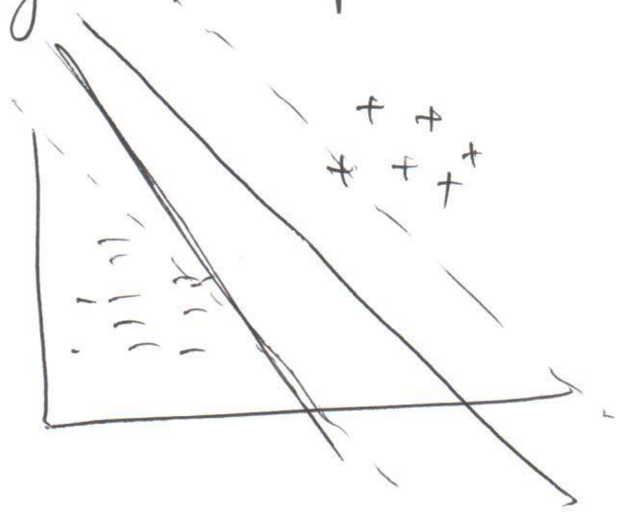
$$H = \frac{d}{dw} \nabla J(w)$$

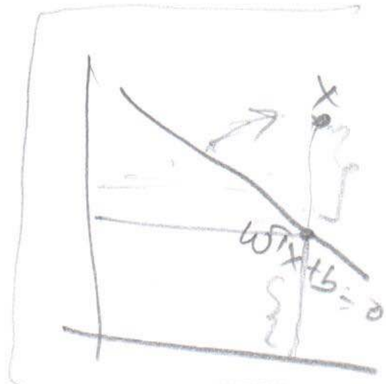
Maximum Margin Principle.

