Soft Margin SVM

$$\underset{A, B, C}{\arg\min} \ \frac{\sqrt{A^2+B^2}}{2}$$

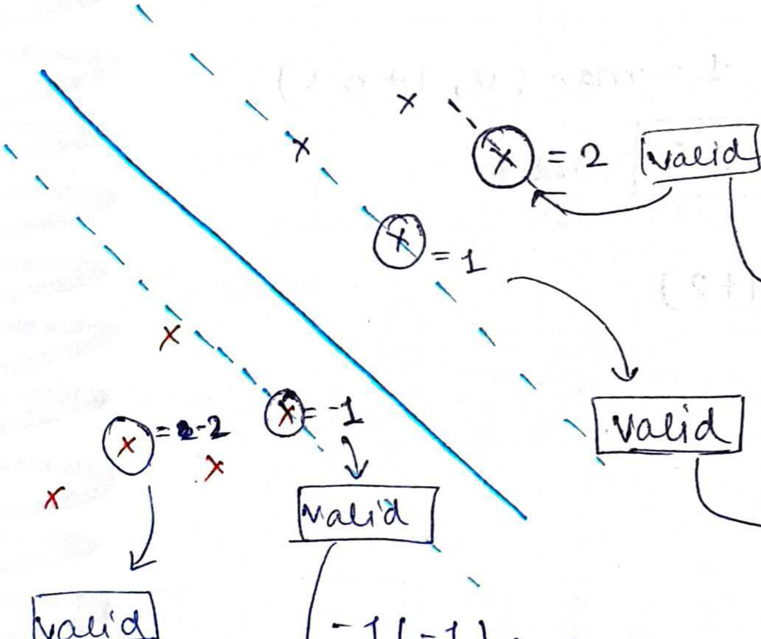
such that

$$\boxed{y_i(Ax_{1i} + Bx_{2i} + C) \geq 1 - \xi_i}$$

for all x_i

such that

$$\xi_i \geq 0$$



$1(2) \geq 1 - \max(0, 1-1(2))$

$\boxed{2 \geq 1}$

$1(1) \geq 1 - \max(0, 1-1(1))$

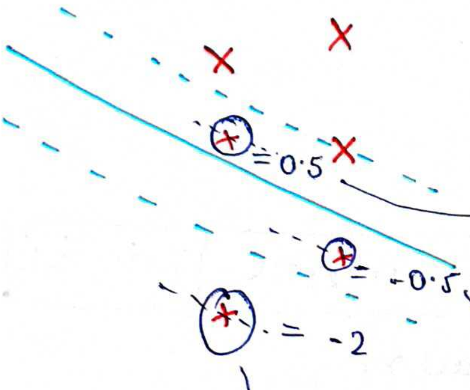
$\boxed{1 \geq 1}$

$-1(-1) \geq 1 - \max(0, 1-1)$

$\boxed{1 \geq 1}$

$-1(-2) \geq 1 - \max(0, 1-2)$

$\boxed{2 \geq 1}$



$$y_i(Ax_{1i} + Bx_{2i} + C) \geq 1 - \xi_i$$

$$1(0.5) \geq 1 - \max(0, 1 - 0.5)$$

$$0.5 \geq 1 - \max(0, 0.5)$$

$$\boxed{0.5 \geq = 0.5}$$ valid

$$1(-0.5) \geq 1 - \max(0, 1 + 0.5)$$

$$\boxed{-0.5 \geq -0.5}$$ valid

$$1(-2) \geq 1 - \max(0, 1 + 2)$$

$$\boxed{-2 \geq -2}$$ valid

Problem

$$\underset{A,B,C}{\arg\min} \ \frac{\sqrt{A^2 + B^2}}{2}$$

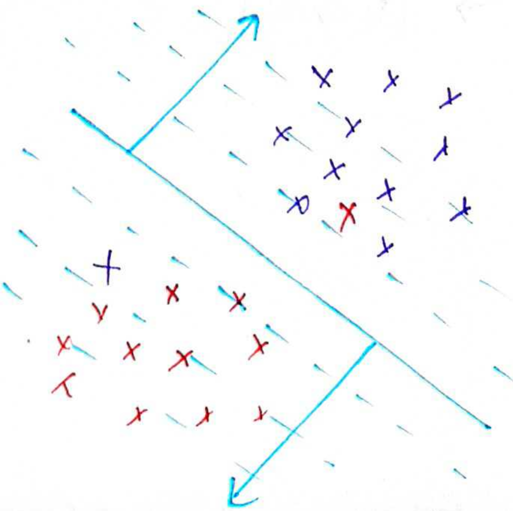
such that

$y_i(Ax_{1i} + Bx_{2i} + C) \geq 1 - \xi_i$

for all x_i

such that $\xi_i \geq 0$

* There 's no constraint to stop the margin because every point is valid and allowed. Margin distance Increasing continously because of no constaint



Solution

$$\underset{A,B,C}{\operatorname{argmin}} \; \overbrace{\boxed{\frac{\sqrt{A^2+B^2}}{2}}}^{\text{margin}\uparrow\uparrow} + \overbrace{\underbrace{\boxed{\frac{1}{n}\sum_{i=1}^{n} \xi_i}}_{\text{reduce}}}^{\text{reduce Avg. missclassification}}$$

$\uparrow$ Big C

such that

$y_i(Ax_{1i} + Bx_{2i} + C) \geq 1 - \xi_i$

for all x_i

and $\xi_i \geq 0$

Both opposite to each other because if margin increase then missclassification also increase. if missclassification decrease than margin also decrease.

Introduction of C

$$\underset{A,B,C}{\operatorname{argmin}} \; \frac{\sqrt{A^2+B^2}}{2} + \overset{\text{hyperparameter}}{\underset{\text{small c}}{C}} \frac{1}{n}\sum_{i=1}^{n} \xi_i$$

such that

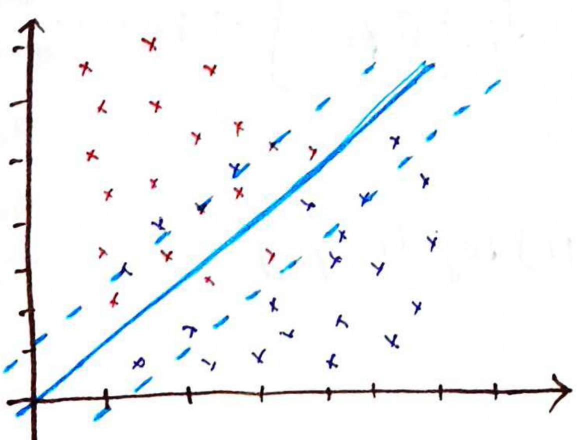
$y_i(Ax_{1i} + Bx_{2i} + C) \geq 1 - \xi_i$

