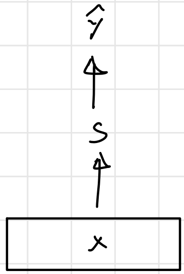
Lecture 13

- Support Vector Machine (SVM)

  - perceptron

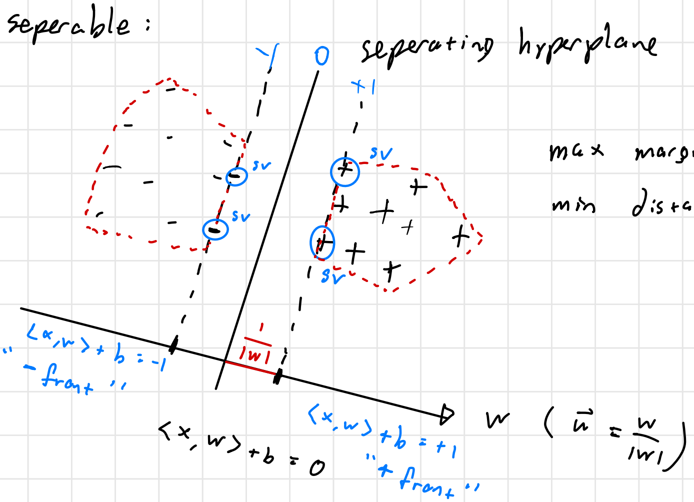


$$s = f(x) = \langle x, w \rangle + b$$

$$\hat{y} = \text{sign}(s) = \begin{cases} +1 & \text{if } s \geq 0 \\ -1 & s < 0 \end{cases}$$

- Geometric Insight : max margin

  - seperable :



max margin : Primal

min distance : Dual

- Primal $\min_{w} \frac{1}{2}|w|^2$

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

  $i = 1, \ldots, n$ primal constraints

  Quadratic Progamming

- Lagrangian:

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

$$\min_{(w,b)} \max_{\alpha=(\alpha_i \geq 0)} \mathcal{L}((w,b), \alpha)$$

dual constraints

no primal constraint

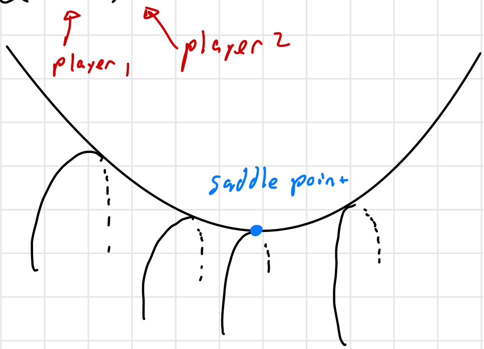
automatically satisfied

$\exists i$, $1 - y_i(\langle x_i, w\rangle + b) > 0$, then

$$\max_{\alpha_i \geq 0} \alpha_i(\ \ > 0) \longrightarrow +\infty$$

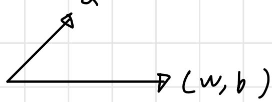
which can not be solution to $\min_{(w,b)} \max_{\alpha>0} \mathcal{L} < +\infty$

$\mathcal{L}((w,b),\alpha)$ ← The loss of player 1 is the gain of player 2.



- not a faithful plot

actually linear in $\alpha \geq 0$



concave

convex

- concave - convex :

  - Von Neumann: $\min_{(w,b)} \max_{(\alpha > 0)} \mathcal{L} = \max_{(\alpha > 0)} \min_{(w,b)} \mathcal{L}$ Solution of game

    anticipate worst case scenario

- Assume $b = 0$:

$$\mathcal{L} = \frac{1}{2}|w|^2 + \sum_{i=1}^{n} \alpha_i (1 - y_i \langle x_i, w \rangle)$$

$$= \frac{1}{2}|w|^2 - \langle \sum_{i=1}^{n} \alpha_i y_i x_i , w \rangle + \sum_{i=1}^{n} \alpha_i$$

$$\min_w : \frac{\partial}{\partial w} = 0 , \quad \boxed{\hat{w}} = \sum_{i=1}^{n} \alpha_i y_i \boxed{x_i}$$