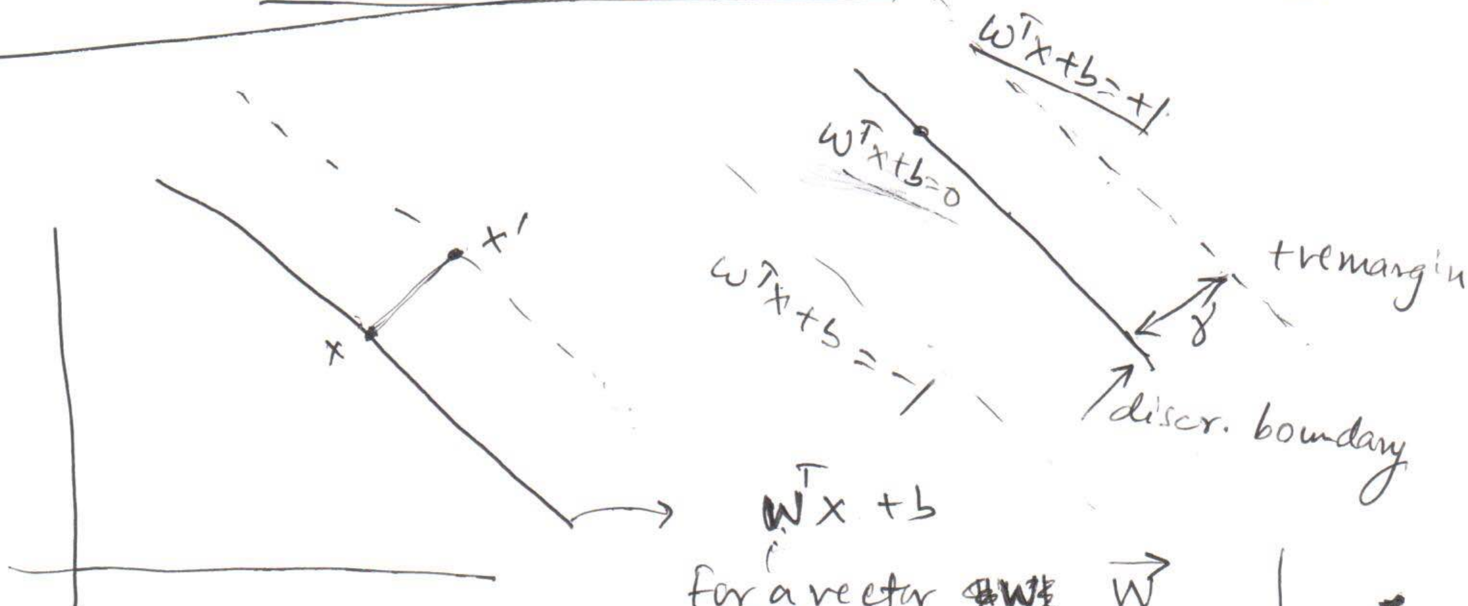


02/15/19

Hinge loss = 0

$\max(0, 1 - y_i(w^T x_i + b))$





$w^T x + b$

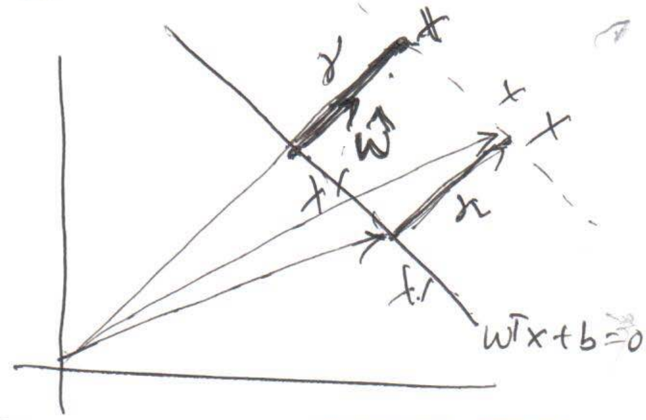
for a vector ~~w~~ $\vec{w}$

l_2 norm: $\sqrt{\sum_{i=1}^{d} w_i^2} = \|w\|$

$w^T w = \|w\|^2$

$\hat{w} = \frac{w}{\|w\|}$ unit vector pointing in the same direction as w

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



$r = x - x'$

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

$w = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}$

$\|w\| = \sqrt{0.4^2 + 0.6^2}$

$\hat{w} = \begin{bmatrix} \frac{0.4}{\sqrt{.52}} \\ \frac{0.6}{\sqrt{0.52}} \end{bmatrix}$

$w^T x' + b = 0 \longrightarrow x'$ lies on the line.

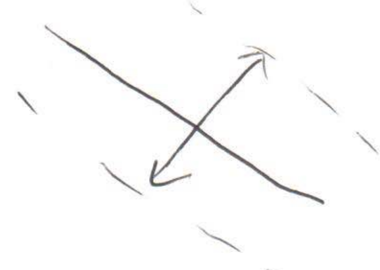
$w^T \left(x - \frac{\gamma w}{\|w\|} \right) + b = 0$

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

$1 = \frac{\gamma w^T w}{\|w\|} = \gamma \|w\|$

$\boxed{\gamma = \frac{1}{\|w\|}}$

For the negative side also: $\gamma = \frac{1}{\|w\|}$



$\boxed{\text{Margin} = \frac{2}{\|w\|}}$

Maximize Margin $\frac{2}{\|w\|}$

or

Minimize $\frac{1}{2}\|w\| \equiv$ Minimize $\frac{\|w\|^2}{2} = \min_w \frac{w^T w}{2}$

$$\max_{x,y} \quad f(x,y) = 2 - x^2 - 2y^2$$

$$\left.\begin{aligned} x &= 0 \\ y &= 0 \end{aligned}\right\}$$

$$\max_{x,y} \quad f(x,y) = 2 - x^2 - 2y^2$$

$$\text{s.t.} \quad h(x,y) = x + y - 1 = 0$$

Lagrangian Multipliers

$$L(x,y) = \min_{x,y} f(x,y)$$

$$h(x,y) = 0$$

$$L(x, \min_{x,y,\beta} f(x,y) + \beta\, h(x,y)$$

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

$$\text{s.t.} \quad x + y - 1 = 0$$

$$L \quad (2 - x^2 - 2y^2) + \beta(x + y - 1)$$

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

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

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

$$\beta = 2x$$

$$x + y = 1 \qquad x = 1 - y$$

$$\beta = 2(1 - y)$$

$$\beta = 4y$$

$$2(1 - y) = 4y$$

$$\boxed{\begin{aligned} y &= \tfrac{1}{3} \\ x &= \tfrac{2}{3} \end{aligned}}$$

$$\min_{x,y} \; f(x,y) = x^3 + y^2$$

$$\text{s.t. } g(x) \quad x^2 - 1 \le 0$$

$$\boxed{\min_{x,y,\alpha} \; x^3 + y^2 + \alpha(x^2-1) \qquad \text{s.t. } \alpha \ge 0}$$

