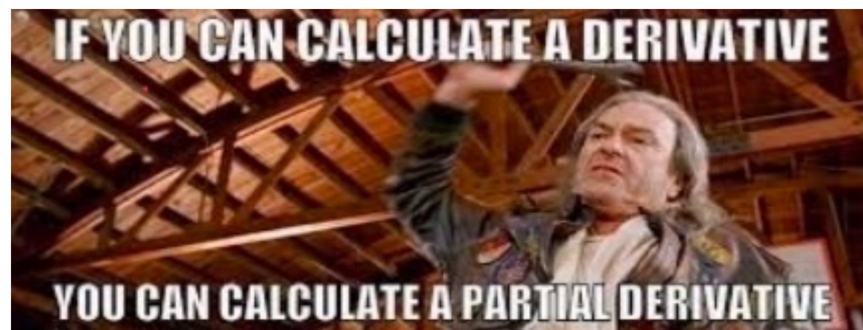
Constraint Optimisation



Agenda :

→ Recap

→ Lagrange Multipliers

→ Unconstraint Opt.

→ Constraint Opt.

Recap

Derivatives: $f(x) \rightarrow$ Continous & differentiable

$$\frac{d\,f(x)}{dx} = f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

Maxima/minima: To find Candidate Points for maxima & minima.

find x, wh: $\boxed{f'(x) = 0}$

if $f''(x) > 0$: minima

$f''(x) < 0$: maxima

Rules:

1. Linearity
2. Product
3. Quotient
4. Chain

Partial Derivatives:

function with multiple inputs. $f(x, y, z)$

$$\frac{d\, f(x)}{dx} \quad Vs. \quad \frac{\partial f(x,y)}{\partial x} \;\; \& \;\; \frac{\partial f(x,y)}{\partial y}$$

$f(w_1, w_2, w_0)$

$$\text{gradient} = \nabla_{\bar{w}} f(\bar{w})$$

→ derivative w.r.t a vector of inputs.

Gradient Descent:

↓

optimisation

↓

minima

1. init Variables $(\bar{w})$ randomly
2. repeat {

$$\bar{w} = \bar{w} - \underset{0.1}{\eta} \cdot \boxed{\frac{\partial L}{\partial \bar{w}}} \quad (\nabla_{\bar{w}} L) = \begin{bmatrix} \frac{\partial L}{\partial w_1} \\ \frac{\partial L}{\partial w_2} \end{bmatrix}$$

}

In what direction does the gradient of a function point at a specific point?

35 users have participated

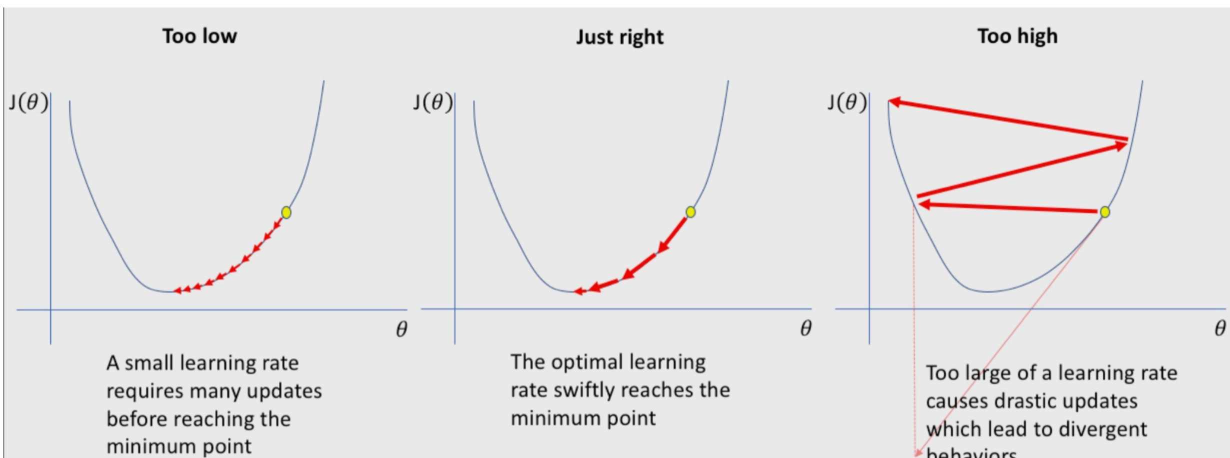
A	It points towards the nearest maximum value of the function.	6%
B	It points towards the nearest minimum value of the function.	43%
C	It points in the direction of the steepest increase of the function at that point.	51%
D	It points in the direction of the x-axis.	0%

