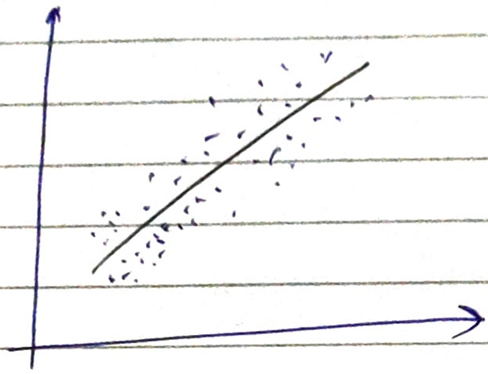
Machine Learning

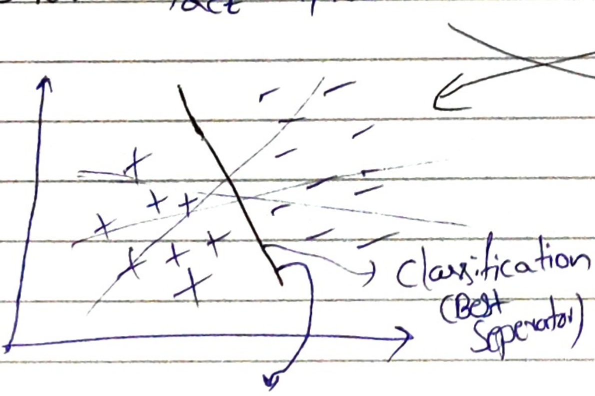
Linear Reg

① Line that bests fits the data



$$Y_{pred} = mx + c$$
$$= w^T x$$

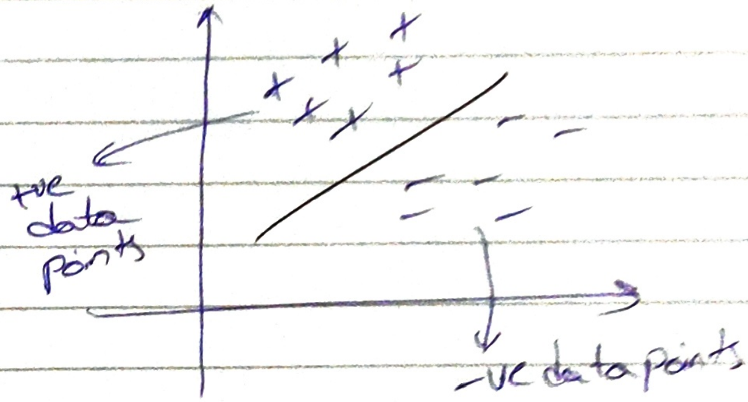
② error = $Y_{act} - Y_{pred}$



Out of all the lines this lines will seperate at best.

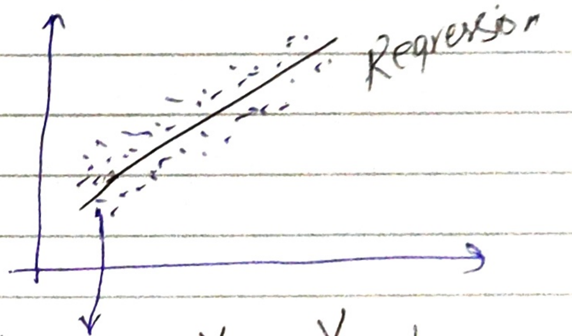
Logistic Reg

① Line that best Seperates the data



$$Y_{pred} = \text{Sign}\begin{pmatrix} mx + c \\ w^T x \end{pmatrix}$$

② Misclassification = $Y_{act} \neq Y_{pred}$



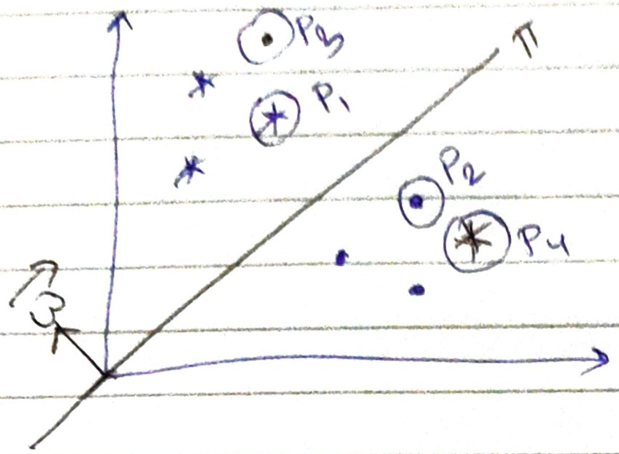
Min error = $Y_{act} - Y_{pred}$

↳ minimize the sum of errors

= min Σ error

Logistic Regression

actual data points $\left\{ \begin{array}{l} * \rightarrow 1 \\ \bullet \rightarrow -1 \end{array} \right\}$