log-barrier method

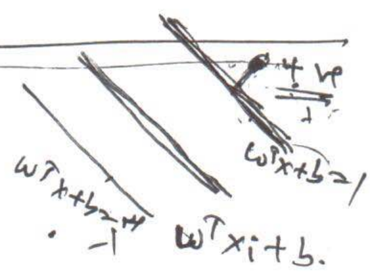
$\cancel{\alpha} = \exp(\tilde{\alpha})$

or $\cancel{\alpha = \log(\tilde{\alpha})}$

$\tilde{\alpha} = \log(\alpha)$

$$\boxed{x^3 + y^2 + \exp(\tilde{\alpha})\,(x^2-1)}$$

SVM



$$\boxed{\min \; \frac{\|w\|^2}{2} \qquad \text{s.t. } \{1 - y_i(w^Tx_i+b) \le 0\} \; \forall i}$$

$$\min_{w} \left[\max_{\substack{\alpha,\beta \\ \alpha_i \ge 0}} L(w,\alpha,\beta) \right] \longrightarrow \theta_p \qquad L(w)$$

Primal formulation

$y_i(w^Tx_i+b) \ge 1$

$$\min_{x,y,z} f(x,y,z) = x^2 + 4y^2 + 2z^2 + 6y + z$$

$$h_1(x,y,z) : \quad x + z^2 - 1 = 0$$

$$h_2(x,y,z) : \quad x^2 + y^2 - 1 = 0$$

$$L = \min_{x,y,z,\beta_1,\beta_2} \left(x^2 + 4y^2 + 2z^2 + 6y + z\right) + \beta_1 (x + z^2 - 1) + \beta_2 (x^2 + y^2 - 1)$$

$$\frac{\partial L}{\partial x} : \quad 2x + \beta_1 + 2\beta_2 x = 0$$

$$\frac{\partial L}{\partial y} : \quad 8y + 6 + 2\beta_2 y = 0$$

$$\frac{\partial L}{\partial z} : \quad 4z + 1 + 2\beta_1 z = 0$$

$$\frac{\partial L}{\partial \beta_1} : \quad x + z^2 - 1 = 0$$

$$\frac{\partial L}{\partial \beta_2} : \quad x^2 + y^2 - 1 = 0$$

Solve for $x, y, z, \beta_1, \beta_2$

$xy + yz + xz$

Quadratic.

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

s.t. $\alpha_i \geq 0$

Minimize L_p w.r.t w, b

$$\frac{\partial L_p}{\partial w} = \cancel{2}w + \sum_{i=1}^{n} \alpha_i \left(-y_i(x_i)\right)$$

$$= \cancel{2}w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0$$

$$\boxed{w = \sum_{i=1}^{n} \alpha_i y_i x_i}$$

$$\frac{\partial L_p}{\partial b} = \boxed{\sum_{i=1}^{n} \alpha_i y_i = 0}$$

$\|w\|^2 = w^T w$

$$\frac{\partial \|w\|^2}{\partial w} = 2w$$

$$\frac{\partial}{\partial w} \underline{\alpha_i (1 - y_i (w^T x_i + b))}$$

$$\frac{\partial}{\partial w}\left(\alpha_i - \alpha_i y_i w^T x_i - \alpha_i y_i b\right)$$

$$= (-\alpha_i y_i x_i)$$

$$\boxed{\max_{\alpha, \beta, \alpha \geq 0} \left[\min_{w} L(w, \alpha, \beta) \right]} \rightarrow \theta_d$$

Dual famulation

$$\boxed{\theta_p \neq \theta_d}$$

Duality Gap. $\left\{ \begin{array}{l} \theta_d \\ \theta_p \end{array} \right.$

If $L(w)$ is convex. then $\boxed{\theta_p = \theta_d}$

SVM obj: fn is convex.

KKT conditions.

If $L(w)$ satisfies KKT conditions

then

$\theta_p = \theta_d$

02/20/19

Support Vector Machines.

