The gradient descent

finally!

Gradient

→ partial derivative
→ vector

e.g: $f = x^2 + y^2$

$$\nabla f_x = \frac{\partial f}{\partial x} \qquad 2x$$
$$\nabla f_y = \frac{\partial f}{\partial y} \qquad 2y$$

$$\lim_{\Delta x \to 0} = \frac{f(x+\Delta x, y) - f(x,y)}{\Delta x} = 2x$$

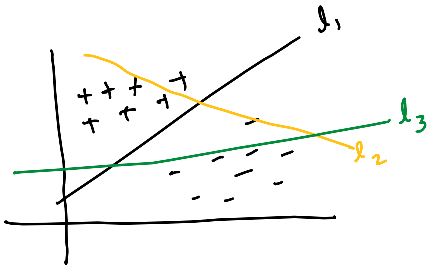
$$\lim_{\Delta y \to 0} = \frac{f(x, y+\Delta y) - f(x,y)}{\Delta y} = 2y$$

$f(x) = 2x^2 + 3 \qquad \frac{df}{dx} = 2 \cdot 2 \cdot x^1 + 0 = 4x$

$f(x,y) = x^2 + y^2 \rightarrow \frac{\partial f}{\partial x} = \frac{\partial (x^2 + c)}{\partial x} = 2x$ ✓

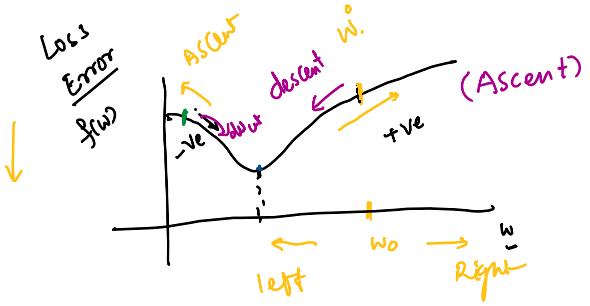
$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}(c + y^2) = 2y$ ✓

Gradient descent

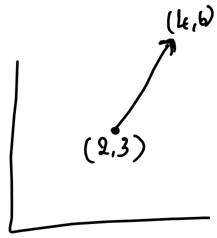


$[l_1, l_2, l_3]$

l_1 - Better 'DB'



Gradient is a Vector ✓



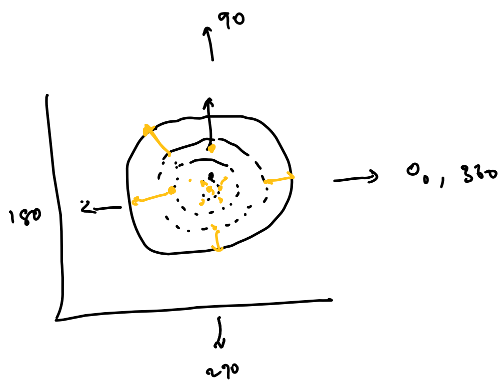
$f(x,y) = x^2 + y^2$

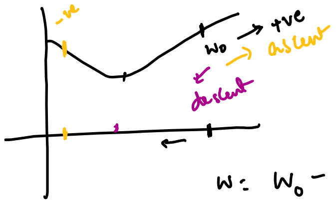
$2x = 4$
$2y = 6$

→ Gradient always points in the direction of 'Ascent'

→ opp of Gradient will point towards "descent"

$f(w_1, w_2)$: 2D





$w = w_0 - \text{gradient}$

$w = w_0 - \text{gradient}$

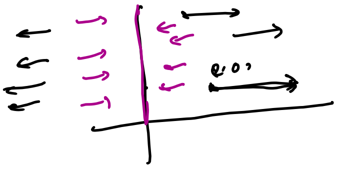
$= w_0 - (-1)$

$\Rightarrow$ addition

$f(x,y) = x^2 + 0$

$\frac{\partial f(x,y)}{\partial x} \rightarrow$ $\begin{bmatrix} 2x \\ 0 \end{bmatrix}$

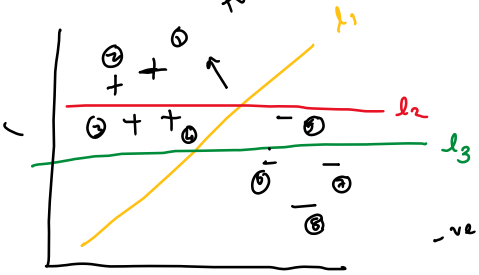
$\frac{\partial f(x,y)}{\partial y} \rightarrow$



$+ \; (w^T x + w_0) > 0$

+ve

Classifier



l_1 - better line

l_3

$-1 \times (w^T . x_{⑤} + w_0)$

-ve $\quad -(w^T + w_0) \} > 0$

$y_i\, -1$

$l_1 \rightarrow$ No missclassification

$l_2 \rightarrow$ ③④ miss classified

$l_3 \rightarrow$ ⑤ mis classified

for Correctly classified observations

$y_i \cdot (w^T x + w_0) > 0$ ✓

wrong classification $\rightarrow$ $[\, y_i \cdot (w^T . x + w_0) < 0 \,]$

wrongly classified observations $\quad y_i (w^T x + w_0) < 0$

Minimize $\rightarrow$ $-(y_i (w^T x + w_0))$ ✓

$l_2 \begin{cases} w^T x_{③} + w_0 = -10x + 1 \\ w^T x_{④} + w_0 = -10x + 1 \end{cases}$

$-10, -10$

$-(-20) = 20$

$-20 \rightarrow \underline{l_2}$ ③, ④ mis classified.

