SVMs

margin $\min_i \frac{y_i \overbrace{f_{w,b}(x_i)}^{w^T x_i + b}}{\|w\|}$

$$\max_{w,b}\left(\min_i \frac{y_i f_{w,b}(x_i)}{\|w\|}\right)$$ ← maximize minimum distance (margin)

let w & b be scaled so that

$$y_i(w^T x_i + b) \geq 1$$ ← equality for at least two points

just minimize $\frac{1}{2}\|w\|^2$

subject to $y_i(w^T x_i + b) \geq 1 \quad \forall i$

Lagrange

Aside

we have a function $f(x,y)$ for which we want to find max or min

$$\frac{\partial}{\partial x \partial y} f(x,y) = 0$$ ← solve for x & y & check

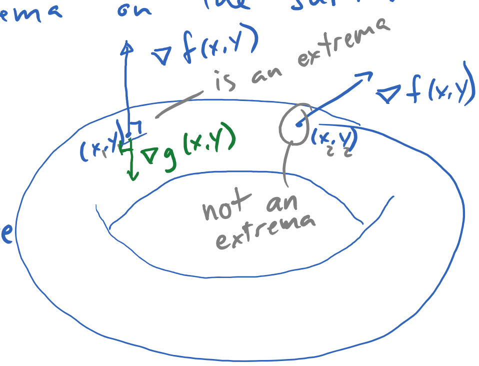
if there is a constraint on x, y it is harder

e.g. max $f(x,y)$ subject to $\hat{g}(x,y) = c$

we can only search for extrema on the surface defined by $\hat{g}(x,y) = c$

Define $g(x,y) = \hat{g}(x,y) - c = 0$

$\nabla g(x,y)$ is $\perp$ to the surface



at an extrema point

$$\nabla f(x,y) = -\lambda \nabla g(x,y)$$

$$\boxed{L(x,y,\lambda) \equiv f(x,y) + \lambda g(x,y)}$$ ← Lagrangian function

$$\boxed{L(x,y,\lambda) \equiv f(x,y) + \lambda g(x,y)}$$

Lagrangian function

$$\nabla L(x,y,\lambda) = 0$$

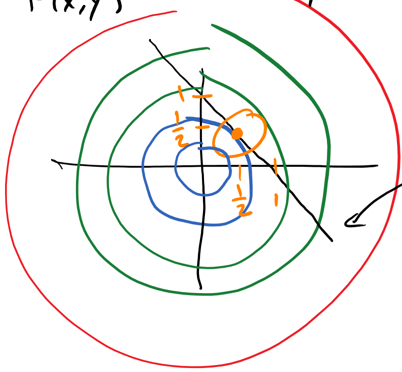
$$\nabla_x f(x,y) + \lambda \nabla_x g(x,y) = 0$$
$$\nabla_y f(x,y) + \lambda \nabla_y g(x,y) = 0$$
$$\nabla_\lambda L(x,y,\lambda) = g(x,y) = 0$$

solving this yields our optimum point

Example

$f(x,y) = \cancel{1} - x^2 - y^2$

constraint is $g(x,y) = x + y - 1 = 0$



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

$-2x + \lambda = 0 \Rightarrow \lambda = 2x$

$-2y + \lambda = 0 \Rightarrow -2y + 2x = 0 \Rightarrow x = y$

$x + y - 1 = 0$

$x + x - 1 = 0$

$x = 1/2 \quad y = 1/2$

What if the constraint is an inequality?

e.g. $g(x,y) \geq 0$

Two possible situations

① $g(x,y) > 0$ ← stationary in the region

our constraint is inactive

$\lambda = 0$

or ② $g(x,y) = 0$ ← we are on the surface $g(x,y) = 0$

either case

$$\boxed{\begin{aligned} &\lambda g(x,y) = 0 \\ &\lambda \geq 0 \\ &g(x,y) \geq 0 \end{aligned}}$$

KKT conditions

$$\boxed{g(x,y) \geq 0}$$

$$L(w,b,\alpha) = \frac{1}{2}\underbrace{\|w\|^2}_{w^T w} - \sum_{n=1}^{N} \alpha_n \{ \underbrace{y_n (w^T x_n + b) - 1}_{g_n(x)} \}$$

$$\nabla_w L = 0 \quad \Rightarrow \quad \boxed{w = \sum_{n=1}^{N} \alpha_n y_n x_n} \quad 1$$

$$\nabla_b L = 0 \quad \Rightarrow \quad \boxed{\sum_{n=1}^{N} \alpha_n y_n = 0} \quad 2$$

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

$$* \quad \underset{\lambda_i}{\alpha_i} [\underbrace{y_i (w^T x_i + b) - 1}_{g_i(x)}] = 0$$