Representer

- Note we could have solved this by realizing:

$$\mathcal{L} = \frac{1}{2} \left| w - \sum_{i=1}^{n} \alpha_i y_i x_i \right|^2 + \underbrace{\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \left| \sum_{i=1}^{n} \alpha_i y_i x_i \right|^2}_{Q(\alpha)}$$

- $Q(\alpha) = \min_{w} \mathcal{L}(w, \alpha)$

$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \left| \sum_{i=1}^{n} \alpha_i y_i x_i \right|^2$$

$$\max_{\alpha \geq 0} Q(\alpha) \quad : \text{dual problem.}$$

$$Q(\alpha) = \sum \alpha_i - \frac{1}{2} \langle \sum \alpha_i y_i x_i , \sum \alpha_j y_j x_j \rangle$$

$$= \sum \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j \underbrace{y_i y_j \langle x_i , x_j \rangle}_{Q_{ij}}$$

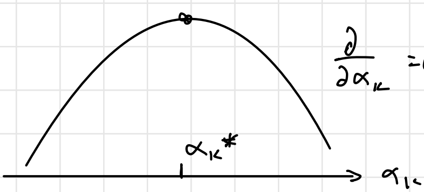
- Dual Coordinate Ascent:
  - Each iteration

for k in $1 \ldots n$

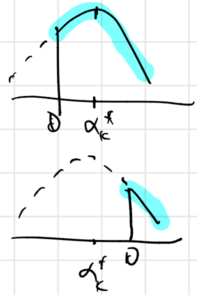
$\max_{\alpha_k \geq 0} Q$ while fixing α_i $(i \neq k)$

$$Q(\alpha_k) = \alpha_k - \frac{1}{2} \alpha_k^2 Q_{kk} - \alpha_k \sum_{i \neq k} \alpha_i Q_{ik}$$

$$= -\frac{1}{2} Q_{kk} \alpha_k^2 + \alpha_k \left(1 - \sum_{i \neq k} \alpha_i Q_{ik} \right)$$



$$\frac{\partial}{\partial \alpha_k} = 0 \qquad \alpha_k^* = \frac{1 - \sum_{i \neq k} \alpha_k Q_{ik}}{Q_{kk}}$$



- If $\alpha_k^* \geq 0$ then $\hat{\alpha}_k = \alpha_k^*$

- If $\alpha_k^* < 0$ then $\hat{\alpha}_k = 0$

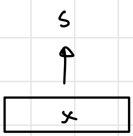
So $\hat{\alpha}_k = \max(0, \alpha_k^*)$

- Thus $\hat{w} = \sum_{i=1}^{n} \hat{\alpha}_i \, y_i \, \boxed{x_i}$

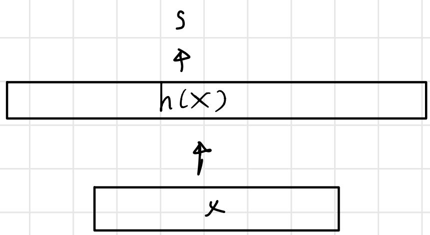
- Classifier $\hat{y} = \text{sign}(s)$

$$s = f(x) = \langle x, \hat{w} \rangle$$
$$= \sum_{i=1}^{n} \hat{\alpha}_i \, y_i \, \langle x_i, x \rangle$$

- Recall Kernel Trick:



$s = f(x) = \langle x, w \rangle$



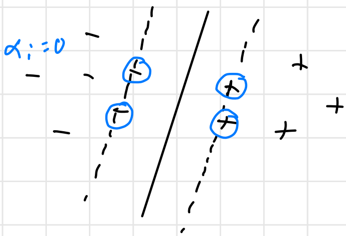
$$s = f(x) = \langle h(x), w \rangle$$
$$K(x, x') = \langle h(x), h(x') \rangle$$

learning : $Q(\alpha) = \sum \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j \overbrace{\boxed{y_i y_j K_{ij}}}^{Q_{ij}}$

Testing : $f(x) = \sum_{i=1}^{n} \alpha_i \, y_i \, K(x_i, x)$

- $$f(x) = \sum_{i:\alpha_i>0} \alpha_i y_i K(x_i, x)$$

  SV



- complementary slackness:

$$y_i(\langle x_i, w\rangle + b) > 1 \Rightarrow \max_{\alpha_i} \alpha_i = 0$$

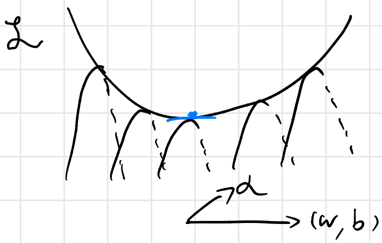
If $\alpha_i > 0 \Rightarrow y_i(\langle x_i, w\rangle + b) = 1$

$$\alpha_i(1 - y_i(\langle x_i, w\rangle + b)) = 0$$ (KKT condition)

Karush-Kuhn-Tucker

- Add back b:

$$\mathcal{L} = \mathcal{L}_{old} - \sum \alpha_i y_i b$$



stationary (KKT condition)

$$\frac{\partial \alpha}{\partial w} = 0$$

$$\frac{\partial \alpha}{\partial b} = 0$$

So $\sum \alpha_i y_i = 0$.

Otherwise if $\sum \alpha_i y_i > 0$

$$\min_{b \to -\infty} \longrightarrow -\infty$$

$$\sum \alpha_i y_i < 0$$

$$\min_{b \to \infty} \longrightarrow -\infty$$

Can not be solution to $\max_{\alpha \geq 0} \min_{(w,b)} \mathcal{L} > -\infty$

- Dual Optimization:

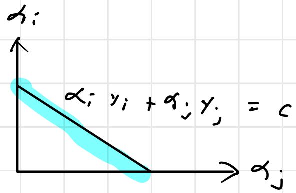
$$\max_{\substack{\alpha \geq 0 \\ \sum_{i=1}^{n} \alpha_i y_i = 0}} Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \left\| \sum_{i=1}^{n} \alpha_i y_i \boxed{x_i} \right\|^2$$

- minimal sequential optimization (MSO)
  each step:

$$\max_{\alpha_i \geq 0 \ \alpha_j \geq 0} Q(\alpha)$$

$$\alpha_i y_i + \alpha_j y_j = - \sum_{k \neq i,j} \alpha_k y_k = c$$

(α_k fixed)



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

$$\sum_{i: y_i = +} \alpha_i + \sum_{i: y_i = -} (-\alpha_i) = 0$$

$$\sum_{i: y_i = +} \alpha_i = \sum_{i: y_i = -} \alpha_i = s$$

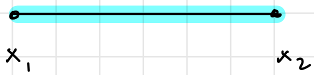
$$\text{So } Q(\alpha) = 2s - \frac{1}{2} s^2 \left\| \sum_{i:+} \overset{c_i}{\frac{\alpha_i}{s}} \boxed{x_i} - \sum_{i:-} \overset{c_i}{\frac{\alpha_i}{s}} \boxed{x_i} \right\|^2$$