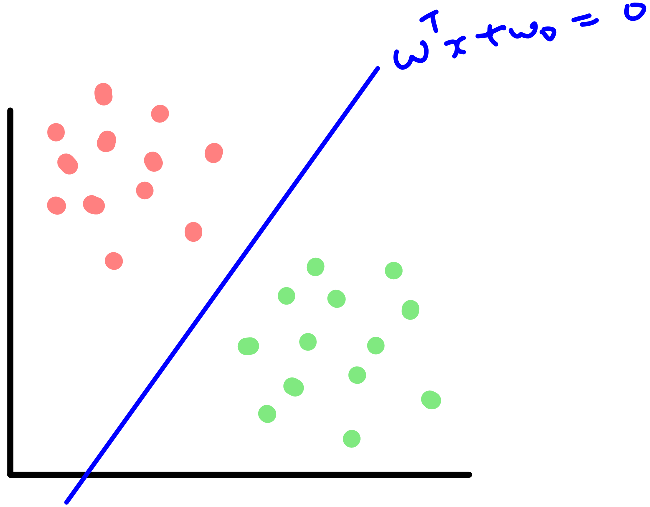
– gradient

In gradient descent, what does the learning rate determine?

34 users have participated

A	The size of the dataset used for training.	0%
B	The number of iterations required for convergence.	6%
C	The accuracy of the model's predictions.	6%
D ✓	The step size taken while updating parameters in the direction of the gradient.	88%





$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ $\quad w_0$

$\downarrow$

Gain function

$\downarrow$

Loss function

$$w^*, w_0^* = \underset{w, w_0}{\arg\min} \text{ Loss func}$$

Compute gradient / Vector calculus

$$f(x_1, x_2, x_3) = \bar{a}^T . \bar{x}$$
$$= a_1 x_1 + a_2 x_2 + a_3 x_3$$

$$\frac{d\,2x}{dx} = 2$$

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\nabla_{\bar{x}} f(\bar{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \bar{a}$$

$$f(\bar{x}) = a^T x$$

$$\nabla_{\bar{x}} f(\bar{x}) = \bar{a}$$

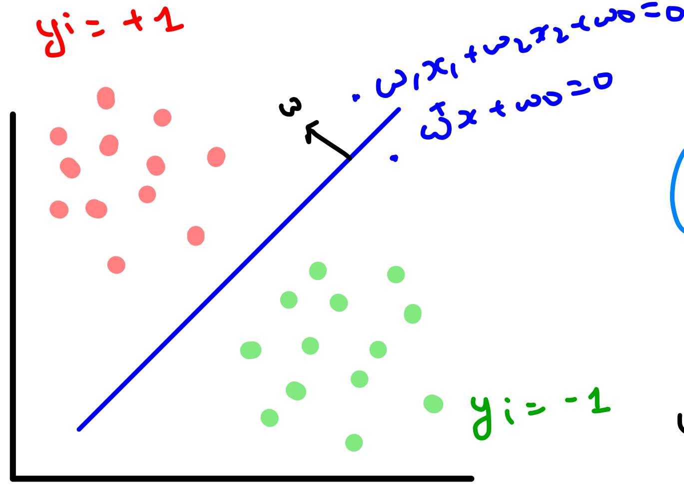
2. $f(\bar{x}) = x^T x$

$f(x_1, x_2, x_3 \ldots x_d) = \sum_{i=1}^{d} x_i^2 = x_1^2 + x_2^2 + \ldots + x_d^2$

$$\nabla_{\bar{x}} f(\bar{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_d} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_d \end{bmatrix} = 2 \cdot \bar{x}$$

$f(\bar{x}) = x^T x$

$\nabla_{\bar{x}} f(\bar{x}) = 2\bar{x}$



$$\text{Gain}(\bar{w}, w_0) = \frac{1}{n}\sum_{i=1}^{n} \frac{(w^T x^i + w_0)}{\|w\|} \cdot y^i$$

$$w^*, w_0^* = \underset{w, w_0}{\text{argmax}} \ \text{Gain}(w, w_0)$$

$$\text{Loss}(w, w_0) = -\frac{1}{n}\sum_{i=1}^{n} \left(\frac{w^T x^i + w_0}{\|w\|}\right) y^i$$

$$w^*, w_0^* = \underset{w, w_0}{\text{argmin}} -\frac{1}{n}\sum_{i=1}^{n} \left(\frac{w^T x^i + w_0}{\|w\|}\right) \cdot y^i$$

$$x^* = \underset{x}{\text{argmin}} (x-5)^2$$

$$w^*, w_0^* = \underset{w, w_0}{\operatorname{argmin}} \; -\frac{1}{n} \sum_{i=1}^{n} \left(\frac{w^T x^i + w_0}{\|w\|} \right) \cdot y^i$$

x → is not variable
is const
(data)

w → is variable
(line-change)

