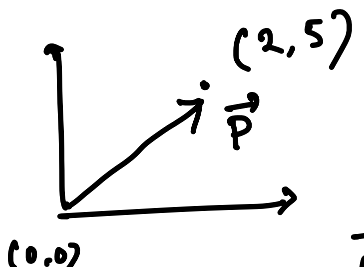
Sep. 19

Support Vector Machines (SVM)
Support Vector Networks

Vladimir Vapnik (1995 – 1999)

- linear classification
- Kernel tricks.
- is a part of max - margin models

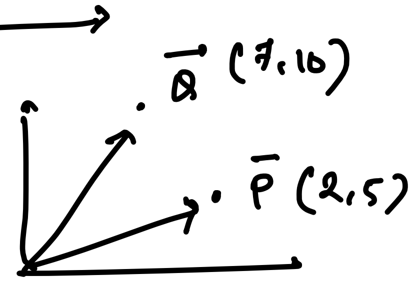
Coordinate Geometry



magnitude = $\sqrt{2^2+5^2} = \sqrt{29}$

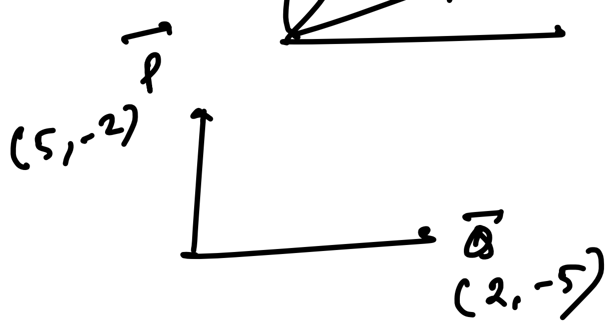
Angle / direction = $\tan\theta = y/x$

$\theta = \tan^{-1}(y/x)$



$\vec{PQ} = ((7-2),\ (10,5))$

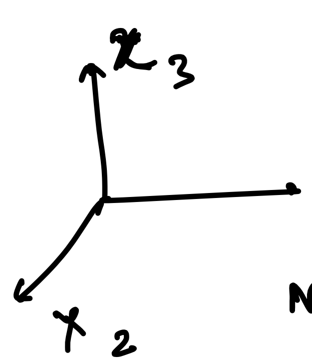
$= (5,5)$



$\bar{P}\cdot\bar{Q} = 0$ perpendi

$= 10 - 10 = 0$

Plane



Eq. of a plane:

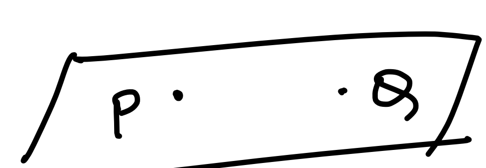
$$w_1x_1 + w_2x_2 + w_3x_3 + w_0 = 0$$

$$w^Tx + w_0 = 0$$

Normal vector of the plane:

(w_1, w_2, w_3)

$w^Tx + w_0 = 0$



$w^Tp + w_0 = 0$ $\quad w^TQ + w_0 = 0$

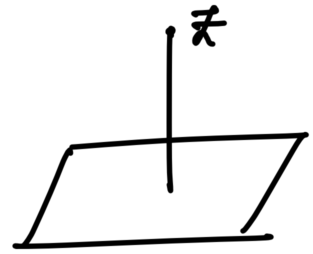
$w^Tp - w^TQ = 0$

$w^T(\vec{P} - \vec{Q}) = 0$

distance of a point from the plane

$\underline{\hspace{6em}}$

z



$$\frac{w^T z + w_0}{\|w\|} = d = \gamma$$

= geometric margin

Functional margin $= w^T z + w_0 = \hat{\gamma}$

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

$$d = \begin{cases} +ve : & \text{if } z \text{ is in the direction of the normal vector} \\ -ve : & \text{otherwise,} \end{cases}$$

