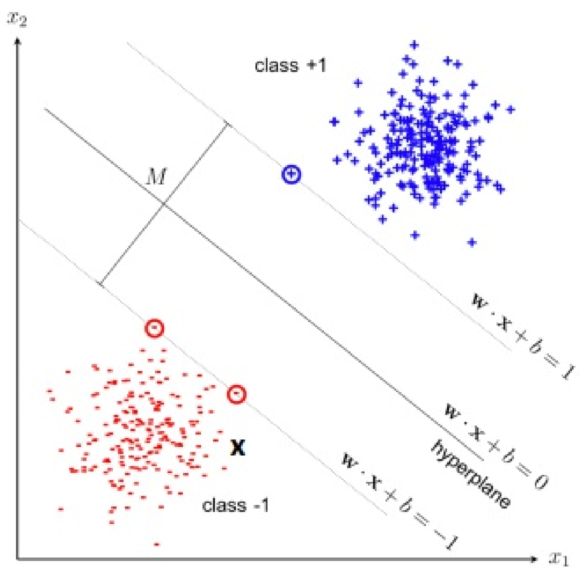
Support Vector Machine



Model ที่ใช้จำแนกประเภทข้อมูล (Classification) โดยสร้าง hyperplane ขึ้นมาเพื่อจำแนกข้อมูลออกจากกัน

Hyperplane

เพื่อหาให้ได้ระยะห่าง ระหว่าง 2 Class มากสุด
โดยหาจาก Lagrange Multipliers (LM), Karush Kuhn Tucker (KKT)

$X1$	$X2$	$\cdots$	XD	Y
x_1^1	x_1^2	$\cdots$	x_1^D	y_1
x_2^1	x_2^2	$\cdots$	x_2^D	y_2
x_3^1	x_3^2	$\cdots$	x_3^D	y_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_N^1	x_N^2	$\cdots$	x_N^D	y_N

X คือ ตัวแปรต้น (Feature)

Y คือ ตัวแปรตาม (Target)

N คือ จำนวนตัวอย่างทั้งหมด

D คือ จำนวน Feature ทั้งหมด

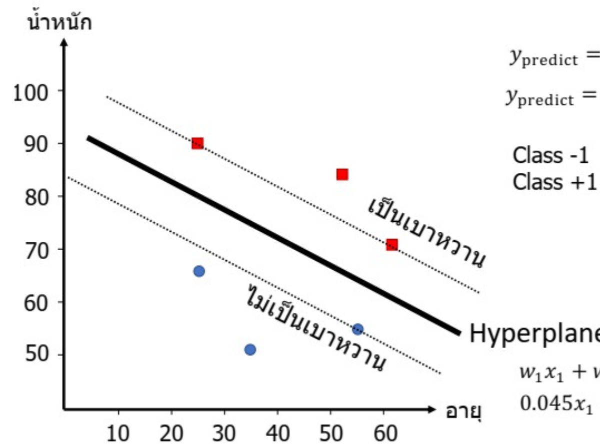
Ex.

$Data =$

อายุ	น้ำหนัก	เบาหวาน?
34	50	ไม่เป็น
25	90	เป็น
62	70	เป็น
53	85	เป็น
27	65	ไม่เป็น
42	55	ไม่เป็น

(+,-)

$$X = \begin{bmatrix} 34 & 50 \\ 25 & 90 \\ 62 & 70 \\ 53 & 85 \\ 27 & 65 \\ 42 & 55 \end{bmatrix}, Y = \begin{bmatrix} \text{ไม่เป็น} \\ \text{เป็น} \\ \text{เป็น} \\ \text{เป็น} \\ \text{ไม่เป็น} \\ \text{ไม่เป็น} \end{bmatrix}$$



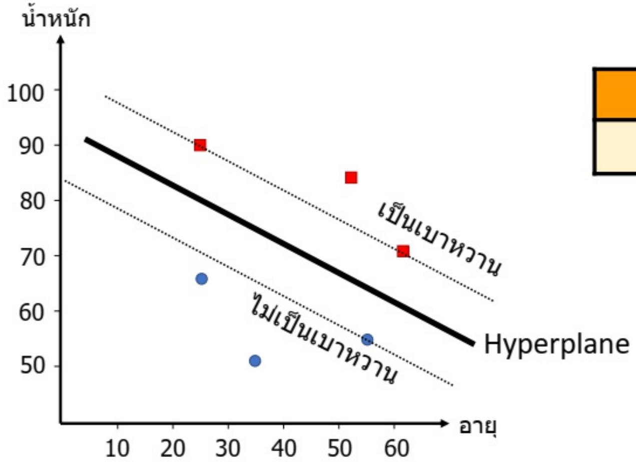
$$y_{\text{predict}} = \text{sgn}(w_1x_1 + w_2x_2 + b)$$

$$y_{\text{predict}} = \text{sgn}(0.045x_1 + 0.083x_2 - 7.65)$$

Class -1 : ไม่เป็นเบาหวาน
Class +1 : เป็นเบาหวาน

$$w_1x_1 + w_2x_2 + b = 0$$

$$0.045x_1 + 0.083x_2 - 7.65 = 0$$



อายุ	น้ำหนัก	เบาหวาน?
45	90	?

$$y_{\text{predict}} = \text{sgn}(0.045x_1 + 0.083x_2 \; - \; 7.65)$$

$$y_{\text{predict}} = \text{sgn}(0.045(45) + 0.083(90) \; - \; 7.65)$$

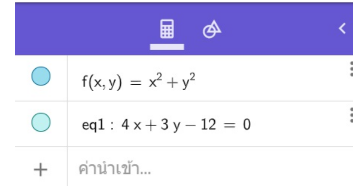
$$y_{\text{predict}} = \text{sgn}(1.845)$$

$$y_{\text{predict}} = +1$$

อายุ	น้ำหนัก	เบาหวาน?
45	90	เป็น

https://www.geogebra.org/3d?lang=th

Lagrange Multipliers (LM)

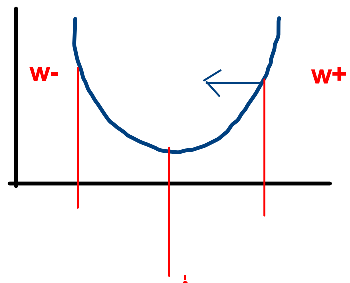


จำนวนเชิงบวก = []^2 + []^2 + []^2 +[]^2

99 = 9^2 + 4^2 + 1^2 +1^2

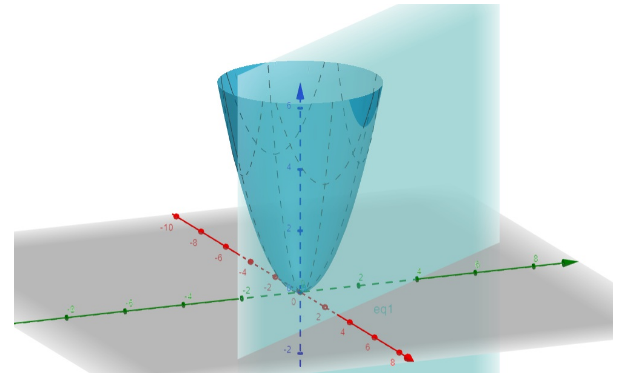
เลขยกกำลัง 2 สี่ตัว จะเท่ากับจำนวนเชิงบวก (Lagrange)
ทุกจำนวนสามารถทำในสมการนี้ได้

ความชัน = อัตราการเปลี่ยนแปลง x,y => Diff
(ความชันใรทุกทิศทาง หรืออัตราการเปลี่ยนแปลง)
[ปรับ w คือ ปรับความชัน]



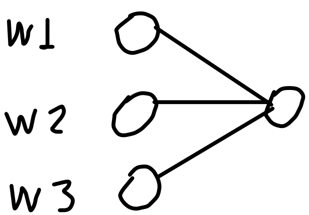
error ต่ำสุด

พื้นที่สัมผัส เกิดความชั้น



จุดที่สัมผัสกันเกาส์เซียน(Gaussia)
อยู่ในแนวเดียวกัน

(1) พื้นที่สัมผัสกัน = จะได้จุดต่ำสุด / สูงสุด



(ไว้หาความชันรายตัว)

$$W = W - \alpha \frac{\partial Error}{\partial w}$$

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_3 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_3 \end{bmatrix} - \begin{bmatrix} \frac{\partial Error}{\partial w} \\ \vdots \\ \frac{\partial Error}{\partial w} \end{bmatrix}$$

ความชันในทุกมิติ

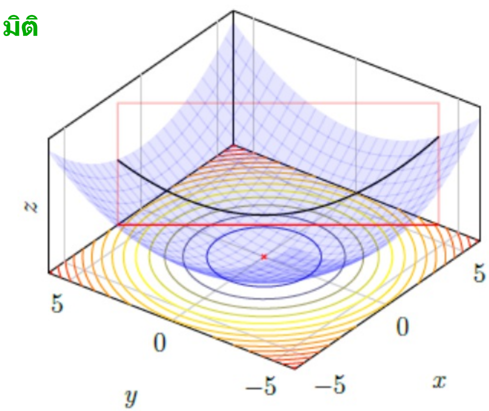
(ไว้หาความชันหลายๆ ตัว)

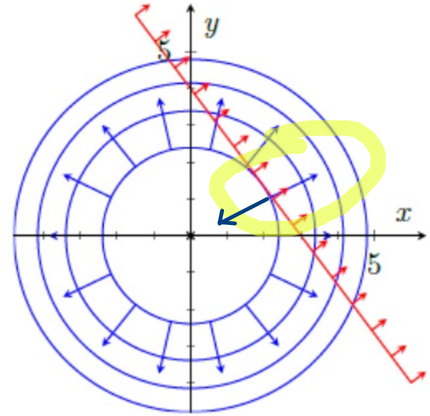
$$\boxed{W = W - \nabla Error}$$

Gradient

Gradient มาช่วยคำนวณ => LM

3 มิติ





(a)