— Linear Classifier $(w^T x + b) \geq 0 \quad y = +1$

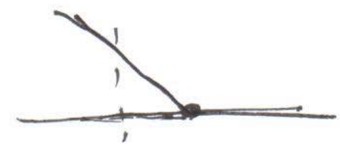
$< 0 \quad y = -1$

↳ Squared loss

logistic loss

PA1, Prob. 6 ← Hinge loss — SVM

$\max\{0, 1 - y_i(w^T x_i + b)\}$



Maximizing the margin → insurance

$$\min_{w} \frac{w^T w}{2} \equiv \frac{\|w\|^2}{2} \qquad \frac{2}{\|w\|} \rightarrow \text{margin}$$

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

→ lagrange.

$$\begin{matrix} f(x) \\ \text{s.t. } h(x) = 0 \end{matrix} \equiv f(x) + \beta h(x)$$

$$\begin{matrix} f(x) \\ g(x) \leq 0 \end{matrix} \equiv \begin{matrix} f(x) - \alpha g(x) \\ \alpha \geq 0 \end{matrix}$$

$$\boxed{\underset{w,\alpha_i \; \alpha_i \geq 0}{L(w)} = \frac{w^T w}{2} - \alpha_i \left[1 - y_i(w^T x_i + b)\right]}$$

Primal

Dual

Primal

$$\boxed{\min_{w} \left[\max_{\alpha_i ; \alpha_i \geq 0} L(w, \alpha_i) \right]} \longrightarrow w_P^*$$

$$\left[\max_{\alpha_i} \left[\min_{w} L(w, \alpha_i) \right] \right] \longrightarrow w_D^*$$

$$\boxed{\begin{array}{l} \|w\|^2 \\ = w^T w \\ \frac{d\, w^T w}{dw} \\ = 2w \end{array}}$$

$w^*_D = w^*_P$ if $L(w)$ is convex

$$L(w) \quad \frac{\|w\|^2}{2} + \sum \alpha_i \left(1 - y_i(w^T x_i + b)\right)$$

$$\boxed{\frac{\partial}{\partial w} w^T w = 2w}$$

$$\frac{\partial L}{\partial w} = w + \sum \alpha_i (-y_i) x_i$$

$$= w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0$$

$$\boxed{w = \sum_{i=1}^{n} \alpha_i y_i x_i}$$

$$\frac{\partial L}{\partial b} = -\sum \alpha_i y_i = 0 \qquad \boxed{\sum_{i=1}^{n} \alpha_i y_i = 0}$$

$$\max_{\alpha_i} \quad \frac{w^T w}{2} + \sum \alpha_i \left(1 - y_i (w^T x_i + b)\right)$$

$$= \left(\sum \alpha_i y_i \underline{x_i}\right)^T \left(\sum \alpha_i y_i x_i\right)$$

$$+ \Big[\underline{\sum \alpha_i} - \sum \alpha_i y_i \left(\sum \alpha_i y_i x_i\right)^T$$

$$- \sum \alpha_i y_i b \Big]$$

$$\max_{\alpha_i} \left[\sum \alpha_i - \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_j \alpha_k y_j y_k (x_j^T x_k)\right]$$

Dual optimization problem is a function of α_i's

Solve the dual problem. → offload it to a solver.

↳ get α_i's that maximize it.

use these to get w and b.

---

for any lagrangian

$$\min_{w, \alpha_i} \boxed{L(w) - \sum \alpha_i g_i(w)} \quad \text{s.t.} \quad \alpha_i \geq 0$$

$\underline{w, \alpha_i}$ How do we know ~~that~~ that is the right answer? $\underline{\text{KKT}}$

KKT #1 $\frac{\partial}{\partial w} L(w) = 0$

~~KKT #2 $\frac{\partial}{\partial b} L(w) = 0$~~

KKT #2 $g_i(w) \leq 0$

KKT #3 $\alpha_i \geq 0$

KKT #4 $\boxed{\alpha_i g_i(w_i) = 0}$

KKT #5 $\boxed{\frac{d}{d\alpha_i} L(w) = 0}$

$w - \sum \alpha_i y_i x_i = 0$

$\sum \alpha_i y_i = 0$

$\forall i \quad 1 - y_i(w^T x_i + b) \leq 0$

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

$\forall i \quad \alpha_i(1 - y_i(w^T x_i + b)) = 0$

If I have $\{\alpha_i\}_{i=1}^{n}$

$w = \sum \alpha_i y_i x_i$

$\alpha_i(1 - y_i(w^T x_i + b)) = 0$

Solve for b

$\alpha_i y_i (w^T x_i + b) = \alpha_i$

$b = \boxed{\quad}$

In practice you will get n b's.