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

if $\alpha_i = 0$ the constraint is considered inactive so we just choose the set of data (x_i's) for which $\alpha_i \neq 0$ and $y_i (w^T x_i + b) = 1$

This set is the "support vectors"

rewriting $L(w,b,\alpha)$

$$L(w,b,\alpha) = \frac{1}{2} \|w\|_2^2 - \sum_{n=1}^{N} \alpha_n \left(y_n (w^T x_n + b) - 1 \right)$$

$$= \frac{1}{2} w^T w - w^T \underbrace{\cancel{\sum_{n=1}^{N} \alpha_n y_n x_n}}_{w \text{ by } 1} - b \underbrace{\cancel{\sum_{n=1}^{N} \alpha_n y_n}}_{0 \text{ by } 2} + \sum_{n=1}^{N} \alpha_n$$

$$= \sum_{n=1}^{N} \alpha_n - \frac{1}{2} w^T w \qquad \text{by } 1 \quad w = \sum_{n=1}^{N} \alpha_n y_n x_n$$

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

$$\boxed{\begin{array}{l} L(\alpha) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m x_n^T x_m \\ \text{subject to} \quad \alpha_n \geq 0 \quad \& \quad \sum_{n=1}^{N} \alpha_n y_n = 0 \end{array}}$$

$$\nabla_\alpha L(\alpha) = 0$$

$$y_n (w^T x_n + b) - 1 \geq 0$$

$$\alpha_n \left(y_n (w^T x_n + b) - 1 \right) = 0 \qquad \text{from KKT}$$

implying $\alpha_n = 0$ or $y_n (w^T x_n + b) = 1$

How to find b?

we know that for any support vector

$$y_n (w^T x_n + b) = 1$$

$$w = \sum_n \alpha_n y_n x_n$$

recall that $\alpha_n = 0$ for all but the support vectors

$$b = y_n - w^T x_n \quad \text{for } n \in \text{support set}$$

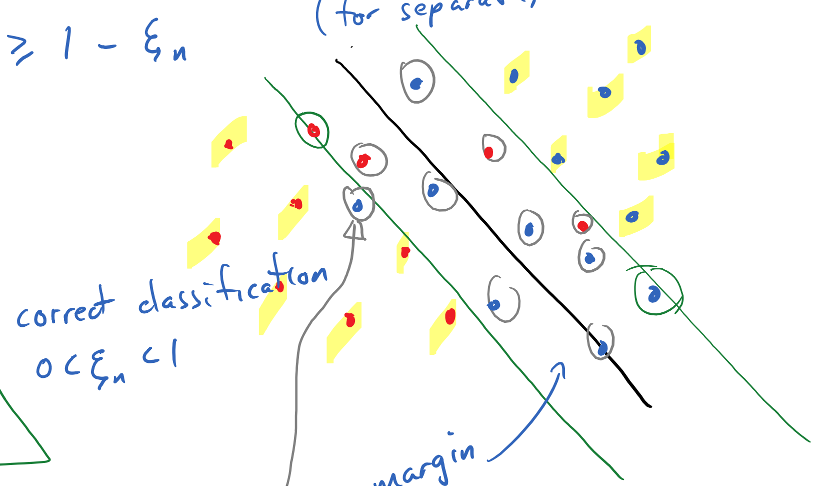
in practice $b = \frac{1}{m} \sum_{k \in S} (y_k - w^T x_k)$

What if the data is not separable?

$$y_n (w^T x_n + b) \geq 1 - \xi_n$$

(for separable, $\xi_n = 0$)

$$\boxed{\begin{array}{l} \text{we now minimize} \\ \frac{1}{2} \|w\|_2^2 + C \sum_{n=1}^{N} \xi_n \\ \xi_n \geq 0 \\ y_n (w^T x_n + b) \geq 1 - \xi_n \end{array}}$$



$$y_n(w^T x_n + b) \geq 1 - \xi_n$$

$\xi_n = 0$ for vectors beyond on or argin boundaries

exception

margin support vectors are inside on on the margin

$$L(w, b, \alpha, \mu, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{n=1}^{N} \xi_n - \sum_{n=1}^{N} \alpha_n\left(y_n(w^T x_n + b) - 1 + \xi_n\right) - \sum_{n=1}^{N} \mu_n \xi_n$$