ความชันไม่เท่า
ไปในแนวเดียวกัน

(2) ณ จุดสัมผัสกัน ความชัน (Slope) & เกรเดียนต์
(Gradient <u>สนามเวกเตอร์</u>) อยู่ในทิศทางเดียวกัน
หรือตรงข้ามก็ได้ (ความชันไม่เท่ากัน)
แต่จะพุ่งออกจากจุดเดียวกัน

Gradient fubction

$$\nabla f(x, y) = 0$$

ดังนั้น จึงสร้างตัวแปรขึ้นมาอีกตัว
เป็นค่าที่บ่งบอกถึงอัตราส่วนของขนาดของเวกเตอร์ทั้งสอง

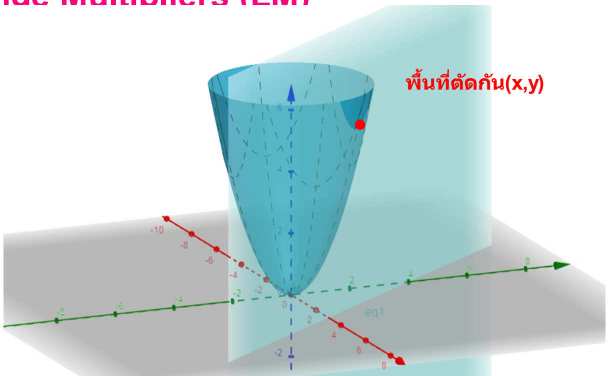
Lagrange Multipliers (LM)

Gradient ∇ ใส่ทั้งสอง เพราะไปในแนวเดียวกัน

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

แรมด้าติดลบ เพื่อจัดรูปง่าย
เวลาย้ายข้าง = 0

Lagrange Multipliers (LM)



$\nabla f(x,y)$ อยู่ในระนาบเดียวกัน $\nabla g(x,y)$

$\nabla f(x,y) \propto \nabla g(x,y)$

แปรผัน

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

$$\therefore \nabla f(x,y) + \lambda \nabla g(x,y) = \emptyset$$

$$\nabla f(x,y) + \sum_{i=1}^{M} \nabla g(x,y) = \emptyset$$

กรณีมีหลายตัว

หา x,y ต่ำสุดได้

step1 $\nabla\left(f(x,y) + \lambda g(x,y)\right) = \emptyset$

$L = f(x,y) + \lambda g(x,y)$

step2 $\nabla L = \emptyset \left\{ \frac{\partial L}{\partial x}, \frac{\partial L}{\partial y} \right\} = \emptyset$ ได้ x,y ไว้คิด λ

step3 แทน x,y ลงใน constrain g(x,y) เพื่อหา λ

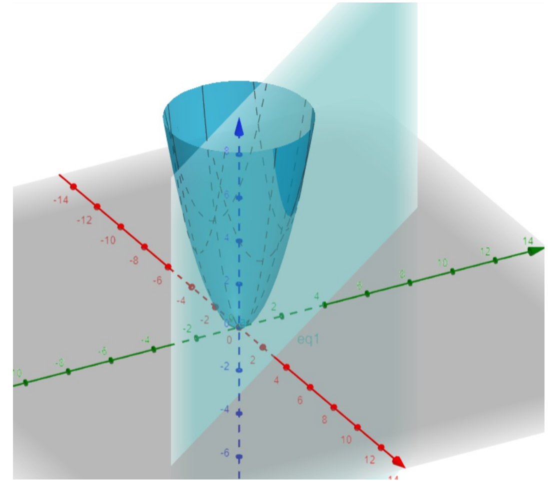
step4 แทนค่า x,y และ λ ในสมกการ f(x,y) ก็จะได้จุดตัด ต่ำสุด x,y ออกมา

ตัวอย่างคำนวณ Lagrange Multipliers (LM)

*Ex*1.

objective function : $f(x,y) = x^2 + y^2$

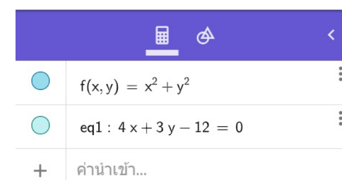
subject to : $g(x,y) = 4x + 3y - 12 = 0$



Gradient fubction

$$\nabla f(x,y) = 0$$

https://www.geogebra.org/3d?lang=th



1) หา Lagrange Function

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

$$\nabla f(x,y) + \lambda \nabla g(x,y) = 0$$

$$\nabla \left(f(x,y) + \lambda g(x,y) \right) = 0$$

$$L = f(x,y) + \lambda g(x,y)$$

$$L = x^2 + y^2 + \lambda(4x + 3y - 12)$$

$$\nabla L = 0$$

2) คำนวณ $\nabla L = \emptyset \rightarrow \frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}$

$$L = x^2 + y^2 + \lambda(4x + 3y - 12)$$

$$\frac{\partial L}{\partial x} = 2x + 4\lambda$$

$$\emptyset = 2x + 4\lambda$$

$$x = -2\lambda$$

$$\frac{\partial L}{\partial y} = 2y + 3\lambda$$

$$\emptyset = 2y + 3\lambda$$

$$y = \frac{-3\lambda}{2}$$

3) แทนค่า x,y ลงใน constraint => g(x,y) เพื่อคำนวณ λ

$$g(x,y) = 4x + 3y - 12 = 0$$

$$4(-2\lambda) + 3\left(\frac{-3\lambda}{2}\right) - 12 = 0$$

$$-8\lambda - \frac{9\lambda}{2} - 12 = 0$$

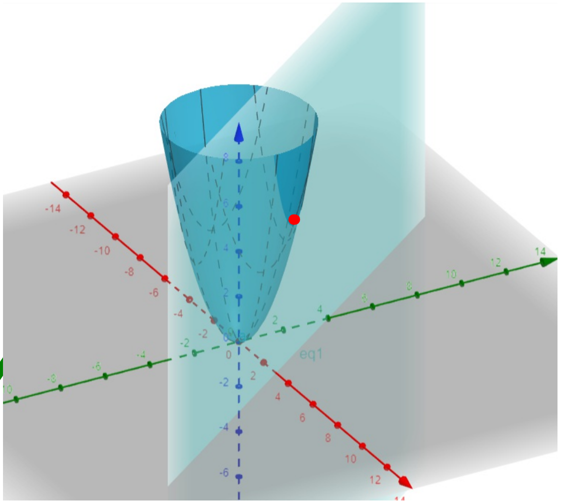
$$\frac{-16\lambda - 9\lambda}{2} - 12 = 0$$

$$\frac{-25\lambda}{2} = 12$$

$$\lambda = \frac{-24}{25}$$

4) คำนวณ x, y จาก λ

$$\rightarrow \quad x = -2\lambda \;,\quad y = \frac{-3\lambda}{2} \;,\quad \lambda = \frac{-24}{25}$$



$$x = -2\left(\frac{-24}{25}\right) = \frac{48}{25}$$

$$y = \frac{-3}{2}\left(\frac{-12}{25}\right) = \frac{36}{25}$$

จุดต่ำสุดที่สัมผัส (x, y)

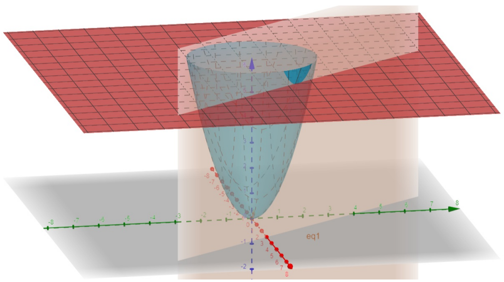
$$\therefore \quad f(x,y) = x^2 + y^2$$

$$= \left(\frac{48}{25}\right)^2 + \left(\frac{36}{25}\right)^2$$

$$= \frac{2304}{625} + \frac{1296}{625} = \frac{3600}{625}$$

$$= \frac{144}{25}$$

ระนาบ



$$G(x,y) = \left(\frac{48}{25}\right)^2 + \left(\frac{36}{25}\right)^2$$

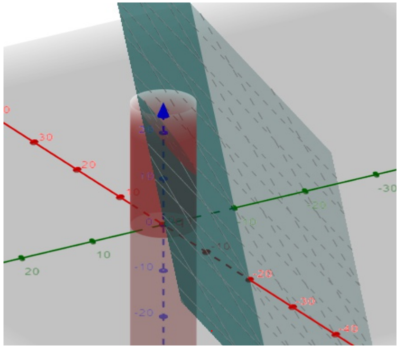
$$f(x,y) = x^2 + y^2$$

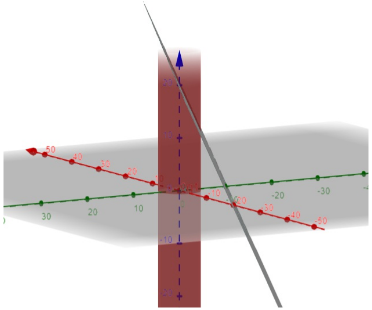
$$\text{eq1}: 4x + 3y - 12 = 0$$

Ex 2.

objective function : $f(x, y) = x + 2y + 20$

subject to : $g(x, y) = x^2 + y^2 - 16 = 0$





$f(x,y) = x + 2y + 20$

eq1: $x^2 + y^2 - 16 = 0$

1) หา Lagrange Function

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

$$\nabla f(x,y) + \lambda \nabla g(x,y) = 0$$

$$\nabla \left(f(x,y) + \lambda g(x,y) \right) = 0$$

$$L = f(x,y) + \lambda g(x,y)$$

$$L = x + 2y + 20 + \lambda(x^2 - y^2 - 16)$$

$$\nabla L = 0$$

2) คำนวณ $\nabla L = \emptyset \rightarrow \frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}$