$-(-6) = 6$

$l_3 : w^T x_{⑤} + w_0 = 6x - 1$

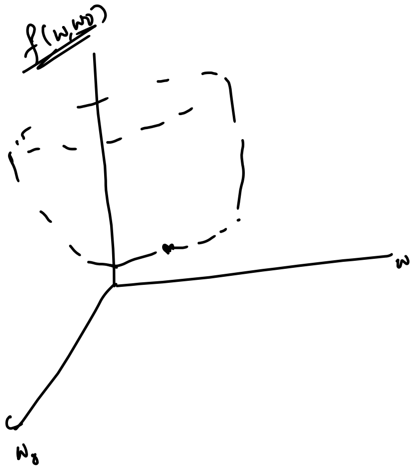
-6

$-6 \rightarrow l_3 \rightarrow$ ⑤ misclassified.

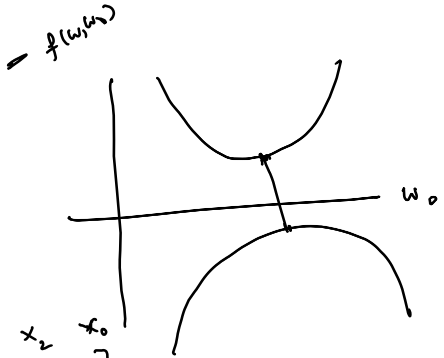
$\text{Min} \left(-y_i (w^T x + w_0) \right) \downarrow$ 'Better decision DB'

$f(w, w_0)$ because x_1, x_2 (inputs) fixed.

$E = \sum_{i=1}^{n} -y_i (w^T . x + w_0)$ only for missclassified observations



$f(w,w_0) = -y_i\,(w^T.x + w_0)$



$w^T.x$

l_3 – 5 $\quad x_1 \; x_2 \qquad x_1 \; x_2 \; x_0$

5 $[2, 3] \qquad [2, 3 \; 1]$

⇒ gradients $\quad [w_1, w_2] \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

$= -2y_5 w_1 - 3y_5 w_2 - w_0 y_5$

⇒ $= -(2w_1 + 3w_2 + w_0)\, y_5$

$\nabla L_{w_1} = -2y_5$

$\nabla L_{w_2} = -3y_5$

$\nabla L_{w_0} = -y_5$

$-y_5 \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = -y_5 \overset{x_5}{[\;\cdot\;]}$

$= -y_i \overset{x_i}{[\;\;]}$

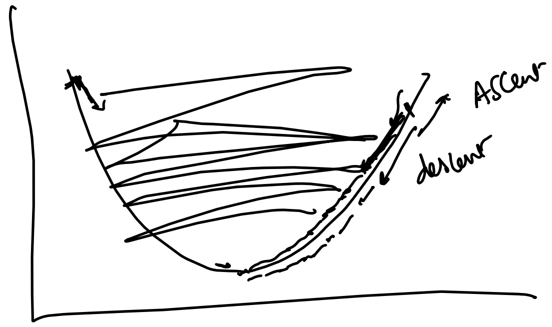
$$\left[\begin{aligned} w_1 &= w_1 - (-y_5 \cdot 2) \\ w_2 &= w_2 - (-y_5\, 3) \\ w_0 &= w_0 - (-y_5) \end{aligned}\right] \leftarrow$$

l_2

3, 4

$= -y_3 \overset{x_3}{[\;\;]} - y_4 \overset{x_4}{[\;\;]} =$ (Combined gradient)

$\nabla L_{w_1} = \Sigma - y_i \overset{x_i}{[\;\;]}$

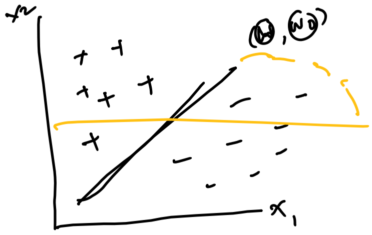


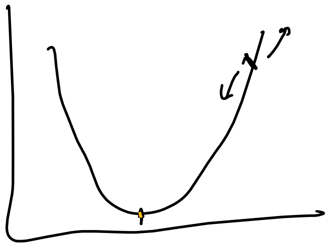
10 → obs (9 – →)

$w^1 = w^0 - \gamma(\nabla)$ $\qquad \gamma$ – learning rate

$$\frac{dy}{dx} = 0$$

$$E = -y_i(w^T \cdot x + w_0) < 0 \qquad ((w^T \cdot x + w_0)y_i < 0)$$





d

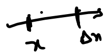
$$f(x) = x^2 + 2x + 1$$

$$\frac{df(x)}{dx} = \boxed{2x + 2 = 0}$$

$$2x = -2$$

$$x = -1$$

$$dc = 0$$



$$f(x,y,z) = \frac{x^2 - 3y^3 + z^2}{}$$

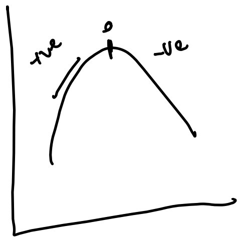
$$\frac{\partial f}{\partial x} = x^2 - c + p = 2x$$

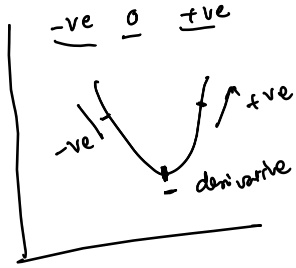
$$\frac{\partial f}{\partial y} = c - 3y^3 + k = -3 \cdot 3y^2 = -9y^2$$

→ Maxima

$f'(x) = 0 \quad f''(x) > 0$ Minimum

$f'(x) = 0 \quad f''(x) < 0$ Maximum





Local minima