Relationship with logistic Regression

$$\underset{\beta_0\, \beta_1\, \beta_2}{\arg\min}\ \frac{\sqrt{\beta_1^2+\beta_2^2}}{2} + C\frac{1}{n}\sum_{i=1}^{n}\xi_i \quad \text{compare}$$

L2 Reg $\quad \text{logloss} + \lambda\left(\sqrt{\beta_1^2+\beta_2^2}\right)$

$$\boxed{\lambda \propto \frac{1}{C}}$$

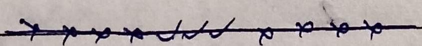
Session-3

→ The problem with SVC/Soft margin SVM and Hard margin SVM is that they only work with linear dataset.

To solve it we have SVM

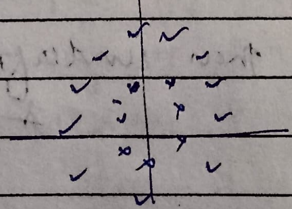
SVM = SVC + Kernels

1)

→ 

↳ It is 1D. So we have to classify the data using a point. Not possible

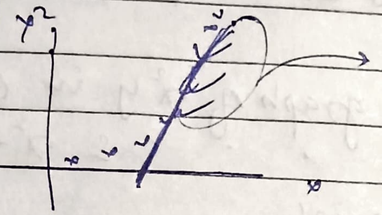
Similarly

2)  → we can't classify it using a line

to Solve it we have kernels.

1) kernels takes input data to higher dimension
2) Data become linearly separable in higher dimension
3) Apply SVC there
4) Project data in lower dimension

- applying kernel x^2 in 1)



Now by using a line we can classify them

in 2) apply kernel $x_1^2 + x_2^2$ or $e^{-(x_1^2 + x_2^2)}$

- Types of kernels –
  - ↳ linear
  - ↳ poly
  - ↳ rbf
  - ↳ Sigmoid

- It is called kernel trick not kernel transformation as we actually not convert data to higher dimension, we do all these things in lower dimension. This saves computation.

→ Mathematics of SVM –

- optimization problem –

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

↳ we have to minimize the loss f^n.
This is optimization problem.

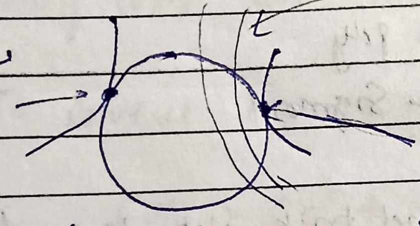
But when we have to do optimizatn and Constraint is given then it will be Constrained optimizatn problem

If $\underset{x,y}{\operatorname{argmax}}\ x^2 y$ such that $x^2+y^2=1$

we want to solve this.

→ we draw graph of x^2y in calcs.
→ " " " " & $x^2+y^2=1$ in same graph.
→ we want that values of x, y for which z is max.

These lines are coming from contour graph



x^2+y^2

This is also

at this pt we got max value of z which satisfies constraint.

• The gradiant of a fn at a pt is a vector that pts in the direcn of the steepest ascent of fn at that pt

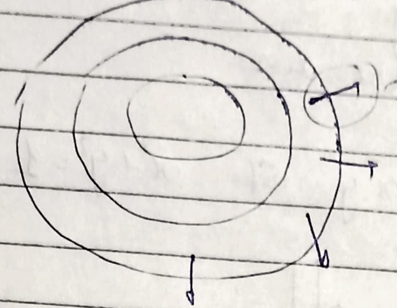
↳ the gradiant at a pt is ⊥ as to the contour line passing through that pt

→ gradient descent works as a slope in SI

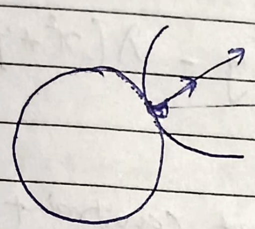
Page No.
Date

- Constraint → $x^2+y^2=1$
  Let $g(x,y) = x^2+y^2$

if we plot contour plot of $g(x,y)$ we get



Now



So. $\nabla f(x,y) = \lambda \nabla g(x,y)$
↳ Lagranges multiplier

$\nabla f(x,y) \Rightarrow \frac{\partial f}{\partial x} = 2xy \quad , \quad \frac{\partial f}{\partial y} = x^2$

$\nabla g(x,y) \Rightarrow \frac{\partial g}{\partial x} = 2x \quad \frac{\partial g}{\partial f} = 2y$

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

$\Rightarrow 2xy = \lambda 2x \qquad \Rightarrow y = \lambda$

$x^2 = 2\lambda y \qquad \Rightarrow x^2 = 2y^2$

$x^2+y^2 = 1 \qquad \Rightarrow 2y^2+y^2 = 1$

$y = \pm \frac{1}{\sqrt{3}} = \lambda$

$x = \pm \sqrt{\frac{2}{3}}$

Page No.	
Date	

(1) (2) (3) (4)

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

$z = x^2 y$

1, 3 are ans

$\Rightarrow \underset{x,y}{\text{argmax}}\ x^2 y$ st $x^2 + y^2 = 1$ [constraint opn]

$\Downarrow$

$L(x, y, \lambda) = \underset{x,y,\lambda}{\text{argmax}}\ x^2 y - \lambda(x^2 + y^2 - 1)$

(optimize without constraint)

- Session - 4 :-

→ SVM in n-dimension

x_1	x_2	---	x_n	y
w_1	w_2	. .	w_n	

Eqn of hyperplane $\Rightarrow w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b = 0$