$L = x + 2y + 20 + \lambda(x^2 - y^2 - 16)$

$$\frac{\partial L}{\partial x} = 1 + 2\lambda x$$

$$\emptyset = 1 + 2\lambda x$$

$$x = -\frac{1}{2\lambda}$$

$$\frac{\partial L}{\partial y} = 2 + 2\lambda y$$

$$\emptyset = 2 + 2\lambda y$$

$$y = \frac{-2}{2\lambda} = \frac{-1}{\lambda}$$

3) แทนค่า x,y ลงใน constraint => g(x,y) เพื่อคำนวณ λ

$g(x, y) = x^2 + y^2 - 16 = 0$

$$\left(\frac{-1}{2\lambda}\right)^2 + \left(\frac{-1}{\lambda}\right)^2 - 16 = \emptyset$$

$$\frac{1}{4\lambda} + \frac{4}{4\lambda} = 16$$

$$\frac{5}{4\lambda^2} = 16$$

$$\lambda^2 = \frac{5}{64}$$

$$\lambda = \pm\frac{\sqrt{5}}{8}$$

4) คำนวณ x, y จาก λ

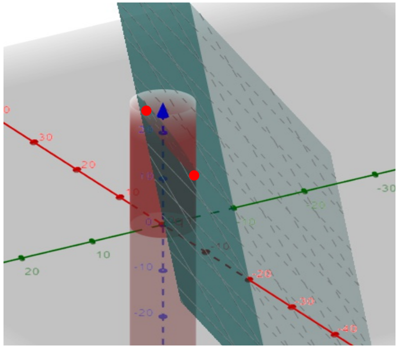
Case1 $\lambda = \frac{\sqrt{5}}{8}$, $x = \frac{-1}{2\lambda}$, $y = \frac{-1}{\lambda}$

$$x = -\frac{1}{2} \div \frac{\sqrt{5}}{8} = -\frac{1}{2} \times \frac{8}{\sqrt{5}} = -\frac{4}{\sqrt{5}}$$

$$y = -1 \div \frac{\sqrt{5}}{8} = -1 \times \frac{8}{\sqrt{5}} = -\frac{8}{\sqrt{5}}$$

Case2 $\lambda = \frac{-\sqrt{5}}{8}$

$$x = \frac{4}{\sqrt{5}}, \quad y = \frac{8}{\sqrt{5}}$$



แทนค่าหา min,max จาก f(x,y)

$$f(x,y) = x + 2y + 20$$

Case1

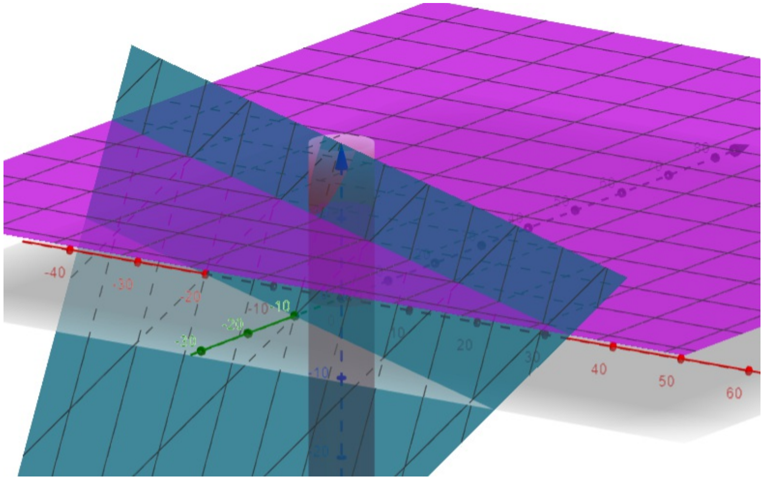
$$-\frac{4}{\sqrt{5}} + 2\left(\frac{-8}{\sqrt{5}}\right) + 20$$

$$= -\frac{20}{\sqrt{5}} + 20 \quad \textbf{min}$$

Case2

$$\frac{4}{\sqrt{5}} + 2\left(\frac{8}{\sqrt{5}}\right) + 20$$

$$= \frac{20}{\sqrt{5}} + 20 \quad \textbf{max}$$

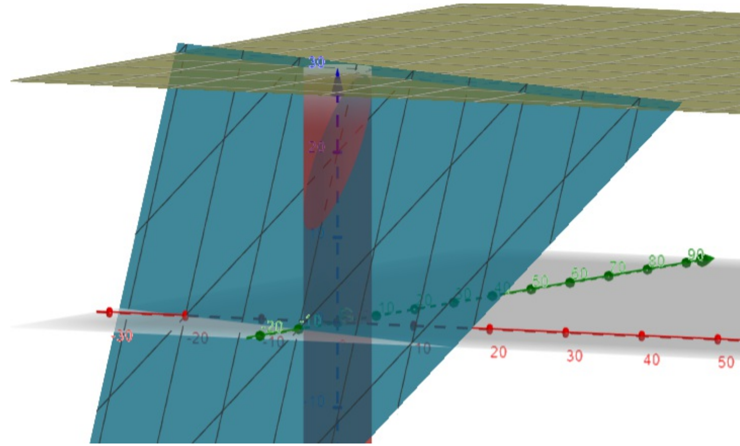


$f(x, y) = x + 2y + 20$

eq1: $x^2 + y^2 - 16 = 0$

$E(x, y) = -\frac{4}{\sqrt{5}} - 2 \cdot \frac{8}{\sqrt{5}} + 20$

min



$f(x, y) = x + 2y + 20$

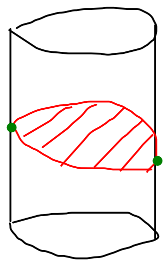
eq1: $x^2 + y^2 - 16 = 0$

$G(x, y) = \frac{4}{\sqrt{5}} + 2 \cdot \frac{8}{\sqrt{5}} + 20$

max

Karush Kuhn Tucker (KKT)

Lagrange

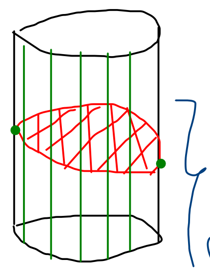


$\nabla f(x,y) = -\lambda \nabla (x,y)$

1) $\nabla L = 0$

2) $g(x,y) = 0$

KKT



ดูเป็นพื้นที่

ไม่ได้หา " = " ได้เสมอ

$\leq \quad \lambda \to \mu$

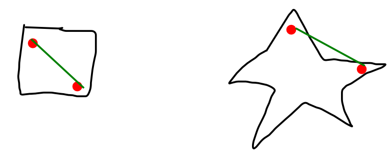
min

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

max

$\nabla f(x,y) = \mu \nabla g(x,y)$

แก้ปัญหารูปทรงแบบ convax (เลขาคณิต)



Yes
ใช้ได้กับ KKT ได้

No

convax = สมการลากเส้นจากจุด 2 จุดที่สั้นที่สุด ไปได้ในพื้นที่เดียวกัน

กฎและเงื่อนไข (KKT)

1) Stationarity $\quad \nabla L(x,\lambda,\mu) = 0 \Rightarrow \nabla L = 0$

2) Primal feasibility $\quad g_i(x) \leq 0,\ h_i(x) = 0 \Rightarrow g(x,y) \leq 0$

3) Dual feasibility $\quad \mu_i \geq 0 \Rightarrow \mu \geq 0$

เพื่อให้ได้ค่า Max / ค่า Min
ก็คิดลบสลับด้านกัน

4) Complementary slackness $\quad \mu_i g_i(x) = 0 \Rightarrow \underline{\mu^* g(x,y) = 0}$

Global Point (G)

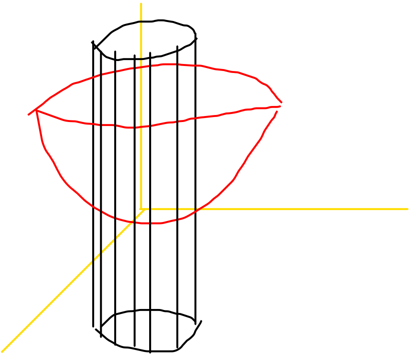
$\nabla f(x,y)$

$\longrightarrow$

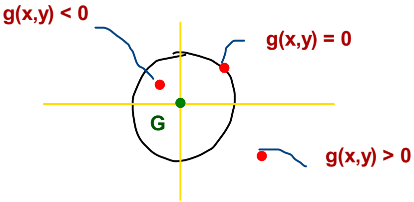
$\longleftarrow \mu \nabla g(x,y)$

$\longrightarrow$

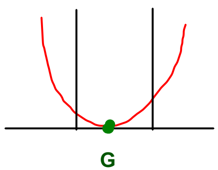
$-\mu \nabla g(x,y)$



Top View



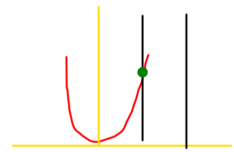
Objctive function จุดต่ำสุด / สูงสุดของ
คือจุด Global Point (G)

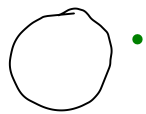


$$\nabla f(x,y) = 0$$

$$\mu = 0$$

$$g(x,y) \neq 0$$





กรณี G อยู่นอก Constraint

$$\mu \neq 0$$

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

step1 $\nabla\left(f(x,y) + \mu g(x,y)\right) = 0$

$$L = f(x,y) + \mu g(x,y)$$

step2 $\nabla L = 0 \; \left\{ \frac{\partial L}{\partial x}, \frac{\partial L}{\partial y} \right\} = 0$ ได้ x,y ไว้คิด μ

step3 แทน x,y ลงใน Objective function (L) ติดค่า μ รูปบบ Dual From -> (LM)

step4 แทนค่า x,y ใน Constrain g(x,y) เพื่อหา Objective function ใน Dual From => ได้ μ