Take an average to get b.

$\boxed{w, b}$

let n = 10

α_1

α_2

α_3

$\vdots$

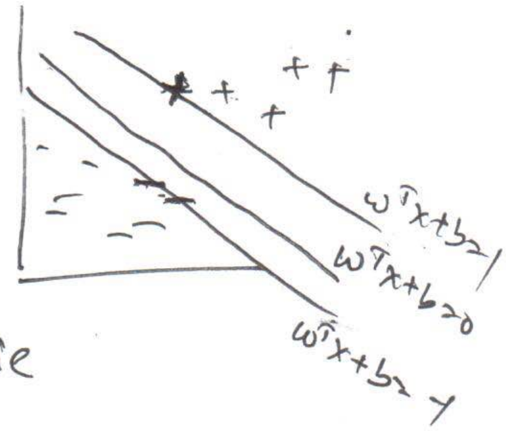
α_{10}

for point that lie on the margin

α_i will be ≥ 0

for point that do not lie on the margin

$\alpha_i = 0$



$$\alpha_i \left(1 - y_i (w^T x_i + b)\right) = 0$$

KKT # 4

for points on the margin $\quad y_i (w^T x_i + b) = 1$

for points not on the margin $\quad y_i (w^T x_i + b) > 1$

$\Rightarrow 1 - (y_i (w^T x_i + b)) < 0$

$\Rightarrow \alpha_i$ will be 0

$\rightarrow$ Known as Support vectors.

For a test point x^*

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

$$y^* = w^T x^* + b$$

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

$$= \sum \alpha_i y_i (x_i^T x^*) + b$$

o.k to ignore points for which $\alpha_i \neq 0$

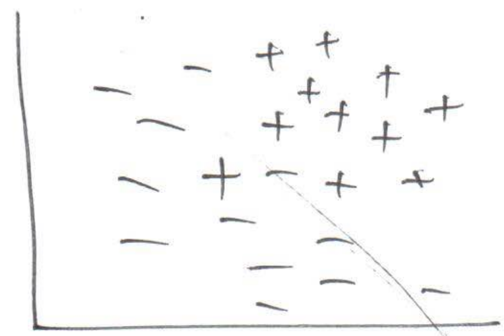
In actual ~~Sup~~ SVM implementation,
we never calculate w & b.

We simply get y^* for an x^* using α_i s.

We do not need to store α_i s and x_i and y_is if $\alpha_i = 0$

Only support vectors are needed.

What if data is not separable

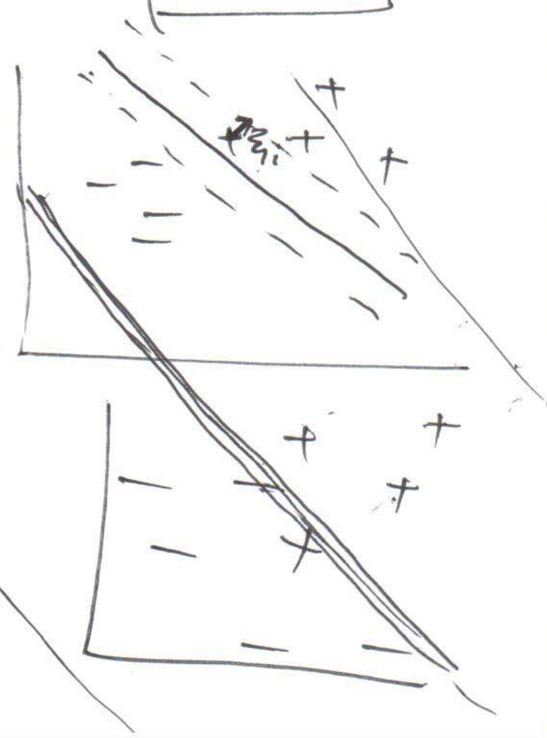


$$\boxed{y_i(w^T x_i + b) \geq 1}$$

will never be met

Slack variable

$$\boxed{y_i(w^T x_i + b) \geq 1 - \xi_i} \qquad \boxed{\xi_i \geq 0}$$



~~Max.~~
$$\min \ \frac{1}{2}\|w\|^2 + C \sum \xi_i$$

Control

s.t. $y_i(w^T x_i + b) \geq 1 - \xi_i$

and $\xi_i \geq 0$