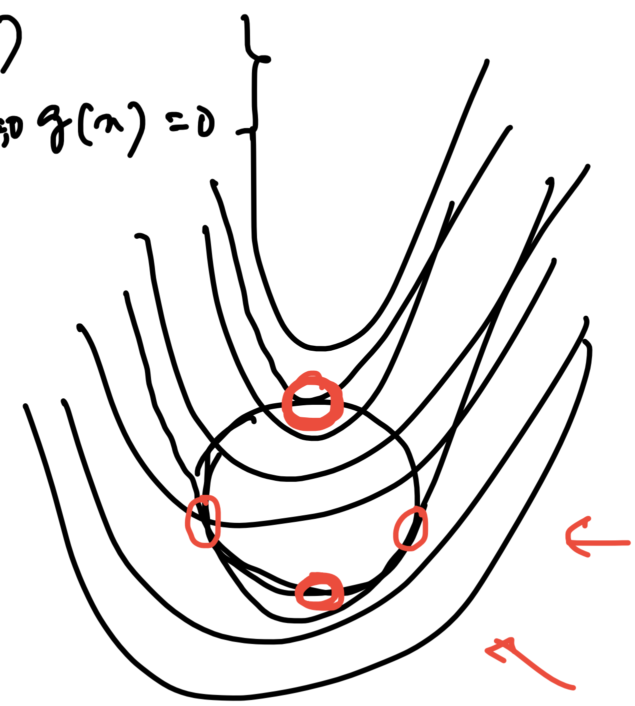
Lagrange Multiplier

Linear Regression : Sq loss

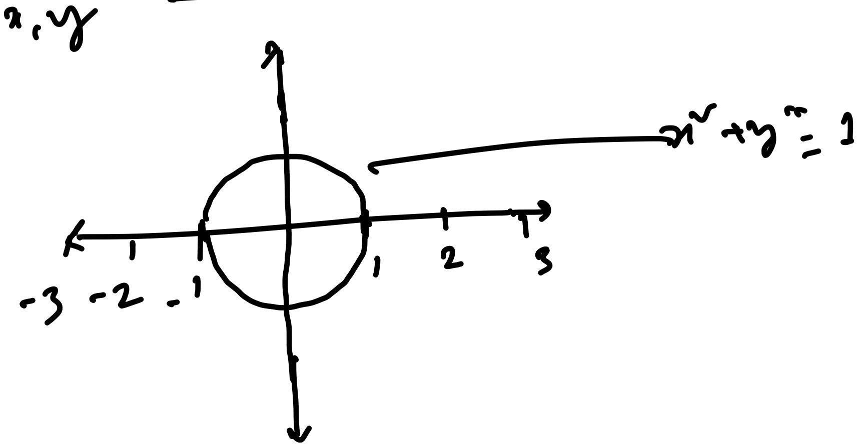
logistic " : log loss

max $f(x)$

$h(x) \leq 0$, $g(x) = 0$



E.g. $\max_{x,y} \; z = x^2 y$ s.t. $x^2 + y^2 = 1$



1) Convert a constraint O.F to an unconstraint O.F.

$f(h, s)$

$H(h, s)$

$$L(h, s, \lambda) = f(h, s) - \lambda H(h, s)$$

λ = Lagrange multiplier.

$$= f(h, s) - \lambda_1 H_1(h, s) - \lambda_2 H_2(h, s)$$

n

1

$$\left\{ \begin{aligned} \frac{\partial L}{\partial h} &= 0 \\ \frac{\partial L}{\partial s} &= 0 \\ \frac{\partial L}{\partial \lambda} &= 0 \end{aligned} \right.$$

E.g. $\quad f(h,s) = 200\, h^{2/3} s^{1/3} \quad$ s.t. $H(h,s) \Rightarrow 20h + 170s = 20{,}000$

$$L(h,s,\lambda) = 200\, h^{2/3} s^{1/3} - \lambda(20h + 170s - 20{,}000)$$

$$\left.\begin{aligned} \frac{\partial L}{\partial h} &= 200 \cdot \frac{2}{3} h^{-1/3} s^{1/3} - \lambda \cdot 20 = 0 \\ \frac{\partial L}{\partial s} &= 200 \frac{1}{3} h^{2/3} s^{-2/3} - 170\lambda = 0 \\ \frac{\partial L}{\partial \lambda} &= -20h - 170s + 20000 = 0 \end{aligned}\right\}$$