for all x_i

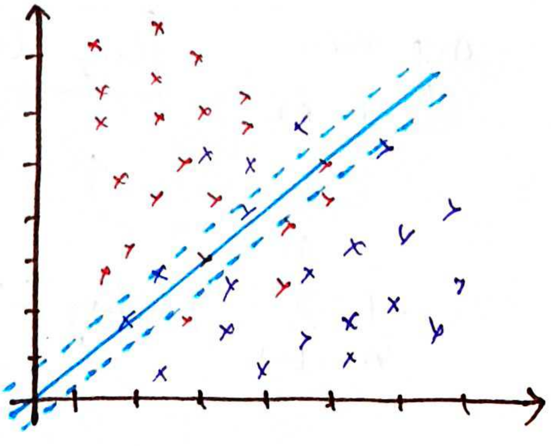
and $\xi_i \geq 0$

if $C\uparrow\uparrow$ = margin $\downarrow\downarrow$

$C\downarrow\downarrow$ = margin $\uparrow\uparrow$



C = 1



C=100

Bias Variance Trade off

$$\underset{A, B, C}{\arg\min} \; \frac{\sqrt{A^2 + B^2}}{2} + C \frac{1}{n} \sum_{i=1}^{n} \xi_i$$

Such that

$y_i(Ax_{1i} + Bx_{2i} + C \geq 1 - \xi_i$

for all x_i

and $\xi \geq 0$

→ C high → overfitting (low bias high variance)

→ C low → underfitting (high bias low variance)

Relationship with Logistic Reg.

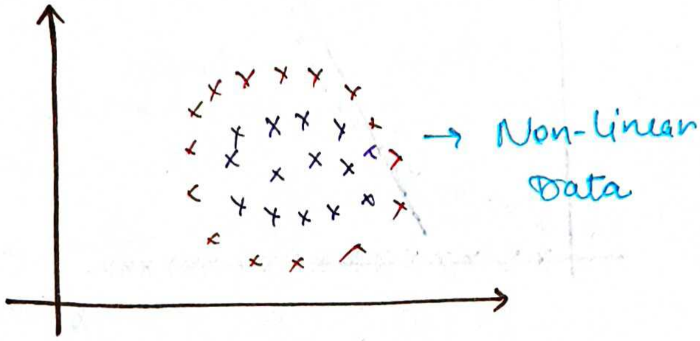
$$\underset{A, B, C}{\arg\min} \; \boxed{\frac{\sqrt{\beta_1^2 + \beta_2^2}}{2}} + \boxed{C \frac{1}{n} \sum_{i=1}^{n} \xi_i}$$ Constrain

Logistic Reg

$$\underset{\beta_0\, \beta_1\, \beta_2}{\arg\min} \; \boxed{\log \text{ loss}} + \boxed{\lambda \sqrt{\beta_1^2 + \beta_2^2}}$$ → regression

$$\frac{-1}{n} \sum_{i=1}^{n} y_i \log(\hat{y_i}) + (1 - y_i) \log(1 - \hat{y_i})$$

Problem with SVC

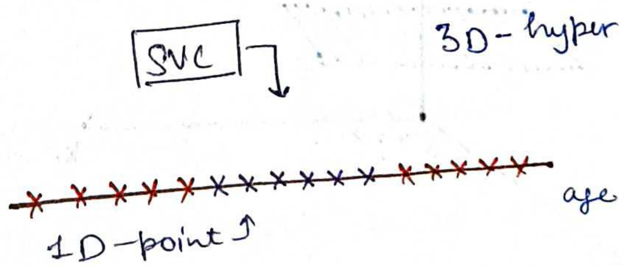


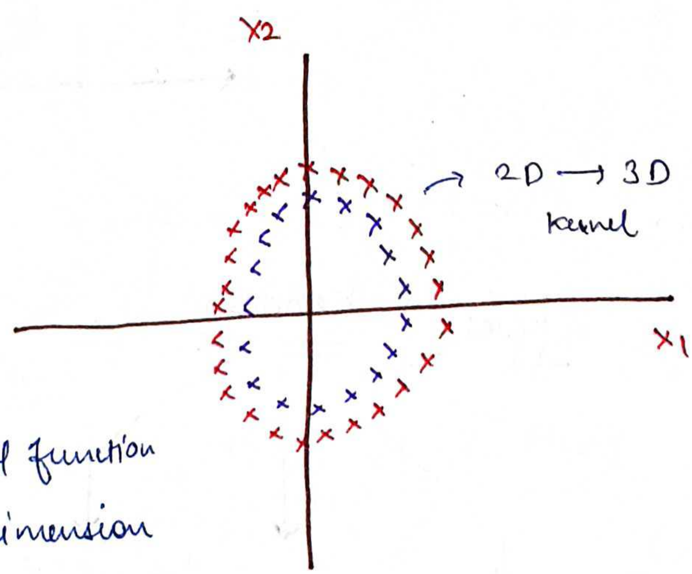
Soft Margin SVC failed in Non-linear Data

