{\frac{w_1^2 + w_2^2}{2} + \alpha_1 (1 + w_1 + w_2 + b) + \alpha_2 (1 - 2w_1 - 2w_2 - b)}$$

$$\boxed{\alpha_1, \alpha_2 \ge 0}$$

$$\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \frac{\partial L}{\partial \alpha_1}, \frac{\partial L}{\partial \alpha_2}, \frac{\partial L}{\partial b}$$

$$L(w, \alpha, \beta)$$

$\min f(x)$

Assume w is fixed

$$\theta_p(w) \quad \boxed{\max_{\alpha,\beta \mid \alpha_i \geq 0} \mathcal{L}(w,\alpha,\beta)}$$

$$\boxed{\min_{w} \theta_p(w)}$$

min-max is going to give us the optimal solution

---

what if we solve:

$$\max_{\alpha,\beta\,;\,\alpha_i \geq 0} \left[\min_{w} \mathcal{L}(w,\alpha,\beta) \right] \rightarrow \underline{\text{dual}}$$

<u>max-min</u>

$$\boxed{\text{max-min} \leq \text{min-max}}$$

"duality-gap"

$f(w_1, w_2)$

$$L_p = \frac{\|w\|^2}{2} + \sum_{i=1}^{N} \alpha_i \{1 - y_i(w^T x_i + b)\}$$

$$\frac{\partial L}{\partial w} = \frac{1}{\cancel{2}} \cancel{*} \cancel{2} w + \sum_{i=1}^{N} \alpha_i (-y_i x_i)$$

$$\boxed{\begin{aligned} &\|w\|^2 \\ &= w^T w \\ &\frac{d}{dw} w^T w = 2w \end{aligned}}$$

$$= w - \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\boxed{\begin{aligned} w &= \sum_{i=1}^{N} \alpha_i y_i x_i \\ \sum_{i=1}^{N} \alpha_i y_i &= 0 \end{aligned}}$$

$$L_p = \frac{\|w\|^2}{2} + \sum \alpha_i \{1 - y_i(w^T x_i + b)\}$$

$$= \frac{1}{2} w^T w + \sum \alpha_i \{1 - y_i(w^T x_i + b)\}$$

Replace w with $\sum_{i=1}^{N} \alpha_i y_i x_i$

Solve the dual to get $\alpha_1, \alpha_2 \dots \alpha_N$

cxprod

Feed the α's to get w

Friday Feb 26 $(w^T x + b)$

$$J(w) \quad \min_{w,b} \frac{||w||^2}{2}$$

$$\text{s.t } 1 - [y_i(w^T x_i + b)] \le 0 \quad i = 1 \dots N$$

$$L_p(w, b, \cancel{\alpha}) = \frac{||w||^2}{2} + \sum_{i=1}^{N} \alpha_i \{1 - y_i(w^T x_i + b)\}$$

$$\text{s.t } \boxed{\alpha_i \ge 0} \quad i = 1, \dots, N$$

Primal

$$L_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{m,n=1}^{N} \alpha_m \alpha_n y_m y_n (x_m^T x_n)$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad i = 1, \cdots N$$

Dual

$$W = \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\begin{bmatrix} x_1 , & y_1 \\ x_2 , & y_2 \end{bmatrix} \quad N = 2$$

$$L_D(\alpha) = (\alpha_1 + \alpha_2) - \frac{1}{2} \Big[\alpha_1 \alpha_1 y_1 y_1 x_1^T x_1 + \alpha_1 \alpha_2 y_1 y_2 x_1^T x_2 + \alpha_2 \alpha_1 y_2 y_1 x_2^T x_1 + \alpha_2 \alpha_2 y_2 y_2 x_2^T x_2 \Big]$$