KKT $\Rightarrow$

$$\alpha_n \geq 0, \quad \mu_n \geq 0$$

$$y_n(w^T x_n + b) - 1 + \xi_n \geq 0$$

$$\alpha_n\left(y_n(w^T x_n + b) - 1 + \xi_n\right) = 0$$

$$\xi_n \geq 0$$

$$\mu_n \xi_n = 0$$

as before

$$\nabla_w L = 0 \Rightarrow w = \sum_{n=1}^{N} \alpha_n y_n x_n$$

$$\nabla_b L = 0 \Rightarrow \sum_{n=1}^{N} \alpha_n y_n = 0$$

$$\nabla_{\xi_n} L = 0 \Rightarrow \boxed{\alpha_n = C - \mu_n}$$

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

but

$$\alpha_n \geq 0, \quad \mu_n \geq 0$$

$$\alpha_n \leq C$$

$$0 \leq \alpha_n \leq C \quad \leftarrow \text{box condition}$$

$$\sum_{n=1}^{N} \alpha_n y_n = 0$$

$N \quad N \quad x^T x$

$$\tilde{L}(\alpha) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \underbrace{x_n^T x_m}$$

the only place x's appear

If we substituted $\phi(x)$ for x the we have $\boxed{\phi^T(x_n)\,\phi(x_m)}$

$$\Downarrow$$

$$K(x_n, x_m)$$

$$\boxed{\tilde{L}(\alpha) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m k(x_n, x_m)}$$

The kernel trick

Support Vector Machines concluded

Given $k(x_i, x_j)$, we can write the classifier as

$$\boxed{\hat{h}(x) = \text{sign}\left(\sum_i \alpha_i^* y_i k(x, x_i) + b^* \right)}$$

where α^* is the solution to

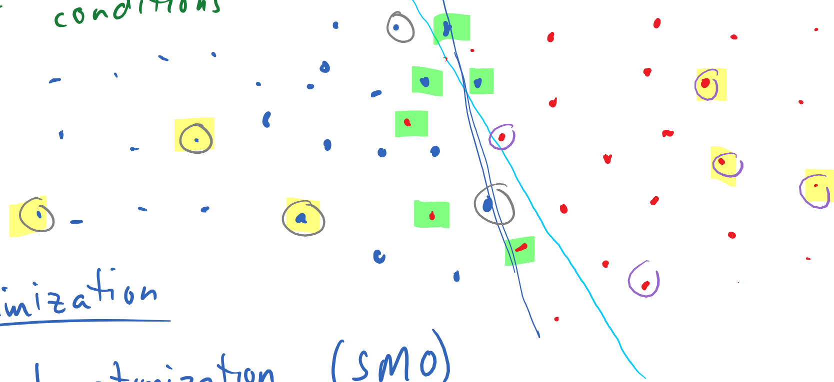
$$\max_{\alpha} \; -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i, x_j) + \sum_i \alpha_i$$

$$\text{s.t.} \quad \sum_i \alpha_i y_i = 0, \qquad 0 \le \alpha_i \le \frac{c}{n} \quad i = 1, \dots, n$$

and $b^* = y_i - \sum_j \alpha_j^* y_j k(x_i, x_j)$ for some i s.t. $0 < \alpha_i^* < \frac{c}{n}$

Training SVMs

- start with an arbitrary data subset (working set)
  ↳ perform optimization
- The support vectors stay in the working but the others are replaced by new vectors that severely violate the KKT conditions
- Repeat



Solving the optimization

Sequential minimal optimization (SMO)

$$\max_{\alpha} \; -\frac{1}{2}\sum_{i,j} \alpha_i \alpha_j q_{ij} + \sum_i \alpha_i \qquad \text{where } q_{ij} \equiv y_i y_j k(x_i, x_j)$$

$$\text{s.t.} \quad \sum_i \alpha_i y_i = 0, \qquad 0 \le \alpha_i \le \frac{C}{n} \qquad i = 1, \dots, n$$

initialize

$\alpha = 0$

repeat

(1) select a pair i, j $\quad 1 \le i, j \le n$

(2) update α_i and α_j by optimizing the dual QP holding all other α_k $k \ne i, j$ fixed

update step

choose α_i, α_j

$$\max_{\alpha_i, \alpha_j} \; -\frac{1}{2}\left(\alpha_i^2 q_{ii} + \alpha_j^2 q_{jj} + 2\alpha_i \alpha_j q_{ij}\right) + c_i \alpha_i + c_j \alpha_j$$

(where $c_i = 1 - \frac{1}{2} \sum_{k \ne i,j} \alpha_k q_{ik}$ and likewise for c_j)

$$\text{s.t.} \quad \alpha_i y_i + \alpha_j y_j = -\sum_{k \ne i,j} \alpha_k y_k$$

$$0 \le \alpha_i, \alpha_j \le \frac{C}{n}$$

running time $\Rightarrow O(n^3)$ but in practice it is usually $O(n^2)$