$h = 666.66$; $s = 39.12$; $\lambda = 3.59$

Max $f = 51777$

if two graphs are tangent at the point then their normal vectors must be parallel. — the two normal vectors must be scalar multiples of each other.

$$\nabla f(\ ) = \lambda \quad g(\ )$$

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

$\nabla f_x = \lambda \nabla g_x \qquad \nabla f_y = \lambda \nabla g_y \qquad \nabla f_z = \lambda \nabla g_z$

$\nabla f_x - \lambda \nabla g_x = 0$

$f(x)\ \ x^2 + y^2 - 4 = 0 \qquad g(x):\ \frac{x^2}{4} + y^2 - 1 = 0$

KKT $\quad \min_{w} f(w)$ $\qquad$ 23/09/2024

s.t. $g_i(w) \le 0$; $i = 1 \ldots K$

$h_i(w) = 0$; $i = 1 \ldots l$

Generalized Langrange Eq.

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{K} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

- Derivative of $\mathcal{L}$ w.r.t. w_is and β_is.
- $\frac{\partial \mathcal{L}}{\partial w_i} = 0 \qquad \frac{\partial \mathcal{L}}{\partial \beta_i} = 0$

$\alpha^* =$ optimal α

- $\alpha_i^* g_i(w) = 0$

$w^* =$ optimal w

- $g_i(w^*) \leq 0$
- $\alpha^* \geq 0$

Primal and Dual Problems

$$\min_{w} f(w) \quad \text{s.t. } g_i(w) \leq 0 \quad i = 1 \ldots k \text{ —— (i)}$$
$$h_i(w) = 0 \quad i = 1 \ldots l \text{ —— (ii)}$$

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

Let us define: $\theta_p(w) = \max_{\substack{\alpha, \beta \\ \alpha \geq 0}} \mathcal{L}(w, \alpha, \beta)$ $\quad p =$ primal

$$= \max_{\substack{\alpha, \beta \\ \alpha \geq 0}} f(w) + \sum \alpha_i g_i(w) + \sum \beta_i h_i(w)$$

Constraints violated:

if $g_i(w) > 0$ then $\alpha_i \rightarrow \infty$ and $\theta_p(w) \rightarrow \infty$

if $h_i(w) \neq 0$ then $\beta_i \rightarrow +\infty / -\infty$ and $\theta_p(w) \rightarrow \infty$

Constraints satisfied:

$\sum \alpha_i g_i(w) = 0$ and $\sum \beta_i h_i(w) = 0$

$$\theta_p(w) = f(w)$$

$$\theta_p(w) = \begin{cases} f(w) & ; \text{ if cond}^n \text{ satisfied} \\ \infty ; & \text{otherwise.} \end{cases}$$

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

Sol^n of the primal prob $\quad p^* = \min_{w} \max_{\substack{\alpha, \beta \\ \alpha \geq 0}} \mathcal{L}(w, \alpha, \beta)$

Dual Problem

$$d^* = \max_{\substack{\alpha, \beta \\ \alpha \geq 0}} \min_{w} \mathcal{L}(w, \alpha, \beta)$$

$$= \max_{\substack{\alpha, \beta \\ \alpha \geq 0}} \theta_d(\alpha, \beta)$$

Theorem Min Max inequality —

$$\max \min f(x) \le \min \max f(x)$$

$$d^* \le p^*$$

Under certain Condition. $d^* = p^*$

$f()$ and $g_i()$ are convex ; $h_i()$ are affine

$g_i()$ are (strictly) feasible.

then must exist w^*, α^*, β^* so that w^* is the sol^n of the primal problem and α^*, β^* are the sol^n of the dual problem.

and $p^* = d^* = L(w, \alpha, \beta)$

and w^*, α^*, β^* satisfy the KKT Cond^n.

- $\partial L / \partial w_i = 0$
- $\partial L / \partial \beta_i = 0$
- $\alpha_i^* g_i(w^*) = 0$
- $g_i(w^*) \le 0$
- $\alpha^* \ge 0$