$$\min \left| \sum_{i:+} c_i x_i - \sum_{i:-} \alpha_i x_i \right|^2$$

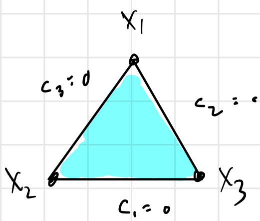
min distance between $\pm$ convex hulls

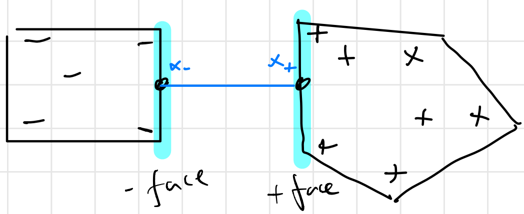
$$\sum_{i:+} c_i = 1 \quad , \quad \sum_{i:-} c_i = 1$$

$$\{c_1 x_1 + c_2 x_2 : c_1 + c_2 = 1, \; c_1 \geq 0, \; c_2 \geq 0\}$$



$$\{c_1 x_1 + c_2 x_2 + c_3 x_3 : c_1 + c_2 + c_3 = 1 \}$$
$$c_1 \geq 0 \quad c_2 \geq 0 \quad c_3 \geq 0$$



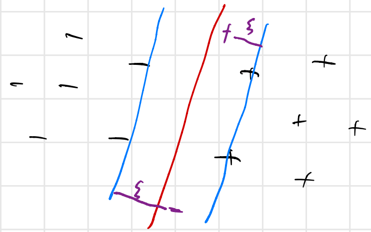


convex hulls

$$\hat{w} \propto x_+ - x_-$$

$$\Big(\hat{w} = \sum \alpha_i y_i x_i$$
$$= \sum_{i \in +} \alpha_i x_i - \sum_{i \in -} \alpha_i x_i$$
$$= s\Big(\sum_{i \in +} c_i x_i - \sum_{i \in -} c_i x_i\Big)$$
$$= s(x_+ - x_-) \Big)$$

- Non-Sepenable:



$$\min_{w,b,\xi} \frac{1}{2}|w|^2 + C\sum \xi_i$$

s.t.

$$y_i(\langle x_i, w\rangle + b) \geq 1 - \xi_i \qquad \text{slackness - penalty}$$

$$\xi_i \geq 0, \quad i = 1, \dots, n$$

- Lagrangian:

$$\mathcal{L}((w,b,\xi),(\alpha,\mu)) = \frac{1}{2}|w|^2 + C\sum \xi_i + \sum \alpha_i(1 - \xi_i - y_i(\langle x_i, w\rangle + b)) + \sum \mu_i(-\xi_i)$$

$$= \underbrace{\frac{1}{2}|w|^2 + \sum \alpha_i(1 - y_i(\langle x_i, w\rangle + b))}_{\mathcal{L}_{old}\ (\text{with } b)} + \underbrace{C\sum_{i=1}^{n} \xi_i - \sum \alpha_i \xi_i - \sum \mu_i \xi_i}_{-\sum_{i=1}^{n}(C - \alpha_i - \mu_i)\xi_i}$$

Set $\frac{\partial \mathcal{L}}{\partial \xi_i} = 0$ so $C - \alpha_i - \mu_i = 0$

$$\alpha_i = C - \mu_i \leq C$$

otherwise, if $C - \alpha_i - \mu_i > 0$

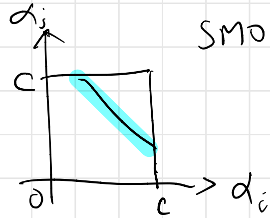
$$\min_{\xi_i \to \infty} -(C - \alpha_i - \mu_i)\xi_i \to -\infty$$

if $C - \alpha_i - \mu_i < 0$

$$\min_{\xi_i \to -\infty} -(C - \alpha_i - \mu_i)\xi_i \to -\infty$$

can not be solution to $\max_{(\alpha,\mu) \geq 0} \min_{(w,b,\xi)} \mathcal{L}$

SMO



- Dual $\max_{\alpha_i \in [0, c]} Q(\alpha)$

box constraint

$\sum \alpha_i y_i = 0$ ← caused by b

affinity

- $f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x)$

query

SV Value key

follow important friends

memorization

\+ nearest neighbors interpolation