S.t $\alpha_1 y_1 + \alpha_2 y_2 = 0$

and $\alpha_1 \geq 0$

and $\alpha_2 \geq 0$

Find α_1 and α_2 that maximize $L_D(\alpha)$

α_1^*, α_2^*

$$W = \alpha_1^* y_1 x_1 + \alpha_2^* y_2 x_2$$

$$\boxed{b = -\frac{w^T x_1 + w^T x_2}{2}}$$

For a test instance x^* $y^* - ?$

$$y^* = \text{sign}(w^T x^* + b)$$

$$= \text{sign}\left(\left(\sum \alpha_i y_i x_i\right)^T x^* + b\right)$$

$$= \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i \underline{(x_i^T x^*)} + b\right)$$

$$J(w, b)$$

$$L_P(w, b, \alpha)$$

$$\underline{L_D(\alpha)}$$

KKT #5 $\underline{\alpha_i(1 - y_i\{w^T x_i + b\})} = 0$ $\forall i$

For any x_i that lies on one of the

margins: $y_i \{ w^T x_i + b \} = 1$

and for any x_i that does not lie on the margin:

$$y_i \{ w^T x_i + b \} > 1$$

α_i will be 0

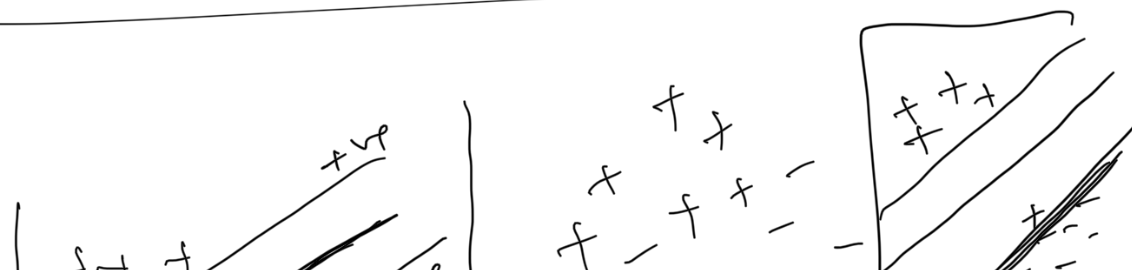
Support vectors:

$$y^* = \text{sign}(w^T x^* + b)$$

$$= \text{sign}\left(\left(\sum \alpha_i y_i x_i\right)^T x^* + b \right)$$

$$= \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i (x_i^T x^*) + b \right)$$

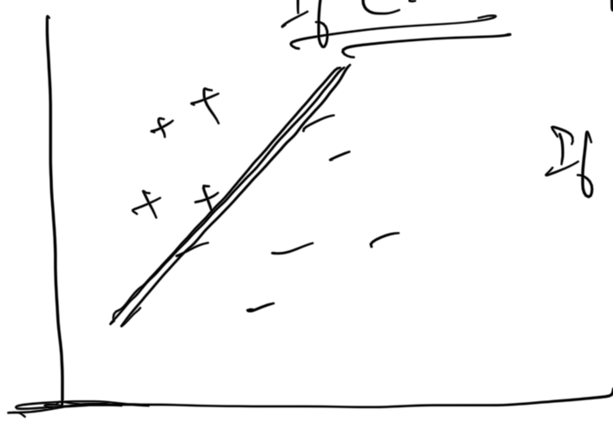
$$\boxed{y^* = \text{sign}\left(\sum_{i \in SV} \alpha_i y_i (x_i^T x^*) + b \right)}$$



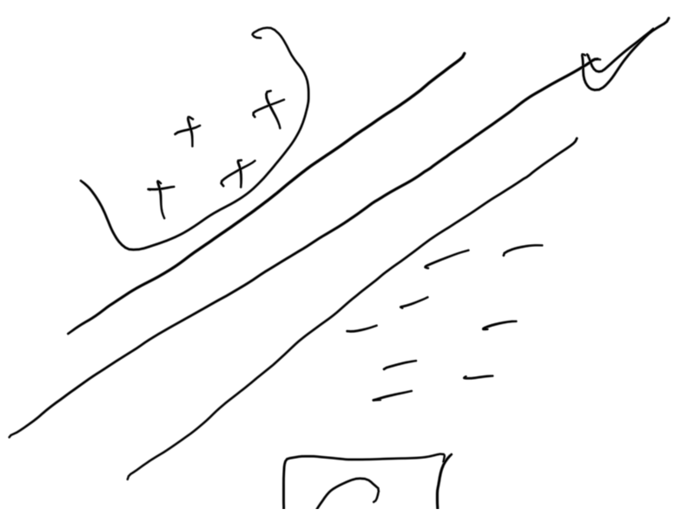
$\xi_i \rightarrow$ slack for training example i

$$\boxed{1 - y_i (w^T x_i + b)}$$

If $C = 0$ $\quad w = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$



If C is very high



$\boxed{\text{Bias-Variance Tradeoff}}$