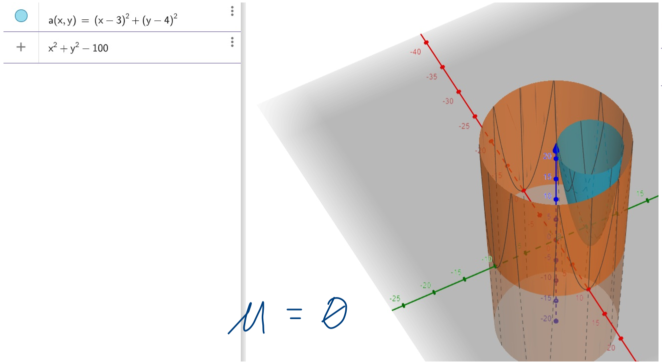
step5 ค่าใน ∇L (ของ Dual From) $\left.\right\}$ $\frac{\partial L}{\partial \mu} = \frac{\partial \text{(สมการ L ใน step3)}}{\partial \mu}$

แทนค่า x,y ในสมการ L (step 3)

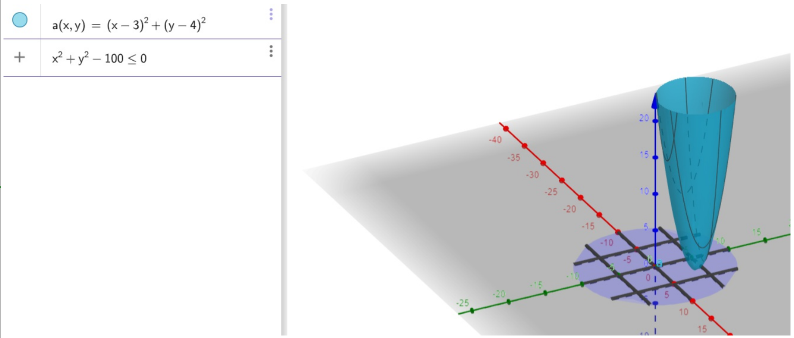
Karush Kuhn Tucker

*Ex*1.

$$\begin{aligned} &\textit{minimize}: && f(x,y) = (x-3)^2 + (y-4)^2 \\ &\textit{subject to}: && g(x,y) = x^2 + y^2 - 100 \le 0 \end{aligned}$$



$\mu = 0$



1) หา Lagrangian Function

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

$\nabla f(x,y) + \mu \nabla g(x,y) = 0$

$\nabla \left(f(x,y) + \mu g(x,y) \right) = 0$

$L = f(x,y) + \mu g(x,y)$

$\therefore L = (x-3)^2 + (y-4)^2 + \mu(x^2 + y^2 - 100)$

$\nabla L = 0$

2) คำนวณ $\nabla L = \emptyset$

หาจุดต่ำสุดแบบทั่วๆไป

$$\frac{\partial f(x,y)}{\partial x} = 0$$

$$2(x-3) = 0$$

$$x = 3$$

$$\frac{\partial f(x,y)}{\partial y} = 0$$

$$2(y-4) = 0$$

$$y = 4$$

$$\frac{\partial L}{\partial x} = 2(x-3) + 2\mu x$$

$$0 = 2(x-3) + 2\mu x$$

$$0 = x - 3 + \mu x$$

$$\mu x + x = 3$$

$$x(\mu+1) = 3$$

$$\boxed{x = \frac{3}{\mu+1}}$$

$$\frac{\partial L}{\partial y} = 2(y-4) + 2\mu y$$

$$0 = y - 4 + \mu y$$

$$\mu y + y = 4$$

$$y(\mu+1) = 4$$

$$\boxed{y = \frac{4}{\mu+1}}$$

3) แทนค่า x,y ลงใน L เพื่อหา Objective function (L) ใน Dual Form

$$\boxed{x = \frac{3}{\mu+1}} \qquad \boxed{y = \frac{4}{\mu+1}}$$

$$\therefore L = (x-3)^2 + (y-4)^2 + \mu(x^2+y^2-100)$$

$$= \left(\frac{3}{\mu+1}-3\right)^2 + \left(\frac{4}{\mu+1}-4\right)^2 + \mu\left[\left(\frac{3}{\mu+1}\right)^2 + \left(\frac{4}{\mu+1}\right)^2 - 100\right]$$

$$= 3^2\left(\frac{1}{\mu+1}-1\right)^2 + 4^2\left(\frac{1}{\mu+1}-1\right)^2 + \mu\left[\frac{25}{(\mu+1)^2} - 100\right]$$

$$= 25\left(\frac{1}{\mu+1}-1\right)^2 + \frac{25}{(\mu+1)^2} - 100\mu$$

$1-(\mu+1)$

$$= 25\left(\frac{1-\mu-1}{\mu+1}\right)^2 + \frac{25\mu}{(\mu+1)^2} - 100\mu$$

$$= \frac{25\mu^2}{(\mu+1)^2} + \frac{25\mu}{(\mu+1)^2} - 100\mu$$

$$= \frac{25\mu^2 + 25\mu}{(\mu+1)^2} - 100\mu$$

$$= \frac{25\mu(\mu+1)}{(\mu+1)^2} - 100\mu$$

$$= \frac{25\mu}{\mu+1} - 100\mu$$

$$= \frac{25\mu - 100\mu(\mu+1)}{\mu+1}$$

$$= \frac{25\mu - 100\mu^2 - 100\mu}{\mu+1}$$

$$L = \frac{-100\mu^2 - 75\mu}{\mu+1}$$

Dual Form

$\boxed{x = \frac{3}{\mu+1}}$ $\boxed{y = \frac{4}{\mu+1}}$

4) แทนค่า x,y ลงใน ลงใน constraint (g(x,y)) เพื่อหา Objective function ใน Dual Form

$\therefore x^2 + y^2 - 100 \le 0$

$\left(\frac{3}{\mu+1}\right)^2 + \left(\frac{4}{\mu+1}\right)^2 - 100 \le 0$

$\frac{9}{(\mu+1)^2} + \frac{16}{(\mu+1)^2} \le 100$

$\frac{25}{(\mu+1)^2} \le 100$

$25 \le 100(\mu+1)^2$

หารด้วย100

$\frac{1}{4} \le (\mu+1)^2$

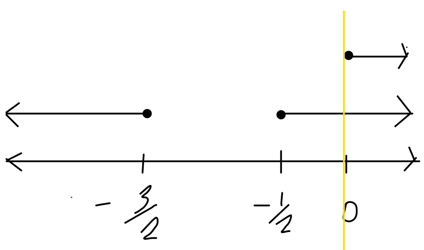
$(\mu+1)^2 - \frac{1}{4} \ge 0$

$4\mu^2 + 8\mu + 3 \ge 0$

$(2\mu+3)(2\mu+1) \ge 0$

$\mu = -\frac{3}{2} \;;\; -\frac{1}{2}$ # KKT $\mu \ge 0$

$\mu \ge 0$



$\mu \in (-\infty, -\frac{3}{2}] \cup [-\frac{1}{2}, \infty)$

$\mu \in [0, \infty)$ # คำตอบ ต่าง

$\mu = 0 \to \infty$

5) คำนวณ $\nabla L = 0$ (L จากขั้นตอนที่ 3) ของ Dual Form

$$L = \frac{-100\mu^2 - 75\mu}{\mu+1}$$

ล่างดิฟบน ลบ บนดิฟล่าง
ส่วนด้วยล่างกำลังสอง

$$\frac{\partial L}{\partial \mu} = \frac{(\mu+1)\frac{\partial(-100\mu^2-75\mu)}{\partial\mu} - (-100\mu^2-75\mu)\frac{\partial(\mu+1)}{\partial\mu}}{(\mu-1)^2}$$

$$0 = \frac{(\mu+1)(-200\mu-75) + (100\mu^2+75\mu)(1)}{(\mu-1)^2}$$

$$0 = (\mu+1)(-200\mu-75) + 100\mu^2 + 75\mu$$

$$= -200\mu^2 - 75\mu - 200\mu - 75 + 100\mu^2 + 75\mu$$

$$= -100\mu^2 - 200\mu - 75$$

$$= 100\mu^2 + 200\mu + 75$$

$$= 4\mu^2 + 3\mu + 3$$

$$= (2\mu+3)(2\mu+1)$$

$\therefore \mu = \frac{-3}{2}, -\frac{1}{2}$

KKT กำหนดให้ $\mu \geq 0$

$\therefore \mu = 0$

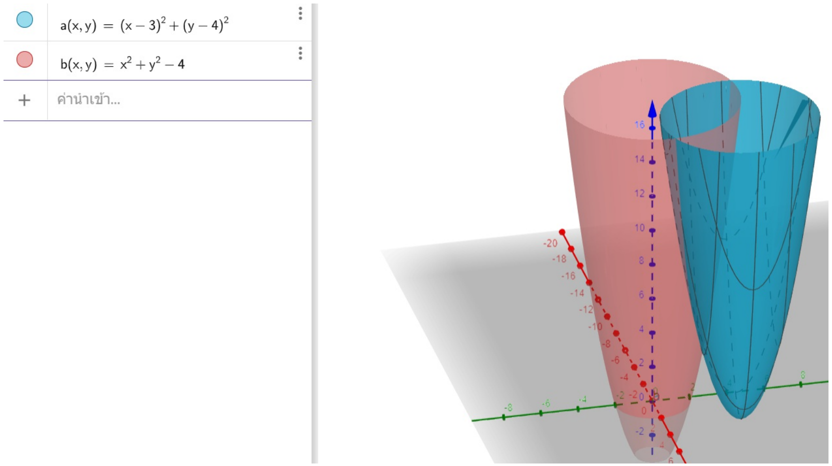
แกนตัดใน x,y จากขั้นตอนที่ 2

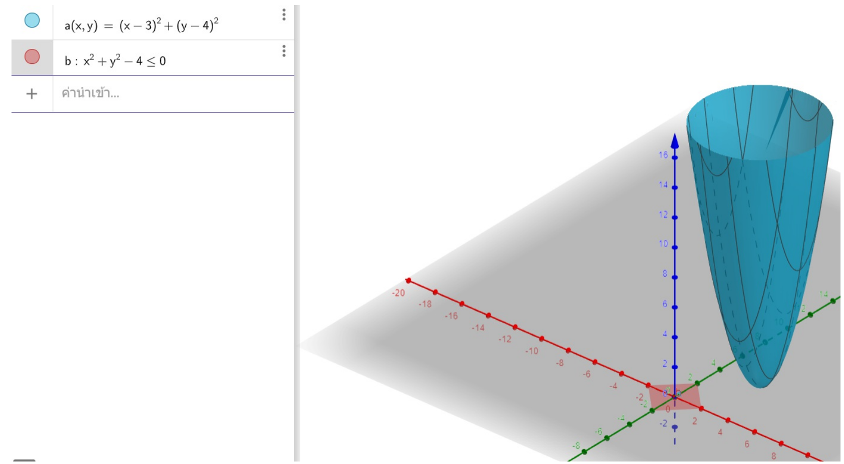
$$x = \frac{3}{\mu+1} = 3$$

$$y = \frac{4}{\mu+1} = 4$$

Ex1.