$\underline{P_1} \Rightarrow Y_{act} * Y_{pred}$

$= (+1) * \vec{w} \cdot \vec{P_1}$

↳ angle b/w $\vec{w}$ & $\vec{P_1}$ is $< 90°$

$\Rightarrow \cos\theta_{< 90°} \Rightarrow +ve$

$= (+1) * (+ve)$

$= +ve$

$P_2 = Y_{act} * Y_{pred}$

$= (-1) * (\vec{w} \cdot \vec{P_2})$

↳ $\cos(\theta)_{w, P_2} \rightarrow -ve$

$= (-1) * (-1)$

$= +ve$

$P_3 \Rightarrow Y_{act} * Y_{pred}$

$= (-1) * w \cdot P_3$

↳ $\cos\theta_{w, P_3}$

$= (-1) * (+ve)$

$= -ve$

$P_4 \Rightarrow Y_{act} * Y_{pred}$

$= (+1) * w \cdot P_4$

↳ $\cos\theta_{w, P_4} \Rightarrow -ve$

$= (+1)(-ve)$

$= \underline{\underline{-ve}}$

$P_1, P_2 \rightarrow$ Correctly classified

$P_3, P_4 \rightarrow$ Not Correctly classified.

If the line is correctly classified then we need to maximize "$Y_{act} * Y_{pred}$."

→ We need to Maximize $Y_{act} * Y_{pred}$

$$= \text{Max} \sum Y_{act} * Y_{pred}$$

$$m^*, c^* = \text{Max} \sum Y_{act} * \underbrace{(mx + c)}_{\to\, \omega^T x + \omega_0}$$

$$m^*, c^* = \underset{m,c}{\text{arg}} \ \text{Max} \sum Y_{act} * (\omega^T x + \omega_0)$$

$$m, c \to \omega, \omega_0$$

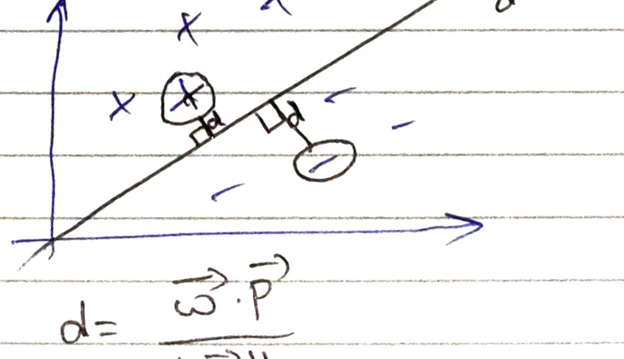
$$\omega^*, \omega_0^* = \underset{\omega,c}{\text{arg}} \ \text{Max} \sum \Big(Y_{act} * \underbrace{(\omega^T x + \omega_0)}_{Y_{pred}}$$

$\omega \to$ May be a Vector

$\omega_0 \to$ Scalar Quantity

$Y_{act} * Y_{pred} = \vec{\omega} \cdot \vec{p}$

$\vec{p}$ is at a distance of $(\vec{\omega} \cdot \vec{p})$ from the plane Π_d



$$d = \frac{\vec{\omega} \cdot \vec{p}}{\|\vec{\omega}\|}$$

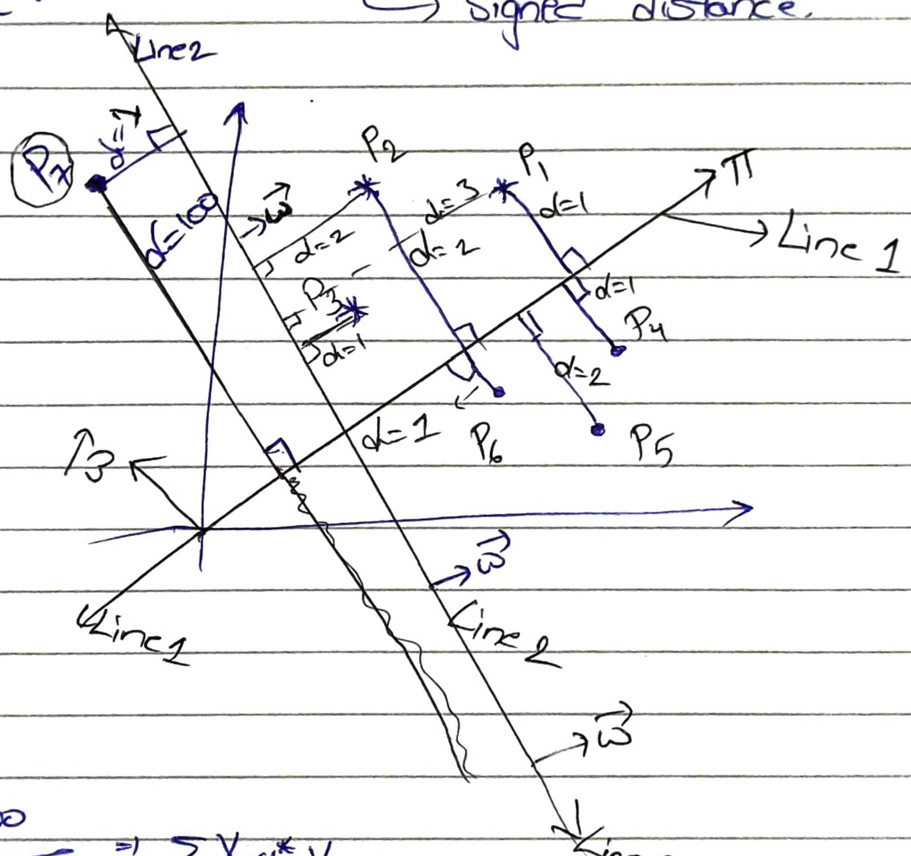
$Y_{act} * Y_{pred}$ ⟶ Signed distance (it has $\cos\theta$ → gives sign & distance (Y_{act} Y_{pred}) ⇓ distance)