Kernel Intuition

age	side effect
72	1
32	0
26	0
5	1

1D - point
2D - line
3D - hyper





Kernel → Mathematical function

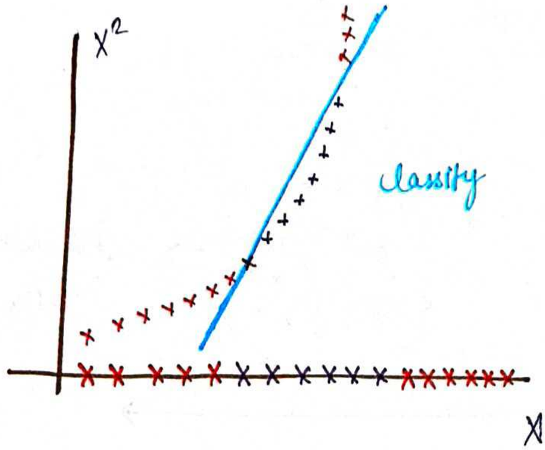
1.) input → higher dimension

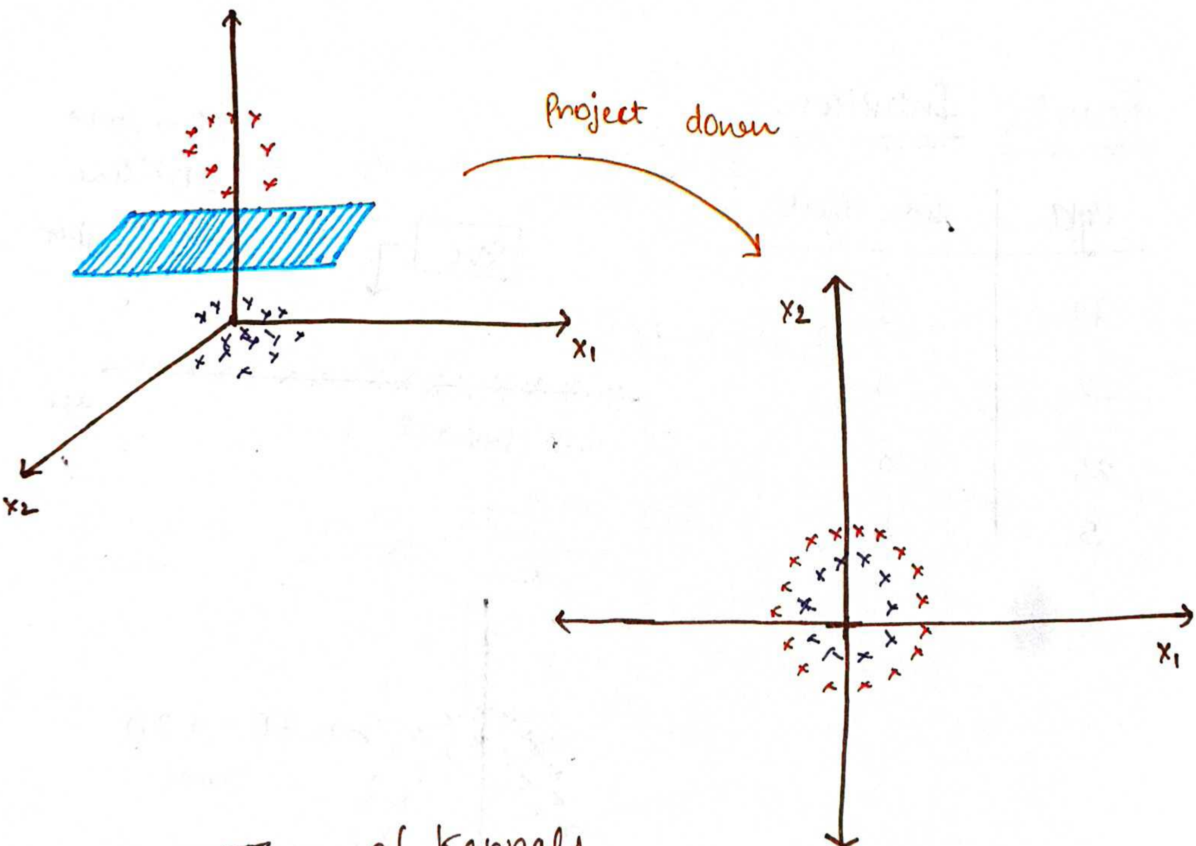
2.) linear seperable

3.) SVC

4.) Project down.

kernel = x^2





Types of Kernals

- linear
- rbf
- poly
- sigmoid

Mathematics of SVM

{ constrained optimization problem }

where we study optimization prob

Optimization prob

↓

linear reg

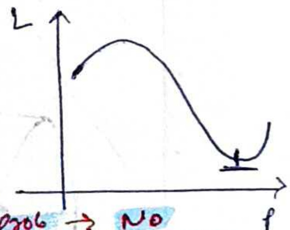
OLS | GD

Logistic Reg

$$L = \underset{(\beta)}{\operatorname{argmin}} \ \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \ \uparrow \min$$

constraint

$$\rightarrow \boxed{\frac{\partial L}{\partial \beta} = 0}$$



in optimization prob → No constraint

constraint optimization problem. (constraint means restriction)

$$\underset{A,B,C}{\operatorname{argmin}} \ \boxed{\frac{\sqrt{A^2+B^2}}{2}} + C\frac{1}{n}\sum_{i=1}^{n} \xi_i \quad \text{such that} \quad \boxed{y_i(Ax_{1i} + Bx_{2i} + C) \geq 1 - \xi_i}$$

Contrained optimization proble

simple example of this

min or max

$$\underset{x,y}{\operatorname{arg\,max}} \ \boxed{x^2 y} \quad \text{such that} \quad \boxed{x^2 + y^2 = 1}$$

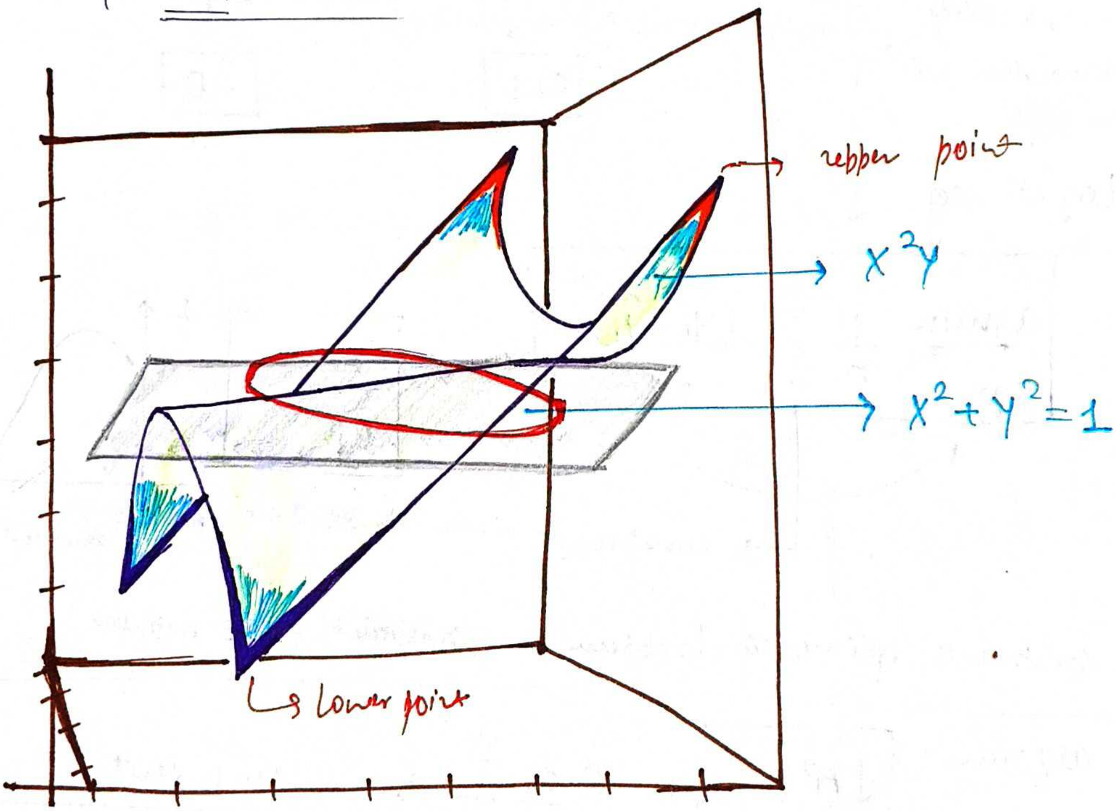
constrain

* for x^2y is max, we need x, y and x, y satisfy eqn $(x^2 + y^2 = 1)$ eg:-

(i) $X = 5$, $Y = 7$ $\quad (5^2) + 7 = 175 \quad (5)^2 + (7)^2 \neq 1$ $\boxed{\text{failed}}$