$$\begin{aligned} &\textit{minimize}: && f(x,y) = (x-3)^2 + (y-4)^2 \\ &\textit{subject to}: && g(x,y) = x^2 + y^2 - 4 \le 0 \end{aligned}$$





1) หา Lagrangian Function

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

$$\nabla f(x,y) + \mu \nabla g(x,y) = 0$$

$$\nabla \left(f(x,y) + \mu g(x,y) \right) = 0$$

$$L = f(x,y) + \mu g(x,y)$$

$$\therefore L = (x-3)^2 + (y-4)^2 + \mu(x^2+y^2-4)$$

$$\nabla L = 0$$

2) คำนวณ $\nabla L = \emptyset$

$$\frac{\partial L}{\partial x} = 2(x-3) + 2\mu x$$

$$\emptyset = 2(x-3) + 2\mu x$$

$$\emptyset = x - 3 + \mu x$$

$$\mu x + x = 3$$

$$x(\mu + 1) = 3$$

$$\boxed{x = \frac{3}{\mu+1}}$$

$$\frac{\partial L}{\partial y} = 2(y-4) + 2\mu y$$

$$\emptyset = y - 4 + \mu y$$

$$\mu y + y = 4$$

$$y(\mu + 1) = 4$$

$$\boxed{y = \frac{4}{\mu+1}}$$

3) แทนค่า x,y ลงใน L เพื่อหา Objective function (L) ใน Dual Form

$$\therefore L = (x-3)^2 + (y-4)^2 + \mu(x^2+y^2-4)$$

$$= \left(\frac{3}{\mu+1} - 3\right)^2 + \left(\frac{4}{\mu+1} - 4\right)^2 + \mu\left[\left(\frac{3}{\mu+1}\right)^2 + \left(\frac{4}{\mu+1}\right)^2 - 4\right]$$

$$= 3^2\left(\frac{1}{\mu+1} - 1\right)^2 + 4^2\left(\frac{1}{\mu+1} - 1\right)^2 + \mu\left[\frac{9}{(\mu+1)^2} + \frac{16}{(\mu+1)^2} - 4\right]$$

$$= 25\left(\frac{1}{\mu+1} - 1\right)^2 + \frac{25}{(\mu+1)^2} - 4\mu$$

$1-(\mu+1)$

$$= 25\left(\frac{1-\mu-1}{\mu+1}\right)^2 + \frac{25\mu}{(\mu+1)^2} - 4\mu$$

$$= \frac{25\mu^2}{(\mu+1)^2} + \frac{25\mu}{(\mu+1)^2} - 4\mu$$

$$= \frac{25\mu^2 + 25\mu}{(\mu+1)^2} - 4\mu$$

$$= \frac{25\mu(\mu+1)}{(\mu+1)^2} - 4\mu$$

$$= \frac{25\mu}{\mu+1} - 4\mu$$

$$= \frac{25\mu - 4\mu(\mu+1)}{\mu+1}$$

$$= \frac{25\mu - 4\mu - 4\mu^2}{\mu+1}$$

$$L = \frac{21\mu - 4\mu^2}{\mu+1}$$

Dual Form

ทำต่อขั้นตอน5

4) แทนค่า x,y ลงใน ลงใน constraint (g(x,y)) เพื่อหา Objective function ใน Dual Form

$$\therefore x^2 + y^2 - 4 \le 0$$

$$\left(\frac{3}{\mu+1}\right)^2 + \left(\frac{4}{\mu+1}\right)^2 - 4 \le 0$$

$$\frac{9}{(\mu+1)^2} + \frac{16}{(\mu+1)^2} \le 4$$

$$\frac{25}{(\mu+1)^2} \le 4$$

$$25 \le 4(\mu+1)^2$$

$$\frac{25}{4} \le (\mu+1)^2$$

$$0 \le (\mu+1)^2 - \frac{25}{4}$$

$$0 \le 1 + 2\mu + \mu^2 - \frac{25}{4}$$

$$0 \le 4 + 8\mu + 4\mu^2 - 25$$

$$0 \le 4\mu^2 + 8\mu - 21$$

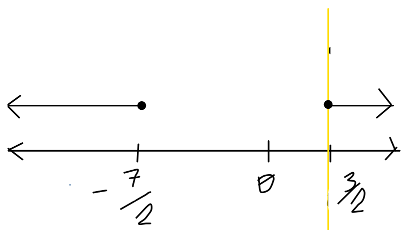
$$0 \le (2\mu+7)(2\mu-3)$$

$$(2\mu-7)(2\mu+3) \ge 0$$

$\mu = -\frac{7}{2}\ ;\ \frac{3}{2}$ # KKT $\mu \ge 0$

$\mu \ge 0$

$= 3/2$



$$\mu \in \left(-\infty, -\tfrac{7}{2}\right] \cup \left[\tfrac{3}{2}, \infty\right)$$

$\mu \in \left[\tfrac{3}{2}, \infty\right)$ # คำตอบค่าไ

5) คำนวณ $\nabla L = 0$ (L จากขั้นตอนที่ 3) ของ Dual Form

$$L = \frac{21\mu - 4\mu^2}{\mu + 1}$$

$$\frac{\partial L}{\partial \mu} = \frac{(\mu+1)(-8\mu+21) - (-4\mu^2 + 21\mu)(1)}{(\mu+1)^2}$$

$$0 = \frac{-8\mu^2 + 21\mu - 8\mu + 4\mu^2 - 21\mu + 21\mu}{(\mu+1)^2}$$

$$0 = \frac{-4\mu^2 - 8\mu + 21\mu}{(\mu+1)^2}$$

$$4\mu^2 + 8\mu - 21 = 0$$

μ ที่ถูกต้อง $\rightarrow \mu = 3/2$

$$\mu = \frac{-8 \pm \sqrt{8^2 - (4)(4)(-21)}}{2(4)} \quad \begin{cases} \frac{-8+20}{8} = 3/2 \\ \frac{-8-20}{8} = -7/2 \end{cases}$$

KKT กำหนดให้ $\mu \geq 0$ $\nearrow$ $(\mu - 1)^2 \rightarrow 1$

$$\mu = -\frac{7}{2} \; ; \; \frac{3}{2} \quad \# \text{ KKT } \mu \geq 0$$

แทนค่าใน x,y จากขั้นตอนที่ 2

$x = \frac{3}{\mu+1} = \frac{6}{5}$	$y = \frac{4}{\mu+1} = 8/5$

Duplicate a(x,y)

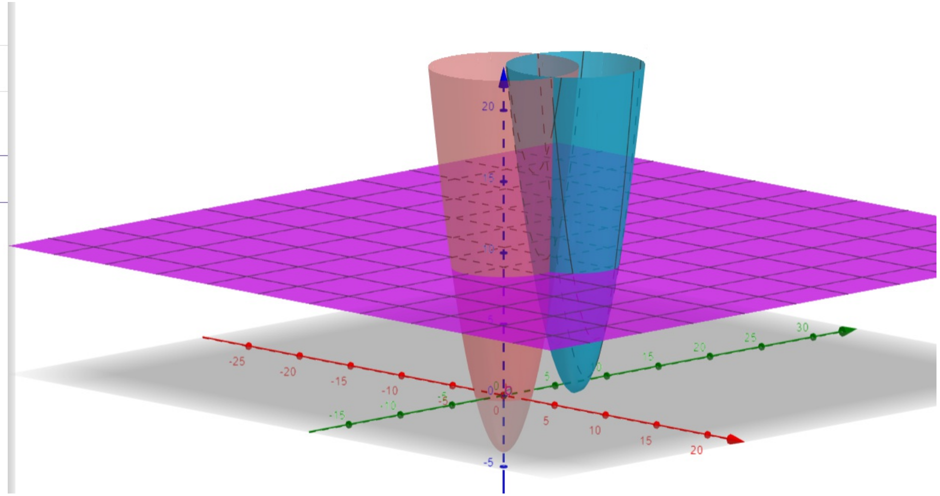
X = 6/5 , Y = 8/5

$a(x, y) = (x-3)^2 + (y-4)^2$

$b(x, y) = x^2 + y^2 - 4$

$c(x, y) = \left(\frac{6}{5} - 3\right)^2 + \left(\frac{8}{5} - 4\right)^2$

+ ค่านำเข้า...