↳ +ve ⇒ Correctly Classified

↳ -ve ⇒ Mis classified

$$\Rightarrow m^*, c^* = \underset{m,c}{\arg\max} \sum (Y_{act} * Y_{pred})$$

↳ signed distance.



Line 1:

$\sum Y_{act} * Y_{pred}$

$P_1 = (+1) * (1) = 1$

$P_2 = (+1) * (2) = 2$

$P_3 = (+1) * (1) = 1$

$P_4 = (-1) * (-1) = 1$

$P_5 = (-1) * (-2) = 2$

$P_6 = (-1) * (-1) = 1$

$P_7 = (-1) * (100) = -100$

$\underline{-92}$ ⇒ $\sum Y_{act} * Y_{pred} = -92$

Line 2:

$\sum Y_{act} * Y_{pred}$

$P_1 = (+1) * (3) = 3$ $\quad$ $P_4 = (-1) * (3) = -3$ $\quad$ $P_7 = (-1) * (-1) = 1$

$P_2 = (+1) * (2) = 2$ $\quad$ $P_5 = (-1) * (2) = -2$ $\quad$ $\sum P_1 + P_2 + \cdots P_7 = 1$

$P_3 = (+1) * (1) = 1$ $\quad$ $P_6 = (-1) * (1) = -1$ $\quad$ $= \sum Y_{act} * Y_{pred} = 1$

Line 1	Line 2
→ 1 Misclassified points	→ 3 Misclassified points
→ $\Sigma Y_{act} * Y_{pred} = -92$	→ $\Sigma Y_{act} * Y_{pred} = -1$

→ Line 1 is good because of 1 misclassified point but my Equation is telling me Line 2 is good
↓
$\Sigma Y_{act} * Y_{pred}$

→ Because this is impacted by outliers.

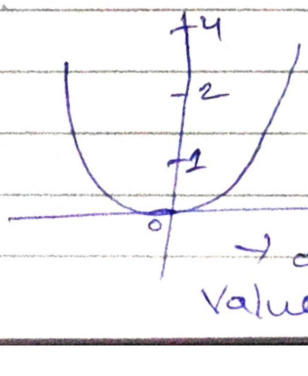
Linear Reg → $\text{Min } \Sigma e^2 \Rightarrow e = Y_{act} - Y_{pred}$

Logistic Reg → Max Σ Signed distance
↳ $Y_{act} * Y_{pred}$

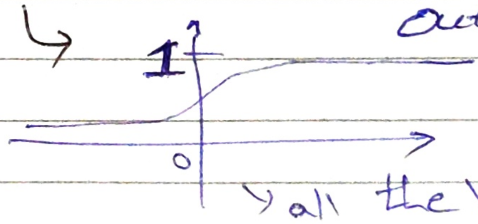
↓ we need to Maximize this equation but considering every point Outliers.

For that we need to apply some ← function like x^2, $\log x$, sigmoid functions to handle outliers.

x^2 →



→ all the Values are increasing rapidly



→ all the Values are Converted to 0 & 1

Sigmoid Function =

$\sigma(f) = g(f) = \frac{1}{1+e^{-f}} \Rightarrow$ Values from 0 to 1

↳ So, if we apply sigmoid function to every data point and Convert it to in the range 0 & 1 then it will clear our problem of outliers.

$$= \text{Max} \sum Y_{act} * Y_{pred}$$

$$= \text{Max} \sum Y_i * w^T x_i$$

Now apply Sigmoid function

$$= \text{Max} \sum \sigma(Y_i * w^T x_i)$$

$$= \text{Max} \sum \frac{1}{1+e^{-Y_i * w^T x_i}}$$

$$w^* = \underset{w}{\text{arg Max}} \sum \frac{1}{1+e^{-Y_i * w^T x_i}}$$

$$\text{Max}(f) = \text{Max} \log(f)$$

$\boxed{\log \frac{A}{B} = \log A - \log B}$

$$w^* = \underset{w}{\text{arg Max}} \sum \log\left(\frac{1}{1+e^{-Y_i * w^T x_i}}\right)$$

$$\boxed{w^* = \underset{w}{\text{arg Min}} \sum \log\left(1+e^{-Y_i * w^T x_i}\right)}$$

↳ logistic Regression Optimization Eqn