$X = \frac{1}{\sqrt{2}}$, $Y = \frac{1}{\sqrt{2}}$ $\quad \frac{1}{2} \times \frac{1}{\sqrt{2}} = \frac{1}{2\sqrt{2}} \quad \left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{2}}\right)^2 = 1$ $\boxed{\text{Pass}}$

Geometric Intuition → code



Img 1

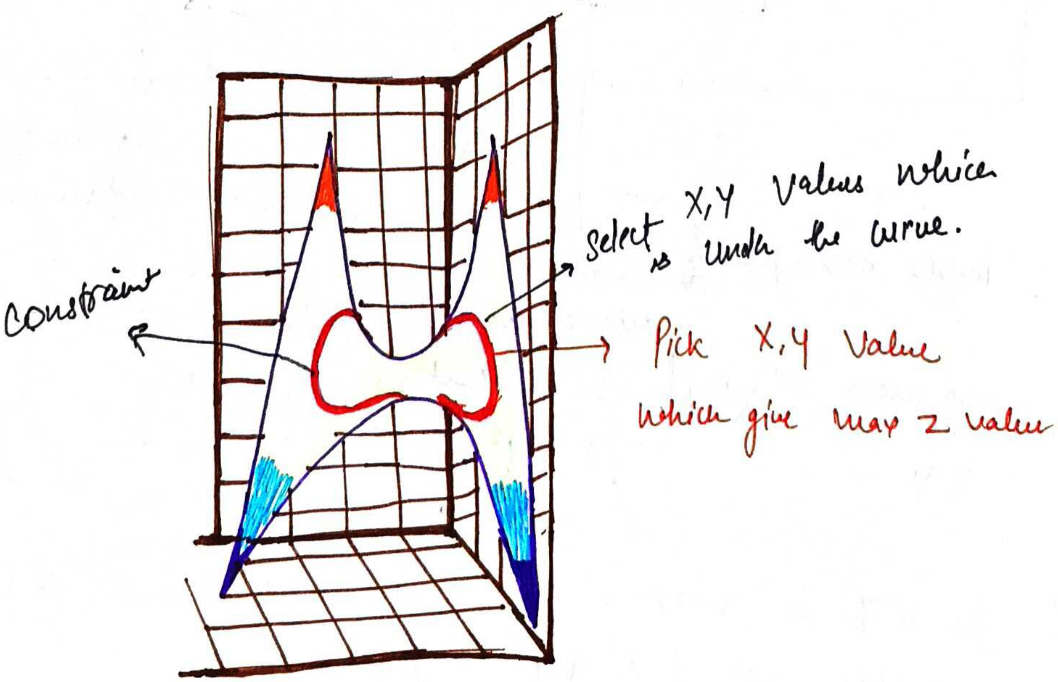
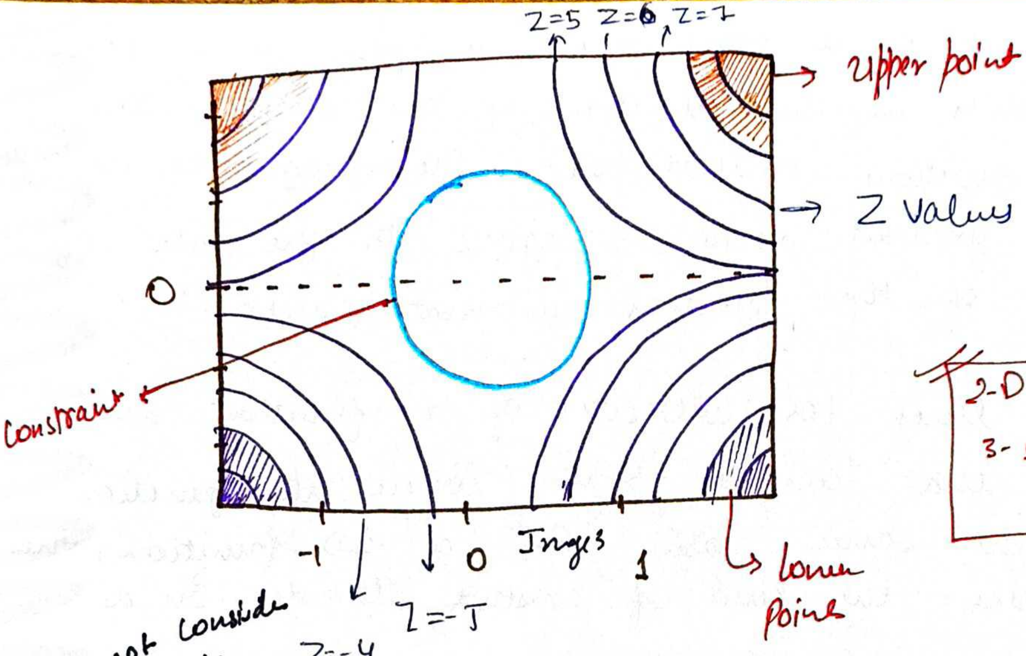
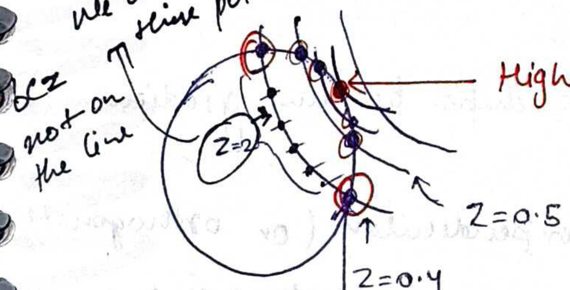
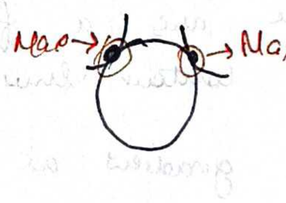


