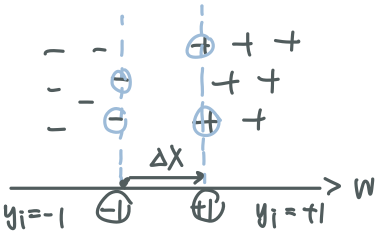
support vector machine (SVM)

primal dual



$y_i = -1$ $\qquad$ $y_i = +1$

$\langle x_i, w\rangle + b \leq -1$ $\qquad$ $\langle x_i, w\rangle + b \geq 1$

infinitely many w & b that gives separation

can rescale w + shape b to get -1 & $+1$

$\langle \Delta x, w\rangle = 2$

$|\Delta x||w|\cos\theta = |\Delta x|\,|w| = 2$

$\downarrow$

$\theta = 0$

$|\Delta x| = \frac{2}{w}$ margin

primal problem: max margin

$$\boxed{\begin{array}{l} \min_{(w,b)} \frac{1}{2}|w|^2 \\ \text{subject to constraints} \\ \quad y_i(\langle x_i, w\rangle + b) \geq 1,\ i = 1, \dots, n \end{array}}$$

simplify $b = 0$

$\min \frac{1}{2}|w|^2$

subject to

$y_i\langle x_i, w\rangle \geq 1 \quad i = 1, \dots, n$

Lagrangian

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

$$\mathcal{L} = (\alpha_i \geq 0,\ i = 1, \dots, n)$$

$$\underset{(\text{constrained})}{\text{primal}} \xrightarrow{\text{translate}} \underset{(\text{unconstrained})}{\min\max \mathcal{L}}$$

$\downarrow$ solution

satisfies all constraints automatically

proof by contradiction

If $\exists$, $1 - y_i\langle x_i, w\rangle > 0$

$$\max_{\alpha_i \geq 0} \alpha_i(1 - y_i\langle x_i, w\rangle) = 0$$

cannot be the solution

At the solution, all constraints are satisfied

$$1 - y_i\langle x_i, w\rangle \leq 0 \qquad \forall i$$

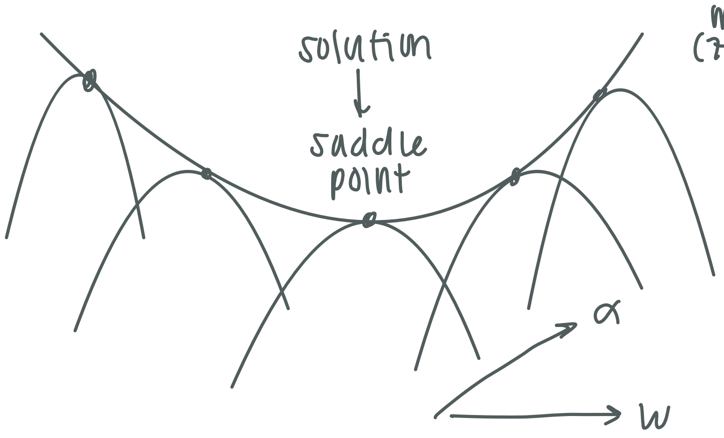
$$\max_{\alpha_i \geq 0} \alpha_i(1 - y_i\langle x_i, w\rangle) = 0$$

complementary slackness:

$$1 - y_i\langle x_i, w\rangle < 0 \Rightarrow \alpha_i = 0$$

$$\alpha_i > 0 \Rightarrow 1 - y_i\langle x_i, w\rangle = 0 \quad (\text{S.V.'s})$$

$\underset{\text{(constrained)}}{\text{primal}} \xrightarrow{\text{translate}} \underset{\text{(unconstrained)}}{\min_{w} \max_{\alpha \geq 0} \mathcal{L}}$



concave − convex

$\overset{\downarrow}{\alpha} \qquad \overset{\downarrow}{w}$

min max gain (zero sum) $\Big\Downarrow$ translate

$$\max_{\alpha \geq 0} \boxed{\min_{w} \mathcal{L}}$$

$\Big\Downarrow$ solution

$\max_{\alpha \geq 0} Q(\alpha)$ dual problem

min w $\mathcal{L}$:

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

$$\frac{\partial}{\partial w}\mathcal{L} = 0$$

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

$$|a-b|^2 = \langle a-b, a-b \rangle = \langle a,a \rangle + \langle b,b \rangle = 2\langle a,b \rangle$$
$$= |a|^2 + |b|^2 = 2\langle a,b \rangle$$

Representer:

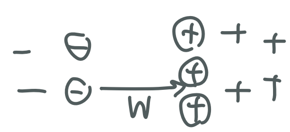
$$\hat{w} = \sum_{i=1}^{n} \alpha_i y_i \boxed{x_i} = \sum_{\alpha_i > 0} \alpha_i y_i \boxed{x_i} = \sum_{\substack{\alpha_i > 0 \\ y_i = +1}} \alpha_i \boxed{x_i} - \sum_{\substack{\alpha_i > 0 \\ y_i = -1}} \alpha_i \boxed{x_i}$$