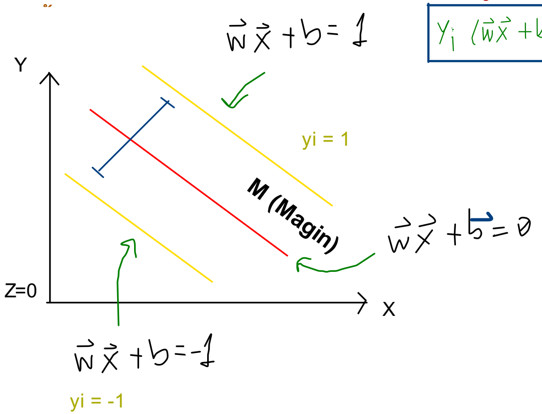


Support Vector Machine

เส้นแบ่งสองกลุ่มให้ได้ระยะห่างมากที่สุด = Error

subject to

$$\boxed{y_i(\vec{w}\vec{x}+b) \geq 1}$$



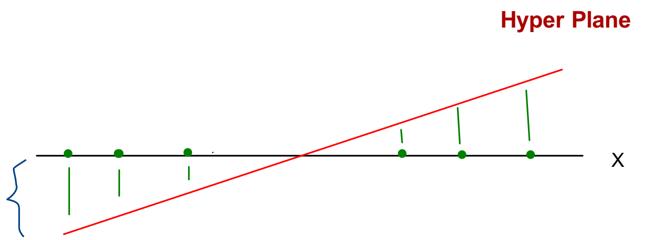
Magin (M) => $\vec{W}$

เราอยากได้ Hyper Plane ที่ทำให้ Magin ห่างมากสุด

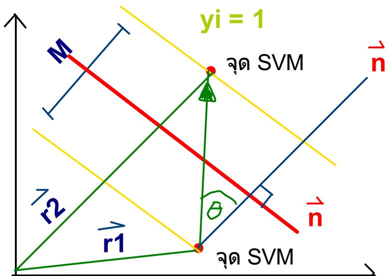
$$w_1x_1 + w_2x_2 + \dots + w_Dx_D + b = 0$$

x เป็นค่าคงที่ รู้จากข้อมูล Data set

จุดที่ใกล้ที่สุดเรียกว่า Support Vector



จุด SVM = จุดที่ใกล้เว้น Hyperplen



subject to

yi = -1

$\boxed{y_i\,(\vec{w}\vec{x}+b) \geq 1}$

สิ่งที่รู้

$$\vec{w}\vec{r}_2 + b = 1$$

$$\vec{w}\vec{r}_1 + b = -1$$

$$M = \|\uparrow\| \cos\theta$$

$$M = \|\vec{r}_2 - \vec{r}_1\| \cos\theta$$

หาขนาด M (Magin) ได้

จากมุม

$$\vec{A}\cdot\vec{B} = \|A\| \times \|B\| \times \cos\theta$$

$$\therefore\ M = \|\vec{r}_2 - \vec{r}_1\| \times \|A\| \times \cos\theta$$

$$= \langle \vec{r}_2 - \vec{r}_1 \rangle \cdot n$$

จาก

$$M = \|\vec{r}_2 - \vec{r}_1\| \cdot 1 \cdot \cos\theta$$

$$= \|\vec{r}_2 - \vec{r}_1\| \cdot \|\vec{n}\| \cdot \cos\theta$$

$$= \langle \vec{r}_2 - \vec{r}_1 \rangle \cdot \vec{n}$$

เวกเตอร์ 1 หน่วยตั้งฉาก

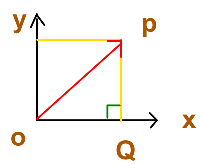
$$= \langle \vec{r}_2 - \vec{r}_1 \rangle \cdot \frac{\vec{w}}{\|\vec{w}\|}$$

เขียนขนาด M ให้อยู่ในรูป เวกเตอร์ W ได้

ค่าคงที่

$\|\vec{w}\|$

Note เวกเตอร์ 1 หน่วยแหนมุมฉาก



$$\overrightarrow{OP} = \overrightarrow{OQ} + \overrightarrow{QP}$$

$$= \vec{x}_i + \vec{y}_j$$

$$= \begin{bmatrix} x \\ y \end{bmatrix} \text{ or } (x, y)$$

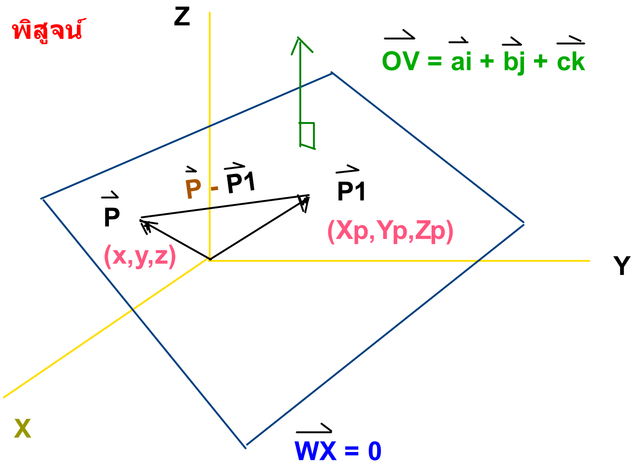
สูตร

$$\vec{u}_v = \frac{\vec{v}}{|\vec{v}|} = \frac{\vec{x}_i + \vec{y}_i}{\sqrt{x^2 + y^2}}$$

$$\therefore\ \frac{\vec{w}}{\|\vec{w}\|} \ \#$$

$$\|\vec{w}\| = \sqrt{w_1^2 + w_2^2 + \dots + w_D^2}$$

พิสูจน์



$$A \bullet B = \|A\| \times \|B\| \times \cos(90^\circ)$$
$$= 0$$

0

$\cos 90^\circ$

$$\vec{p}_1 = x_p\hat{i} + y_p\hat{j} + z_p\hat{k}$$
$$\vec{p} = x\hat{i} + y\hat{j} + z\hat{k}$$
$$\vec{p} - \vec{p}_1 = (x - x_p)\hat{i} + (y - y_p)\hat{j} + (z - z_p)\hat{k}$$
$$\vec{OV} \times (\vec{p} - \vec{p}_1) = 0$$

ตั้งฉาก

$$a(x - x_p) + b(y - y_p) + c(z - z_p) = 0$$
$$ax + by + cz - ax_p - by_p - cz_p = 0$$
$$ax + by + cz = ax_p + by_p + cz_p$$
$$w_1x_1 + w_2x_2 + \dots + w_px_p = -b$$

$$\vec{w}\vec{x} + b = 0 \Rightarrow \vec{w}\vec{x} = -b$$
$$\vec{OV} = \vec{w}$$

$$\vec{n} = \frac{\vec{OV}}{\|\vec{w}\|} = \frac{\vec{w}}{\|\vec{w}\|}$$

ต้องทำให้ขนาดของ(ตัวส่วน) $\|\vec{W}\|$ น้อยสุด เพื่อให้ได้ M มากสุด

เมื่อ

$$\left.\begin{array}{c} M = \max \\ \downarrow \\ \vec{W} = \min \end{array}\right\} \text{(KKT)}$$

จึงได้

minimize $\|\vec{w}\|$

subject to $y_i(\vec{w}\vec{x}+b) \geq 1$

$$\|\vec{w}\| = \sqrt{w_1^2 + w_2^2 + \dots + w_p^2}$$

เพื่อให้ง่ายต่อการ diff $\|\vec{w}\|^2$

เพื่อไม่ให้มีเศษส่วน $\frac{1}{2}\|\vec{w}\|^2$

subject to

$$y_i(\vec{w}\vec{x}+b) \geq 1$$

กรณีไม่มี Error

ซาย $\xi = 0$

กรณีมี Error

$\xi \neq 0$

ξ

$$y_i(\vec{w}\vec{x}+b) \geq 1$$

$$y_i(\vec{w}\vec{x}+b) \geq 1 - [\underline{1 - y(\vec{w}\vec{x}-b)}]$$

$$\boxed{\therefore\ y_i(\vec{w}\vec{x}+b) \geq 1 - \xi}$$

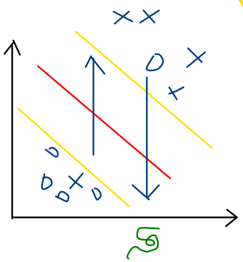
กรณีที่มี error จึงได้

minimize $$\frac{1}{2}\|\vec{w}\|^2 + \left(\sum_{i=1}^{n} \xi_i\right)$$

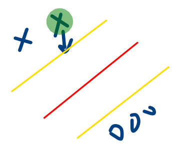
$$\frac{1}{2}\|\vec{w}\|^2 + \underline{C}\left(\sum_{i=1}^{n} \xi_i\right)$$

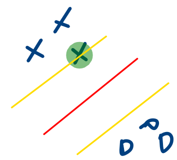
เพื่อลดความ Overfit

$$\therefore \xi_i = \max\left[0, 1 - y_i(\vec{w}\vec{x} + b)\right]$$

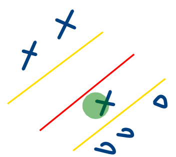


1) ถ้าอยู่ถูกฝั่ง [0 , ติดลบ] => [0 , 1 - (เลขมากกว่าเท่ากับ 1)]

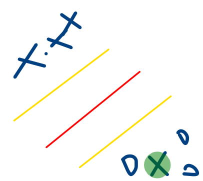




อยู่ถึงฝั่ง บนเส้น เป็น support vector



2) ถ้าอยู่ผิดฝั่ง ระหว่าง magin [0 , บวก] => [0 , 1 - (เลขน้อยกว่า 1)]



3) ถ้าอยู่ผิดฝั่ง [0 , บวก] => [0 , 1 - (น้อยกว่า ไม่เท่ากับ 1)]