Image 2



2-D → slope.
3-D → Gradient

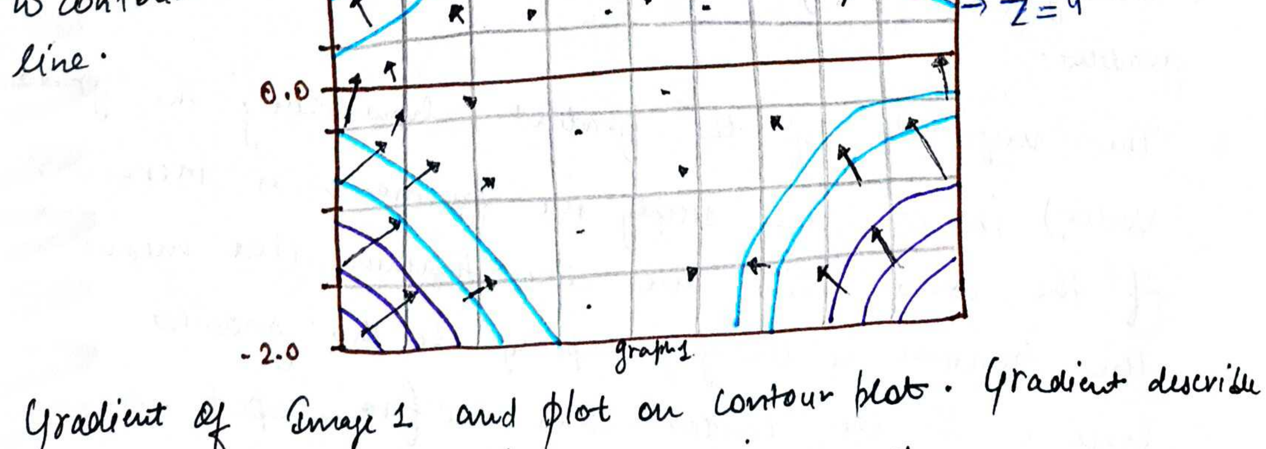




True (X,Y at point is Acceptable)

* points on the circle line are acceptable

Contour plot and Gradient Vector field of $f(x,y) = x^2y$



Gradient of Image 1 and plot on contour plot. Gradient describe the direction of maximum change.

The gradient of a function at a point is a vector that points in the direction of the steepeste ascent of the function at that point. The magnitude (or length) of the gradient vector is equal to the rate of increase of the function in that direction.

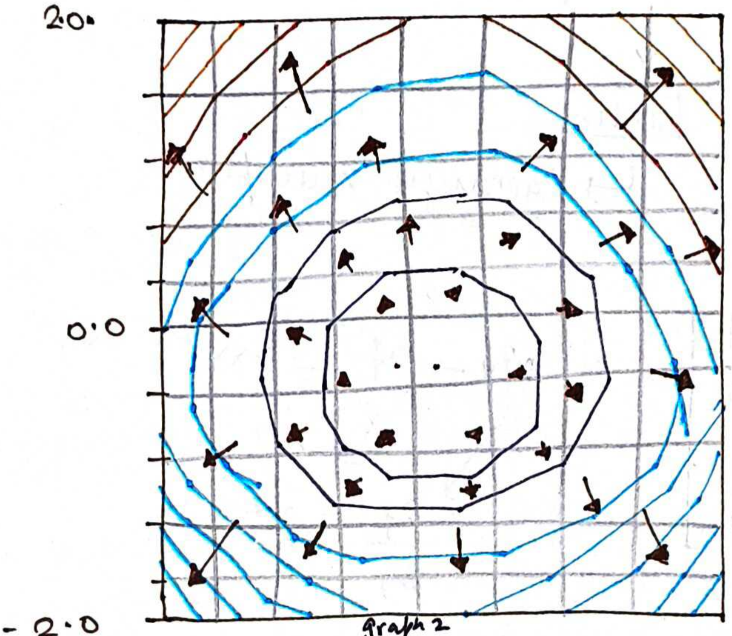
contour lines (or level sets) of a function are curves that connect points where the function has the same value. For a 2D function, these are like the lines of constant altitude on a topographic map.

There are a few key relationships between gradient and contour lines:-

1. The gradient at a point is perpendicular (or orthogonal) to the contour line passing through that point. This is because the contour line represents the direction of maximum change.

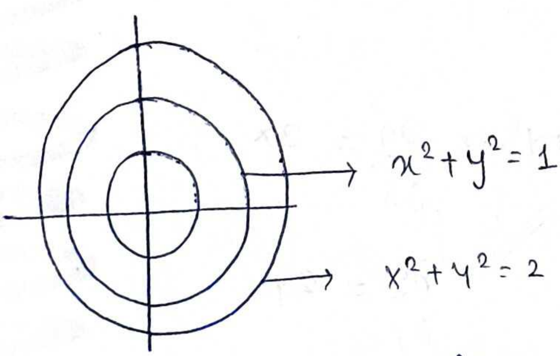
2. The gradient points in the direction where the function increase most rapidly. If you were to walk along the contour line (where the function value doesn't change), the direction you'd need to go to start climbing as steeply as possible is the direction of the gradient.

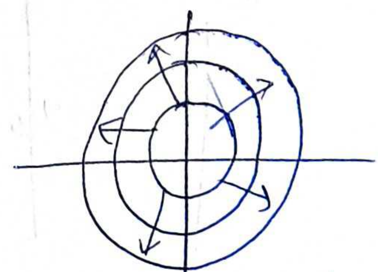
3. The magnitude of the gradient (how long the gradient vector) indicate how steeply the function is increasing. If the contour lines are close together, that means the function is changing rapidly, so the gradient is large. If the contour lines are fare apart, the function is changing slowly, so the gradient is small

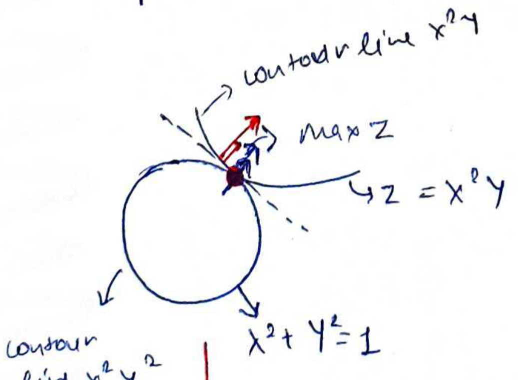


Contour plot and Gradient Descent Vector field of $f(x,y)=x^2+y^2$

contour plot for $g(x,y)=x^2+y^2$







- direction same (arrow direction)
- magnitude difference (arrow magnitude)
- $\boxed{\text{graph 1 and graph 2 compare}}$

$$\underbrace{\nabla f(x,y)}_{f(x,y)=x^2y} = \overset{\text{magnitude}}{\lambda}\, \underbrace{\nabla g(x,y)}_{g(x,y)=x^2+y^2}$$

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

$\downarrow$

$\boxed{\text{Scaler}}$

↳ Lagrange's multiplier

$$\nabla f(x,y) = \begin{bmatrix} \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