$w^T x + b = 0$

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

Such that

$y_i(A x_{1i} + B x_{2i} + c) \geq 1 - \xi_i$

$\xi_i \geq 0$

In n-dim. the eqⁿ will be -

$$\underset{w,b}{\operatorname{argmin}} \frac{\|w\|}{2} + C \sum_{i=1}^{n} \xi_i$$

Such that

$$y_i(w^T x + b) \geq 1 - \xi_i$$

$$\|w\| = \sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}$$

for convenience we will write $\frac{\|w\|^2}{2}$

→ Constrained optimizatⁿ problem with inequality -

$$\min_{x} f(x) = x^2 \text{ Such that } x - 1 \leq 0$$

To solve it we use Karush Kuhn Tucker condtⁿ (KKT)

↳ In this with Lagrangian we also check KKT condtⁿs

→ $f(x) = x^2$ $\left.\begin{matrix} \min x \\ \text{s.t. } x - 1 \leq 0 \end{matrix}\right\}$ - It is primal form

By using this form we calculate dual form

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

~~Condtⁿ~~ I :-

Page No.
Date

Conditn 1 :- derivative of Lagrangian w.r.t primal variable & Lagrange is zero

① $\frac{\partial L}{\partial x} = 0$ & $\frac{\partial L}{\partial \lambda} = 0$

Condⁿ 2 - Primal feasibility -
primal constraint are Satisfied
$x - 1 \leq 0$

Condⁿ 3 - Dual or feasibility -
dual variable must be non-ve
$\lambda \geq 0$

Condⁿ - 4 $\lambda(x-1) = 0$
product of dual variable & constraint is equal to zero.

→ The value of x, λ which Satisfies all these conditns are answers.

eg :- $f(x,y) = x^2 + y^2$ $x + y - 1 \leq 0$
minimize x, y

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

$\frac{\partial L}{\partial x} = 2x - \lambda = 0$ ⇒ $x = \lambda/2$

$\frac{\partial L}{\partial y} = 2y - \lambda = 0$ ⇒ $y = \lambda/2$

$\frac{\partial L}{\partial \lambda}$ ⇒ $x + y = 1$ ⇒ $\lambda = 1$
$x = 0.5$
$y = 0.5$

Page No.	
Date	

(2) $x + y - 1 \leq 0$

$0 \leq 0$

(3) $\lambda \geq 0$

(3) $\lambda(x + y - 1) = 0$

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

- Concept of duality -

↳ we have given a complex optimizatⁿ problem (primal form). we convert that into a form which is easier to solve (dual form). If duality is strong i.e strong relatⁿship b/w primal and dual form then solving dual form can even give us the solⁿ of primal form directly

- Primal form of Hard margin SVM

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

$$\text{Subject to} \quad y_i(w \cdot x_i - b) \geq 1 \qquad i = 1, \dots n$$

Dual form of Hard margin svm

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

$$\text{Subject to} \quad a_i \geq 0 \qquad i = 1, \dots n$$

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

Page No.
Date

· Soft margin svm primal form -

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

Subject to $y_i(w.x_i - b) \geq 1 - \xi_i \quad i = 1 \dots n$

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

Dual form -

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

Subject to

$1 \leq a_i \leq C \quad i = 1, \dots n$

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

→ Dual form derivatⁿ of Hard margin svm -

$x_1, x_2 \dots x_m \mid y$

n rows, m columns

Svm constraint $y_i(w^T x_i + b) \geq 1 \ \forall i$

Its actually n constraint

$\boxed{w^T x_i = w \cdot x_i}$

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

$$\vdots$$

$$- \alpha_n[y_n(w.x_2 + b) - 1]$$

Page No.
Date

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

$$= \frac{||w||^2}{2} - \sum_{i=1}^{n} \alpha_i y_i w \cdot x_i - \sum_{i=1}^{n} \alpha_i y_i b + \sum_{i=1}^{n} \alpha_i$$

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

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

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

So

$$L(w,b,\alpha) = \frac{1}{2}\left(\sum_{i=1}^{n} \alpha_i y_i x_i\right)\left(\sum_{j=1}^{n} \alpha_j y_j x_j\right) - \left[\sum_{i=1}^{n} \alpha_i y_i x_i\right]\left(\sum_{i=1}^{n} \alpha_j y_j x_j\right)$$

$$- 0 + \sum_{i=1}^{n} \alpha_i$$

this comes from $\sum_{i=1}^{n} \alpha_i y_i w \cdot x_i \Rightarrow \sum_{i=1}^{n} \alpha_i y_i x_i \cdot w_i$

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

Subject to

$\alpha_i \geq 0$

$\sum \alpha_i y_i = 0$

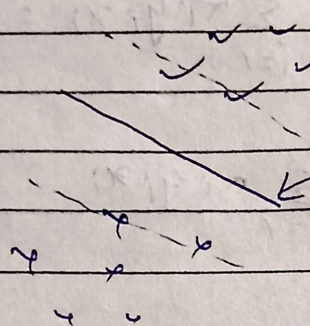
this is dual form

→ why dual form is good –

i) its easy to solve both mathematically & computationally

ii) we can apply kernal here

<u>Observatn</u> –



ξ_i $wx + b = 0$

here w will only depend upon support vector

& we know $w = \alpha_1 y_1 x_1 + \alpha_2 y_2 x_2 + \ldots + \alpha_n y_n x_n$

Suppose $x_1 y_1$ & $x_2 y_2$ are support vector then only α_1 & α_2 is non zero & all other α's are zero.

So for support vector $\alpha > 0$
& for non " " $\alpha = 0$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j)$$

→ it will be non zero only when we have support vector else it will be zero so computations are saved.