Error จะมาก

จึงได้

minimize $\frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{N}\xi_i$

subject to $y_i(\vec{w}\vec{x}_i + b) \geq 1 - \xi_i$

เมื่อ $\xi_i \geq 0$

จัดรูปใหม่

minimize $\frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{N}\xi_i$

subject to $1 - \xi_i - y_i(\vec{w}\vec{x}_i + b) \leq 0$

เมื่อ $-\xi_i \leq 0$

Lagrangan fuction $\nabla f(w_1, w_2, \ldots, w_D, \xi_1, \xi_2, \ldots, \xi_N) = -\sum_{i=1}^{N}\alpha_i \nabla g_i(b, w_1, w_2, \ldots, w_D, \xi_i) - \sum_{i=1}^{N}\beta_i \nabla h_i(\xi_i)$

KKT

Dual From (SVM)

maxinize $\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \vec{x}_i\vec{x}_j$

subject to $\sum_{i=1}^{N}\alpha_i y_i = 0$

เมื่อ $0 \leq \alpha_i \leq C$

เมื่อ Data มีมากกว่า Feature
ทำให้หา Inver ไม่ได้
หรือเกิด Overfit จึงต้องใช้
SMO เข้ามาช่วย

สวัสดี Sequential Minimal Optimization (SMO) Algorithm

แก้ปัญหา Time complexity ของ SVM O(n^3) => SMO อยู่ในช่วงระหว่าง O(n^2)

$\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \vec{x}_i \vec{x}_j$

subject to

$\boxed{y_i (\vec{w}\vec{x} + b) \geq 1}$

SMO จะมีการอัปเดทค่า α ข้อมูล 2 จุด (α_i ตัวตั้งต้น ไม่ได้ถูกเลือกให้เป็นค่าคงที่), (α_j ที่ถูกเลือก)

สรุปสมการที่ใช้ใน SMO

$E_i = \vec{w} \cdot \vec{x}_i + b - y_i$

1) Update α_b : α_j

เมื่อ(eta) $\eta = 2K_{ab} - K_{aa} - K_{bb}$

$= 2\vec{x}_i \vec{x}_j - \vec{x}_i \vec{x}_i - \vec{x}_j \vec{x}_j$

$\alpha_b^{new} = \alpha_b^{old} - \frac{y_b (E_a^{old} - E_b^{old})}{\eta}$

ไม่เท่ากับ 0 ถ้าเป็น 0 ต้องสุ่มใหม่

2) หาขอบใหม่ (Clip)

ให้ L,H เป็นขอบล่างและขอบบน

ถ้า $y_a = y_b$ $(y_i = y_j)$

$L = \max(0, \alpha_i + \alpha_j - C)$

$H = \min(C, C + \alpha_j - \alpha_i)$

ถ้า $y_a \neq y_b$ $(y_i \neq y_j)$

$L = \max(0, \alpha_j - \alpha_i)$

$H = \min(C, C + \alpha_j - \alpha_i)$

เขาสมการเงื่อนไขเพื่ออัปเดท $\alpha_b^{new,Clipped}$

i = a
j = b

เขาสมการเงื่อนไขเพื่ออัปเดท $\alpha_b^{new,Clipped}$

$$\alpha_b^{new,Clipped} = \begin{cases} L \; ; & \alpha_b^{new} < L \\ \alpha_b^{new} \; ; & L \le \alpha_b^{new} \le H \\ H \; ; & \alpha_b^{new} > H \end{cases}$$

α_j

---

3) Update α_a

$(\alpha_i$ **ตัวเริ่มต้น**$)$

$$\alpha_a^{new} = \alpha_a^{old} + S(\alpha_b^{old} - \alpha_b^{new})$$

$$= \alpha_i^{old} + S(\alpha_j^{old} - \alpha_j^{new})$$

$$\boxed{S = y_i y_j = y_a y_b}$$

---

4) Update (b)

ให้เลือกค่าระหว่าง α_a และ α_b ว่าอยู่ในช่วง (0, C) <ไม่รวมจุดต้นและปลาย>

แต่ถ้าอยู่ในช่วง (0,C) ทั้งคู่ $(\alpha_a = \alpha_b)$ ให้นำมาเฉลี่ยกัน $\frac{b_a + b_b}{2}$

$$b = \begin{cases} b_a \; ; & 0 < \alpha_a < C \\ b_b \; ; & 0 < \alpha_b < C \\ \frac{b_a + b_b}{2} \; ; & \text{otherwise} \end{cases}$$

$$b_i^{new} = b_i^{old} - E_i^{old} - (\alpha_i^{new} - \alpha_i^{old}) y_i \vec{x}_i \cdot \vec{x}_i - (\alpha_j^{new} - \alpha_j^{old}) y_j \vec{x}_i \cdot \vec{x}_j$$

$$b_j^{new} = b_j^{old} - E_j^{old} - (\alpha_i^{new} - \alpha_j^{old}) y_j \vec{x}_j \cdot \vec{x}_j - (\alpha_i^{new} - \alpha_i^{old}) y_i \vec{x}_i \cdot \vec{x}_j$$

5) Update (W : weight)

$$W^{new} = W^{old} + (\alpha_a^{new} - \alpha_a^{old}) y_a x_a + (\alpha_b^{new} - \alpha_b^{old}) y_b x_b$$

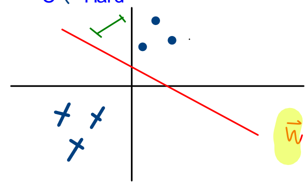
$$= W^{old} + (\alpha_i^{new} - \alpha_j^{old}) y_i \vec{x}_i + (\alpha_j^{new} - \alpha_j^{old}) y_i x_j$$

EX

Index	X1	X2	Y
1	1	1	1
2	2	1	1
3	2	2	1
4	-1	-2	-1
5	-2	-1	-1
6	-2	-2	-1

เวลาดูข้อมูล [1 1] => มองเป็น Vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

C ← Hard



สิ่งที่ต้องหา

$\vec{w}\vec{x} + b = 0$

กำหนดให้

C = 1 b = 0

$\alpha = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ W = $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$

C แปรผันกับ Hard Magin
(C มาก Overfit ได้)

1) คำนวณ $E_i = \vec{w} \cdot \vec{x}_i + b - Y_i$

สุ่มเลือกเอา Index ที่ 1

$$E_i = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0 - 1 = -1$$

มองอยู่ในรูป Vector = [ตั้ง]

Dot product

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 7 \end{bmatrix} = [(1 \times 3) + (3 \times 7)] = 26$$

2) คำนวณ $E_j = \vec{w} \cdot \vec{x}_j + b - Y_j$

สุ่มเลือกเอา Index ที่ 3

$$E_j = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} + 0 - 1 = -1$$

3) คำนวณ(eta) $\eta = 2\vec{x}_i \cdot \vec{x}_j - \vec{x}_i \cdot \vec{x}_i - \vec{x}_j \vec{x}_j$

$$= 2\vec{x}_1\vec{x}_3 - \vec{x}_1\vec{x}_1 - \vec{x}_3\vec{x}_3$$

$$= 2\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}2\\2\end{bmatrix} - \begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}2\\2\end{bmatrix}\begin{bmatrix}2\\2\end{bmatrix}$$

$$\eta = (2\times 4) - 2 - 8 = -2 \ \#$$

---

4) หาขอบล่างของ α_j จาก $\overset{1}{y_i} = \overset{1}{y_j} \Rightarrow L = \max(0, \alpha_i + \alpha_j - C)$

$$= \max(0, 0 - 1) \Rightarrow -1 \ \#$$

0 มีค่ามากกว่า -1
เลือกตัวน้อยสุด

5) หาขอบบนของ α_j จาก $y_i = y_j \Rightarrow H = \min(C, \alpha_i + \alpha_j)$

$$= (1, 0 + 0) \Rightarrow 0 \ \#$$

1 มีค่ามากกว่า 0
เลือกตัวน้อยสุด

6) update $\alpha_j^{new} = \alpha_j^{old} - \dfrac{y_j(E_i - E_j)}{\eta}$

$$\alpha_j^{new} = 0 - \frac{1[(-1)-(-1)]}{-2} = 0 \quad \#$$

7) clip α_j^{new}
$$\begin{cases} L \,; & \alpha_b^{new} < L \\ \alpha_b^{new} \,; & L \le \alpha_b^{new} \le H \\ H \,; & \alpha_b^{new} > H \end{cases}$$

$L = -1, \alpha_j = 0, H = 0$

$\therefore \alpha_j^{new} = 0$

8) update $\alpha_i^{new} = \alpha_i^{old} + s(\alpha_j^{old} - \alpha_j^{new})$

$$= 0 + 1(0 - 0) = 0$$

$\boxed{s = y_a y_b = (1)(1) = 1}$

$\boxed{s = y_i y_j = (1)(1) = 1}$

9) คำนวณ $b_i = b^{old} - E_i - (\alpha_i^{new} - \alpha_i^{old}) y_i \vec{x}_i \cdot \vec{x}_i - (\alpha_j^{new} - \alpha_j^{old}) y_j \vec{x}_i \cdot \vec{x}_j$

$$= 0 - (-1) - (0 - 0)(1)\begin{bmatrix}1\\1\end{bmatrix} \cdot \begin{bmatrix}1\\1\end{bmatrix} - (0 - 0)(1)\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}2\\2\end{bmatrix}$$

$b_i = 1$ #

10) คำนวณ $b_j = b^{old} - E_j - (\alpha_j^{new} - \alpha_j^{old}) y_j \vec{x}_j \cdot \vec{x}_j - (\alpha_i^{new} - \alpha_i^{old}) y_i \vec{x}_i \cdot \vec{x}_j$

$$= 0 - (-1) - (0 - 0)(1)\begin{bmatrix}2\\2\end{bmatrix} \cdot \begin{bmatrix}2\\2\end{bmatrix} - (0 - 0)(1)\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}2\\2\end{bmatrix}$$