$x^2 y \to \frac{\partial f}{\partial x} = 2xy$

$\frac{\partial f}{\partial y} = x^2$

$$\longrightarrow \begin{bmatrix} 2xy \\ x^2 \end{bmatrix}$$

$$\nabla g(x,y) = \begin{bmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial \cancel{x}} \end{bmatrix}$$

$x^2 + y^2 \Rightarrow \frac{\partial g}{\partial x} = 2x$

$\frac{\partial g}{\partial y} = 2y$

$$\longrightarrow \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

$$\begin{bmatrix} 2xy \\ x^2 \end{bmatrix} = \lambda \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

$2xy = \lambda 2x$ — ①

$x^2 = \lambda 2y$ — ②

$x^2 + y^2 = 1$ — ③

from eqⁿ ① $2xy = \lambda \cancel{2x} \Rightarrow \boxed{y = \lambda}$ — ④

eqⁿ ④ put in eqⁿ ② $x^2 = y\,2y \Rightarrow \boxed{x^2 = 2y^2}$ — ⑤

eqⁿ ⑤ put in eqⁿ ③ $2y^2 + y^2 = 1 \Rightarrow \boxed{y^2 = \pm\frac{1}{\sqrt{3}}}$ — ⑥

eqⁿ ⑥ put in eqⁿ ③ $x^2 + \left(\pm\frac{1}{\sqrt{3}}\right) = 1 \Rightarrow \boxed{x^2 = \pm\sqrt{\frac{2}{3}}}$

Possible value

$x = \sqrt{\frac{2}{3}}, -\sqrt{\frac{2}{3}}$

$y = \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}$

$$\begin{array}{cc} x & y \end{array}$$
$$\left(\sqrt{\frac{2}{3}}, \frac{1}{\sqrt{3}}\right), \left(\sqrt{\frac{2}{3}}, -\frac{1}{\sqrt{3}}\right), \left(-\sqrt{\frac{2}{3}}, \frac{1}{\sqrt{3}}\right), \left(-\sqrt{\frac{2}{3}}, -\frac{1}{\sqrt{3}}\right)$$

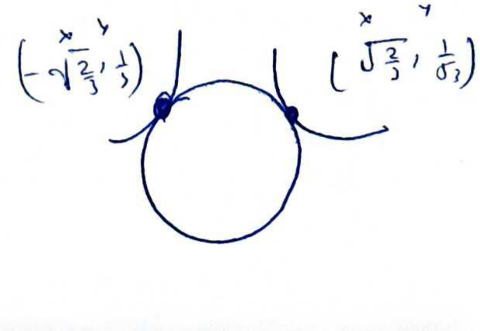
* There's ~~any~~ set which have z max

$z = x^2 y$

$$\begin{array}{cc} x & y \end{array}$$
$\left(\sqrt{\frac{2}{3}}, \frac{1}{\sqrt{3}}\right) = \frac{2}{3} \times \frac{1}{\sqrt{3}} = \frac{2}{3\sqrt{3}}$

$\left(-\sqrt{\frac{2}{3}}, \frac{1}{\sqrt{3}}\right) = \frac{2}{3} \times \frac{1}{\sqrt{3}} = \frac{2}{3\sqrt{3}}$

both z max where given $x^2 + y^2 = 1$



Lagrangian Multiplier → SVM

$$\left[\underset{x,y}{\arg\max} \quad x^2y \quad \text{Such That} \quad x^2+y^2=1\right]$$

$$\nabla \underset{\uparrow\, x^2y}{f(x,y)} = \lambda \underset{\uparrow\, x^2+y^2}{g(x,y)}$$

↓

Optimization prob

Constraint optimization problem convert into optimization problem.

$$\underset{x,y}{\arg\max} \quad f(x,y) - \lambda\,(g(x,y)-1)$$

$$\left[L(x,y,\lambda) = \underset{x,y}{\arg\max} \quad \underbrace{\underset{\uparrow\, x^2y}{f(x,y)} - \lambda\,(\underset{\uparrow\, x^2+y^2}{g(x,y)}-1)}\right]$$

↓ How convert

$$\underbrace{\frac{\partial L}{\partial x} = 0 \qquad \frac{\partial L}{\partial y} = 0 \qquad \frac{\partial L}{\partial \lambda} = 0}_{\text{for findig eqn equal to 0.}}$$

w.r.t x

$$\frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} = 0$$

$$= \frac{\partial f}{\partial \lambda} - \partial \frac{\partial g}{\partial \lambda}$$

$$= \boxed{2xy - \lambda 2x}$$ same as eqn ① prooved

w.r.t y

$$\frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} = 0$$

$$x^2 - \lambda 2y = 0$$

$$\boxed{x^2 = \lambda 2y}$$ same as eqn ② Prooved

w.r.t λ

$$g(x,y) - 1 = 0$$

$$g(x,y) = 1$$

$$x^2 + y^2 = 1$$ prooved

$$\underset{x,y}{\arg\max} \; x^2 y \quad ST \quad x^2+y^2=1$$

Constrained optimisation

↓

optimisation

$$\boxed{L(x,y,\lambda) = \underset{x,y,\lambda}{\arg\max} \; x^2y - \lambda(x^2+y^2-1)}$$

↳ Langrange multiplier

N-dim SVM Soft margin formation

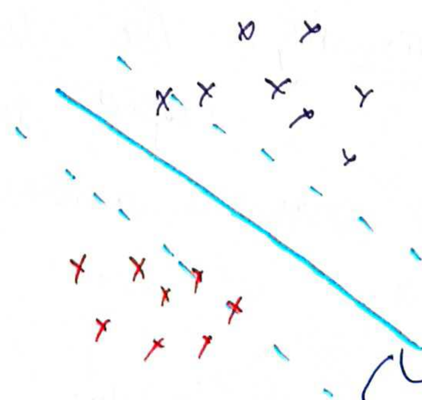
2 dims → line

$\boxed{Ax + By + C}$

$\boxed{A, B, C}$

n- dim

x_1	x_2	x_3	----	x_n	y
w_1	w_2	w_3	----	w_n	



↳ hyperplane

Soft margin

$$\underset{A,B,C}{\arg\min} \; \frac{\sqrt{A^2+B^2}}{2} + C\sum_{i=1}^{n} \xi_i$$

Such that

$$y_i(Ax_{1i} + Bx_{2i} + C) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

$$w_1x_1 + w_2x_2 + w_3x_3 \; ---- \; w_nx_n + b = 0$$

↳ $\boxed{W^T x + b = 0}$

W, b ↳ intercept

coefficient

$$\frac{\sqrt{A^2+B^2}}{2}$$

$$\sqrt{w_1^2+w_2^2+w_3^2 - -}$$

$$\boxed{\underset{w,b}{\operatorname{argmin}} \frac{\|w\|^2}{2} + C\sum_{i=1}^{n} \xi_i \quad \text{such that} \quad y_i(w^Tx+b) \geq 1-\xi_i}$$

↳ n-dim soft margin formation

Karush Kuhn Tucker Conditions (KKT conditions)

They generalize the method of lagrange multipliers to handle inequality constraints. In the content of support vector Machine (SVMs) and many other optimization problems, the KKT condition play a key role in deriving the dual problem from the primal problem.