Dual function

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

Dual problem: $\max_{\alpha_i \geq 0} Q(\alpha)$

w points from NSV to PSV

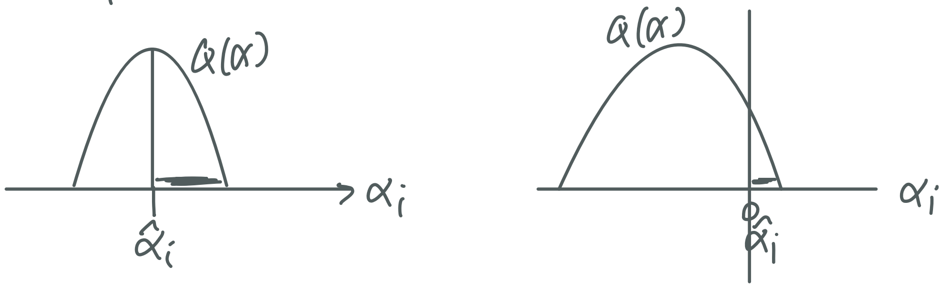


coordinate ascent

each iteration

for i in 1 to n

$\min_{\alpha_i \geq 0} Q(\alpha)$ fix all α_j ($j \neq 1$) at current value



Kernel trick

classification of a training example x

$\hat{y} = \text{sign}(\langle x, \hat{w} \rangle)$

$\underline{\underline{\text{Representer}}}\ \text{sign}\left(\left\langle \sum_{i=1}^{n} \alpha_i y_i \boxed{x_i}, \boxed{x} \right\rangle\right)$

$= \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i \langle x_i, x \rangle \right)$

$\langle \phi(x_i), \phi(x) \rangle = K(x_i, x)$ kernel function

Dual

$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \left\langle \sum_{i=1}^{n} \alpha_i y_i \boxed{x_i}, \sum_{j=1}^{n} \alpha_j y_j \boxed{x_j} \right\rangle$

$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$

$\Big\Downarrow \max_{\alpha \geq 0}$

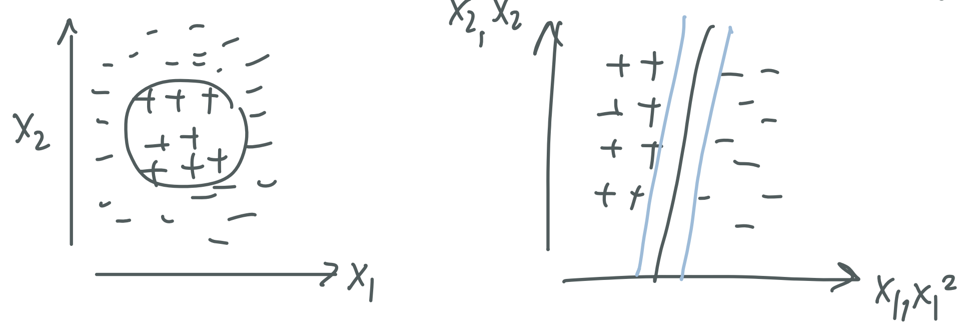
α

$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$

$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$

nonseparable case

$w_1 X_1 + w_2 X_1^2 + w_3 X_2 + w_4 X_2^2 = 0$



x-space $\longrightarrow$ $\phi(x)$ space

$K(x,x') = \exp(-\gamma |x - x'|^2)$ Radial basis kernel (Gaussian kernel)

Decision rule

$\hat{y} = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i \underbrace{K(x_i, x)}_{\text{similarity}}\right)$

nearest neighbor matching

back to linear, add back "b"

$\langle x_i, w \rangle + b$

$\max\limits_{\alpha} \min\limits_{\langle w, b \rangle} \mathcal{L}$

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

$= \frac{1}{2}|w|^2 + \sum_{i=1}^{n} \alpha_i - \left\langle \sum_{i=1}^{n} \alpha_i y_i x_i, w \right\rangle + \overset{0}{\boxed{\sum_{i=1}^{n} \alpha_i y_i}}\, b$

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

if >0, $\min\limits_{b} = -\infty$

if <0 $\min\limits_{b} = -\infty$

cannot be max

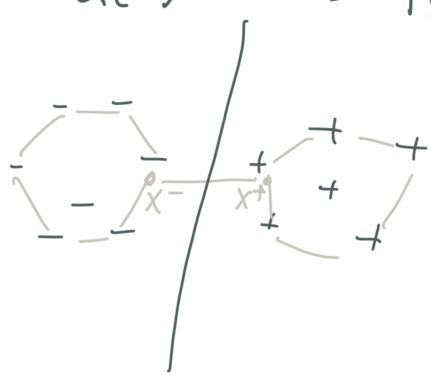
$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\left|\sum_{i=1}^{n} \alpha_i y_i x_i\right|^2$

$\max\limits_{\substack{\alpha_i \geq 0 \;\forall i \\ \sum_{i=1}^{n} \alpha_i y_i = 0}} Q(\alpha)$

$\sum_{y_i=+1} \alpha_i - \sum_{y_i=-1} \alpha_i = s$

Dual problem:

$Q(\alpha) = 2s - \frac{1}{2}s^2 \left| \sum_{y_i=+1} \underbrace{\frac{\alpha_i}{s}}_{c_i \geq 0} x_i - \sum_{y_i=-1} \underbrace{\frac{\alpha_i}{s}}_{c_i \geq 0} x_i \right|^2$



$\downarrow$ min distance = max margin