$b_j = 1$ #

11) update b

$$b = \begin{cases} b_i \;; & \alpha_i \in (0, C),\ \alpha_j \notin (0, C) \\ & 0 < \alpha_i < C \\ b_j \;; & \alpha_i \notin (0, C),\ \alpha_j \in (0, C) \\ & 0 < \alpha_j < C \\ \dfrac{b_i + b_j}{2} \;; & \text{other wise (หาค่าเฉลี่ยทั้งสองตัว)} \end{cases}$$

$$b = \begin{cases} b_a \;; & 0 < \alpha_a < C \\ b_b \;; & 0 < \alpha_b < C \\ \dfrac{b_a + b_b}{2} \;; & \text{otherwise} \end{cases}$$

$\therefore \left.\begin{matrix} \alpha_i = 0,\ \alpha_j = 0 \\ b_i = 1,\ b_j = 1 \end{matrix}\right\}$ อยู่ข้างเดียวกันทั้งคู่ => จึงหาค่าเฉลี่ย $\dfrac{b_i + b_j}{2} = \dfrac{1+1}{2} = 1 \;\#$

12) update W

$$w^{new} = w^{old} + (\alpha_i^{new} - \alpha_i^{old}) y_i \vec{x}_i + (\alpha_j^{new} - \alpha_j^{old}) y_j \vec{x}_j$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + (0 - 0)(1) \begin{bmatrix} 1 \\ 1 \end{bmatrix} + (0 - 0)(1) \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$= 0 \Rightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} \;\#$$

ผลลัพธ์คำนวณรอบแรก w = 0 , b =1

ผลลัพธ์รอบสอง $E_i = 0$, $E_j = 2$, $\eta = -20$, $H = 1$, $L = 0$, $\alpha_j^{new} = 0.1$, $\alpha_i^{new} = 0.1$

$b_i = 1$, $b_j = -1$, $b = 0$

สุ่มเลือก index 2 ตัว => i = 2 , j =5 (index ที่ 2, 5)

$$\vec{w} = \vec{w}^{old} + (\alpha_2^{new} - \alpha_2^{old}) y_2 \vec{x}_2 + (\alpha_5^{new} - \alpha_5^{old}) y_5 \vec{x}_5$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + (0.1 - 0)(1)\begin{bmatrix} 2 \\ 1 \end{bmatrix} + (0.1 - 0)(-1)\begin{bmatrix} -2 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.1\begin{bmatrix} 2 \\ 1 \end{bmatrix} - 0.1\begin{bmatrix} -2 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix} + \begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} \#$$

- Code SVM (SMO) -

1) คำนวณ $E_i = \vec{w} \cdot \vec{x}_i + b - y_i$ และ $E_j = \vec{w} \cdot \vec{x}_j + b - y_j$

```python
def SVM_compute_E(W, x_train, b, y):
    E = np.dot(x_train, W) + b - y
    return E
```

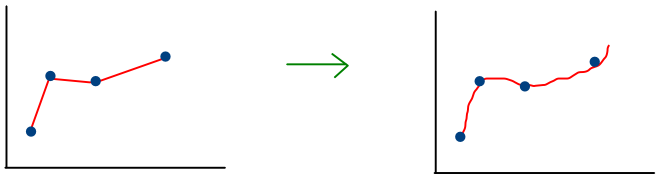
2) คำนวณ(eta) $\eta = 2\vec{x}_i \cdot \vec{x}_j - \vec{x}_i \cdot \vec{x}_i - \vec{x}_j \vec{x}_j$

เปลี่ยนจาก dot เป็น**product ใช้ kernel**

```python
def SVM_compute_eta(xi, xj, kf):
    eta = 2*kernel(xi, xj, kf) - kernel(xi, xi, kf) - kernel(xj, xj, kf)
    return eta
```

```python
def kernel(xi, xj, kf):
    if kf[0] == 'Dot':
        return np.dot(xi, xj)
    elif kf[0] == 'Polynomial':
        p = kf[1]
        return (np.dot(xi, xj) + 1)**p
    elif kf[0] == 'RBF':
        gamma = kf[1]
        degree = np.sum((xi - xj)**2)
        return np.e**(-gamma*degree)
```

RBF (เรเดียนเบสิกฟังก์ชัน) : จะ diff
จำนวนที่ดีที่สุด แบ่งให้ได้ Feature ที่เหมาะสม



diff ได้

---

3) หาขอบล่างของ α_i จาก

$y_i = y_j \quad L = \max(0, \alpha_i + \alpha_j - C)$

$y_i \neq y_j \quad L = \max(0, \alpha_j - \alpha_i)$

```python
def SVM_compute_L(yi, yj, C, alpha_i, alpha_j):
    if yi == yj:
        L = max([0, alpha_i + alpha_j - C])
    elif yi != yj:
        L = max([0, alpha_j - alpha_i])
    return L
```

---

4) หาขอบล่างของ α_i จาก

$y_i = y_j \qquad H = \min(C,\ C + \alpha_j - \alpha_i)$

$y_i \neq y_j \qquad H = \min(C,\ C + \alpha_j - \alpha_i)$

```
def SVM_compute_H(yi, yj, C, alpha_i, alpha_j):
    if yi == yj:
        H = min([C, alpha_i + alpha_j])
    elif yi != yj:
        H = min([C, C + alpha_j - alpha_i])
    return H
```

5) หา α_j^{new}

```
def SVM_compute_alpha_j(alpha_j, yj, Ei, Ej, eta, L, H):
    new_alpha_j = alpha_j - yj*(Ei - Ej)/eta
    if new_alpha_j < L:
        new_alpha_j = L
    elif new_alpha_j > H:
        new_alpha_j = H
    return new_alpha_j
```

$$\begin{cases} L \ ; & \alpha_b^{new} < L \\ \alpha_b \ ; & L \leq \alpha_b^{new} \leq H \\ H \ ; & \alpha_b^{new} > H \end{cases}$$

6) update b

$$
b = \begin{cases} b_i \; ; & \alpha_i \in (0, C),\ \alpha_j \notin (0, C) \\ & 0 < \alpha_i < C \\ b_j \; ; & \alpha_i \notin (0, C),\ \alpha_j \in (0, C) \\ & 0 < \alpha_j < C \\ \dfrac{b_i + b_j}{2} \; ; & \text{other wise (หาค่าเฉลี่ยทั้งสองตัว)} \end{cases}
$$

$$b_i = b^{old} - E_i - (\alpha_i^{new} - \alpha_i^{old}) y_i \vec{x}_i \cdot \vec{x}_i - (\alpha_j^{new} - \alpha_j^{old}) y_j \vec{x}_i \cdot \vec{x}_j$$

$$b_j = b^{old} - E_j - (\alpha_j^{new} - \alpha_j^{old}) y_j \vec{x}_j \cdot \vec{x}_j - (\alpha_i^{new} - \alpha_i^{old}) y_j \vec{x}_i \cdot \vec{x}_j$$

```python
def SVM_compute_b(b, Ei, Ej, new_alpha_i, old_alpha_i, new_alpha_j, old_alpha_j, yi, xi, yj, xj, C, kf):
    bi = b - Ei - (new_alpha_i - old_alpha_i)*yi*kernel(xi, xi, kf) - (new_alpha_j - old_alpha_j)*yj*kernel(xi, xj, kf)
    bj = b - Ej - (new_alpha_j - old_alpha_j)*yj*kernel(xj, xj, kf) - (new_alpha_i - old_alpha_i)*yi*kernel(xi, xj, kf)
    is_sv_i = (new_alpha_i > 0) and (new_alpha_i < C)
    is_sv_j = (new_alpha_j > 0) and (new_alpha_j < C)
    if is_sv_i and not is_sv_j:
        new_b = bi
    elif not is_sv_i and  is_sv_j:
        new_b = bj
    else:
        new_b = (bi + bj)/2
    return new_b
```

9) คำนวณ bi

10) คำนวณ bj

b

7) Run ทั้งหมด
(พยากรณ์)

```
def SVM_fit(X_Train, Y_Train, C, kf, max_iteration = 1000):
    N = X_Train.shape[0]
    D = X_Train.shape[1]
    alpha = np.zeros(N)
    W = np.zeros(D)
    b = 0
    for e in range(max_iteration):
        two_dots_index = random.sample(range(N), 2)
        i = two_dots_index[0]
        xi = X_Train[i, :]
        yi = Y_Train[i, 0]
        j = two_dots_index[1]
        xj = X_Train[j, :]
        yj = Y_Train[j, 0]
        Ei = SVM_compute_E(W, xi, b, yi)
        Ej = SVM_compute_E(W, xj, b, yj)
        eta = SVM_compute_eta(xi, xj, kf)
        if eta >= 0:
            continue
        L = SVM_compute_L(yi, yj, C, alpha[i], alpha[j])
        H = SVM_compute_H(yi, yj, C, alpha[i], alpha[j])
        new_alpha_j = SVM_compute_alpha_j(alpha[j], yj, Ei, Ej, eta, L, H)
        new_alpha_i = alpha[i] + yi*yj*(alpha[j] - new_alpha_j)
        b = SVM_compute_b(b, Ei, Ej, new_alpha_i, alpha[i], new_alpha_j, alpha[j], yi, xi, yj, xj, C, kf)
        W = W + (new_alpha_i - alpha[i])*yi*xi + (new_alpha_j - alpha[j])*yj*xj
        alpha[i] = new_alpha_i
        alpha[j] = new_alpha_j
    return W, b, alpha
```

12) update W

7) ทำนาย Predict (Test)

```python
def SVM_predict(X_Test, W, b):
    Zhat_Test = np.dot(X_Test, W) + b
    Yhat_Test = np.sign(Zhat_Test)
    return Yhat_Test.reshape(-1, 1)
```

$\vec{W}\vec{X} + b = 0$

$X / \sqrt{X^* X}$

$\begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$

row ไม่กำหนด
column = 1