The KKT condition are:

1. **Stationarity**: The derivaties of the lagrangian with respect to the primal variables, the dual variables associated with inequality constraints ~~and~~ ~~that~~ are all zero.

2. **Primal feasibility**: All the primal constraints are satisfied.

3. **Dual feasibility**: All the dual variables associated with inequality constraint are nonnegative.

4. <u>Complementary slackness</u>: The product of each dual variable and its associated inequality constraint is zero. This means that at the optimal solution, for each constraint, either the constraint is active (equality holds) and the dual variable can be non zero, or the constraint is active (strict inequality holds) and the dual variable is zero

(KKT condition) ← SVM dual problem

$$\boxed{\min_{x} \quad f(x) = x^2 \quad \text{such that } x - 1 \le 0}$$

Primal form

Inequality

$$\boxed{f(x) = x^2 \quad \min_{x} \qquad x - 1 \le 0}$$

convert

Langrangian form

$$\boxed{L(x, \lambda) = x^2 - \lambda(x-1)}$$

<u>Condition</u> :- 1. Stationarity :- differentiate w.r.t x and λ

$$\frac{\partial L}{\partial x} = 0 \qquad \frac{\partial L}{\partial \lambda} = 0$$

2. Primal feasibility :- $\boxed{x - 1 \leq 0}$

3. Dual feasibility : $\boxed{\lambda \geq 0}$

4. Complementary slackness: $\boxed{\lambda(x-1) = 0}$

* Through langrangian form, we can find the value x and λ. Apply all 4 condition on x, λ, all condition is true or false

<u>Example</u>

$$\boxed{\min_{x,y} \quad f(x,y) = x^2 + y^2 \qquad \text{subject to } x + y - 1 \leq 0}$$

↳ langrangian form

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

Condition :- ① $\frac{\partial L}{\partial x} = 0 \quad \frac{\partial L}{\partial y} = 0 \quad \frac{\partial L}{\partial \lambda} = 0$

② $x + y - 1 \leq 0$

③ $\lambda \geq 0$

④ $\lambda(x + y - 1) = 0$

$$\frac{\partial L}{\partial x} = 0 \Rightarrow \boxed{2x - \lambda = 0} \qquad \frac{\partial L}{\partial y} = 0 \Rightarrow \boxed{2y - \lambda = 0}$$

$$\frac{\partial L}{\partial \lambda} = -x - y + 1 = 0$$

$$\boxed{x + y = 1}$$

$x = \frac{\lambda}{2}$ $\quad y = \frac{\lambda}{2}$ $\quad x + y = 1$

$$\frac{\lambda}{2} + \frac{\lambda}{2} = 1 \Rightarrow \boxed{\lambda = 1}$$

$x = \frac{1}{2}$ $\quad y = \frac{1}{2}$ $\quad \boxed{\lambda = 1}$

$\boxed{x = 0.5}$ $\quad \boxed{y = 0.5}$

Condition = 2 $\quad x + y - 1 \leq 0$

$0.5 + 0.5 - 1 \leq 0$

$1 - 1 \leq 0$

$0 = 0$ True

Condition 3 = $\quad \lambda \geq 0$

$\cancel{1} \geq 0$ True

Condition 4 = $\quad \lambda(x + y - 1) = 0$

$1(0.5 + 0.5 - 1) = 0$

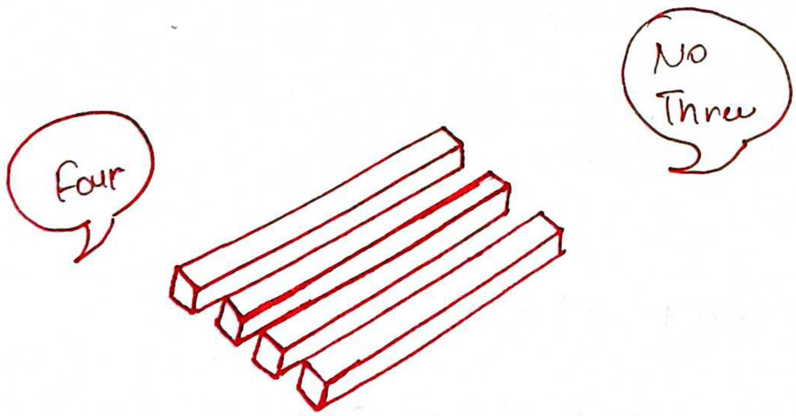
$1(1 - 1) = 0$

$0 = 0$ True

$\boxed{x = 0.5 \quad y = 0.5}$ → satisfy our eqn

Concept of Duality

The duality principle fundamental in optimization theory. It provides a powerful tool for solving difficult or complex optimization problem by transform them into a simpler or easier-to-solve problems. The solution to the dual problem. The solution to the dual problem provides a lower bound on the solution of the primal problem. If strong duality holds (i.e the optimal value of the primal and and dual problem are equal), then solving the dual problem can directly give the solution to the primal problem.



The primal problem is the original optimization problem that you are trying to solve. It involves finding the minimum or maximum of a objective function. Subject to certain constraint.

The dual problem is a related optimization problem that is derived from the primal problem. It provides a lower or upper bound on the solⁿ to the primal prob.

SVM (w, b)

derive $\Rightarrow$ primal $\to$ dual

SVM Dual Problem

The primal form of hard margin SVM is given by:

$$\text{minimize}_{w,b} \ \frac{1}{2}\|w\|^2$$

$$\text{subject to } y_i(w \cdot x_i - b) \geq 1, \quad i = 1, 2, \ldots, n$$

$w, b \to$ primal variable

dual form Hard margin $\to$

$$\text{maximize}_{\underline{a}} \ \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \underbrace{y_i y_j}_{\text{output}} a_i a_j \underbrace{(x_i \cdot x_j)}_{\text{training data}}$$

dual variable

$$\text{subject to } a_i \geq 0, \quad i = 1, \ldots, n$$

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

The primal form of soft margin SVM is given by:

$$\text{minimize}_{w, b, \xi} \ \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i$$

$$\text{subject to } y_i(w \cdot x_i - b) \geq 1 - \xi_i, \quad i = 1, \ldots, n$$

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

Zeta

Dual form of Soft Margin

$$\text{maximize}_a \quad \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j a_i a_j (x_i \cdot x_j)$$

subject to $\quad 0 \le a_i \le C \qquad i = 1, \dots, n$

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

Dual problem Derivation

Primal form

$$\underset{w,b}{\text{argmin}} \ \frac{\|w\|^2}{2} \quad \text{such that} \quad \underbrace{y_i (w^T x_i + b)}_{} \ge 1 \ \forall i$$

↳ for every row → 1 constraint

$$L(w,b,\alpha) = \frac{\|w\|^2}{2} - \alpha_1 [y_1 (w^T x_1 + b) - 1] - \alpha_2 [y_2 (w^T x_2 + b) - 1] \dots - \alpha_n [y_n [w^T x_n + b - 1]$$

↓

$$L(w,b,\alpha) = \frac{\|w\|^2}{2} - \sum_{i=1}^{n} \alpha_i [y_i (w^T x_i + b) - 1]$$

↓

$$\frac{\partial L}{\partial w} = 0 \qquad \frac{\partial L}{\partial b} = 0$$

$$L(w, b, \alpha) = \frac{\|w\|^2}{2} - \sum_{i=1}^{n} \alpha_i y_i w^T x_i + \alpha_i y_i b - \alpha_i$$

$$L(w, b, \alpha) = \frac{\|w\|^2}{2} - \sum_{i=1}^{n} \alpha_i y_i w \cdot x_i - \sum_{i=1}^{n} \alpha_i y_i b + \sum_{i=1}^{n} \alpha_i \quad -(1)$$

differentiate w.r.t w

$$\frac{\partial L}{\partial w} = \frac{2w}{2} - \sum_{i=1}^{n} \alpha_i y_i x_i = 0$$

$$\boxed{w = \sum_{i=1}^{n} \alpha_i y_i x_i}$$

differentiate w.r.t b

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

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

Putting value in eqn (1)

$$= \frac{1}{2}\left(\sum_{i=1}^{n} \alpha_i y_i x_i\right)\left(\sum_{j=1}^{n} \alpha_j y_j x_j\right) - \sum_{i=1}^{n} \alpha_i y_i x_i \left(\sum_{j=1}^{n} \alpha_j y_j x_j\right) + \sum_{i=1}^{n} \alpha_i$$

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

$$\boxed{= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)} \rightarrow \text{dual form}$$

Constraint

$\alpha_i \geq 0$ (KKT condition)

$\sum_{i=1}^{n} (\alpha_i y_i = 0)$ above defined

* we can not write complementary skewness because their use primal but use during finding solⁿ.

$$\left\{ \begin{array}{ccc} \min & \rightarrow & \max \\ \max & \rightarrow & \min \end{array} \right\}$$ dual form

Observation

The primal form of Hard margin SVM is given by:

$$\text{minimize}_{w,b} \ \frac{1}{2} \|w\|^2$$

subject to $y_i(w \cdot x_i - b) \geq 1, \quad i = 1, \ldots, n$

$$\text{maximize}_{a} \ \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j a_i a_j (x_i x_j)$$

subject to $a_i > 0, \quad i = 1, \ldots, n$

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

* Alpha only for support vectors → the eqn is not as dangerous as it seems

* easy to solve

* Dot product

* kernel friendly.

$W = \alpha_1 Y_1 X_1 + \alpha_2 Y_2 X_2 \ldots \alpha_n Y_n X_n$

↳ Acc to all points for (margin) (Hard margin SVM)

dual → Only take support vector for margin

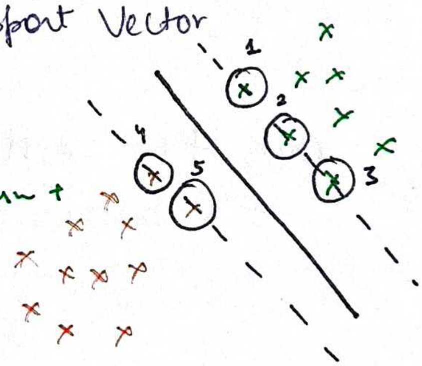
SVM Dual Formulation

$$\max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \underbrace{\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i x_j)}_{\text{make grid while solving}}$$

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

$\alpha_i = 0$ for all non support vectors

$\alpha_i > 0$ for all Support Vector



$$\max_{\alpha_i} \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 - \frac{1}{2}[\text{1 term} + \text{2 term} + \ldots + \text{25 term}]$$

if 5 support vectors then 25 term form

after solving Green line solution.

	X_1	X_2	Y
Considered those points which is support vector	X_{11}	X_{12}	1
	X_{21}	X_{12}	-1
	X_{31}	X_{32}	
	.	.	
	.	.	

$x_i \cdot x_j = X_{11} \cdot X_{12} = X_{11}X_{11} + X_{12}X_{12}$

$y_i \cdot y_j = Y_1 \cdot Y_1 = 1 \times 1$

$x_i \cdot x_j = \cancel{X_{21} \cdot X_{22}} = X_{21} \cdot X_{21} + X_{22} \cdot X_{22}$

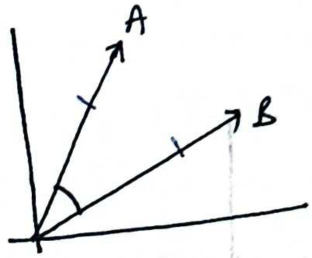
$y_i \cdot y_j = Y_2 \cdot Y_2 = (-1) \times (-1)$

(j) → 5 support vector

(i)	1	2	3	4	5
1	-	-	-	-	-
2	-	-	-	-	-
3	-	-	-	-	-
4	-	-	-	-	-
5	-	-	-	-	-

The Similarity Perspective

$$\text{cosine similarity} = \frac{A \cdot B}{\|A\| \, \|B\|}$$



$\|A\| = 1$ $\|B\| = 2$

$= \boxed{A \cdot B} \rightarrow$ dot product of A and B

* find the similarity between two vectors we use cosine similarity

$$\underbrace{\max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \boxed{(x_i \cdot x_j)}}$$

$\rightarrow$ similarity

$\rightarrow$ maximize the similarity of SV based on main sign

Kernel SVM

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

$\rightarrow$ Kernel SVM

$K(x_i, x_j) \rightarrow$ Kernel $\rightarrow$ similar between x_i and x_j

$\rightarrow \boxed{x_i \cdot x_j}$ $\rightarrow \boxed{x_i \cdot x_j}$ $\rightarrow$ Creative Version

$\rightarrow$ Linear SVM

$\rightarrow$ Polynomail

$\rightarrow$ RBF