Apply "G.D"

1. randomly init w & w_0

2. repeat {

$$w = w - \eta \cdot \nabla_{\bar{w}} L(w, w_0)$$

$$w_0 = w_0 - \eta \cdot \frac{\partial L}{\partial w_0}$$

}

find

$$\bar{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\left\{ \begin{aligned} & L(w, w_0) = -\frac{1}{n} \sum_{i=1}^{n} (w^T x^i + w_0) \cdot y^i \\ & \text{s.t.} \quad \underbrace{\|w\| = 1}_{\text{Constraint}} \end{aligned} \right\}$$

Constraint optimization Problem

s.t → Such that / Subject to

$$\sqrt{w_1^2 + w_2^2} = 1$$

$\Rightarrow$ Lagrange multiplier

Q→ Optimize $y = x^2 - 3x - 3$

s.t $-x^2 + 2x + 3 = 0$

$x = -1 \rightarrow f(x) = 1$

minima $x = 3 \rightarrow f(x) = -3$

Ans

↑ minimum

Constraint optimization $\rightarrow$ $\boxed{\text{lagrange multiplier}}$ $\rightarrow$ Unconstraint optimization

$$\underset{x}{\operatorname{argmin}}\ f(x) \quad \text{s.t. } g(x) = 0 \quad \Rightarrow \quad \underset{x,\lambda}{\operatorname{argmin}}\ f(x) + \lambda \cdot g(x)$$

$\lambda \downarrow$ lagrange multiplier

$$\underset{x}{\operatorname{argmin}}\ f(x)$$

$$\text{s.t. } g_1(x) = 0 \quad \cdots \quad g_n(x) = 0$$
$$g_2(x) = 0 \quad \cdots$$

$$\underset{x,\lambda_1,\lambda_2\ldots\lambda_n}{\operatorname{argmin}}\ f(x)+\lambda_1\cdot g_1(x)+\lambda_2 g_2(x)+\ldots+\lambda_n\cdot g_n(x)$$

~~Example~~ :

$$\underset{x}{\operatorname{argmin}}\ x^2-3x-3$$

$$\text{s.t}\quad -x^2+2x+3=0$$

$\Downarrow$ lagrange multipler

$$x^*,\lambda^* = \underset{x,\lambda}{\operatorname{argmin}}\left[x^2-3x-3+\lambda\left(-x^2+2x+3\right)\right]$$

$\Downarrow$ Apply gradient Descent

result

$\Downarrow$ $\boxed{x=3}$

1. x & λ randomly.

2. $$x = x - \eta \cdot \frac{\partial L}{\partial x} \Rightarrow \left[2x - 3 + 0 - 2\lambda x + 2\lambda + 0 \right.$$
$$\left. = 2x - 2\lambda x + 2\lambda - 3 \right]$$

$$\lambda = \lambda - \eta \cdot \frac{\partial L}{\partial \lambda} \Rightarrow \left[0 + 0 + 0 - x^2 + 2x + 3 \right]$$

$\boxed{\frac{\partial L}{\partial x} = 0}$: $2x - 2\lambda x + 2\lambda - 3 = 0$

$\boxed{\frac{\partial L}{\partial \lambda} = 0}$: $-x^2 + 2x + 3 = 0$

Var: x & λ

⋮

$$-x^2 + 3x - x + 3 = 0$$

$$-x(x+3) - 1(x-3) = 0$$

$$(-x-1)(x-3) = 0$$

$$\boxed{\begin{aligned} x &= -1 \\ x &= 3 \end{aligned}}$$

Optional

Manual X
Calculation

$$L(w, w_0) = -\frac{1}{n}\sum_{i=1}^{n} (w^T x^i + w_0) \cdot y^i$$

S.t $\|w\| = 1$

such that

subject to

Constraint optimization Problem

Constraint

$\sqrt{w^T w} = 1$

$\Downarrow$

$\underbrace{\sqrt{w^T w} - 1}_{g(x)} = 0$

$\Rightarrow$ Lagrange multipler

Convert into unconstraint Opt. Problem.

$$w^*, w_0^*, \lambda^* = \underset{w, w_0, \lambda}{\operatorname{argmin}} \; -\frac{1}{n} \sum_{i=1}^{n} (w^T x^i + w_0) \cdot y^i + \lambda \left(\sqrt{w^T w} - 1 \right)$$

$\downarrow$ Apply Gradient Descent

1. randomly init w, w_0, λ
2. repeat {

$$w = w - \eta \cdot \boxed{\nabla_w L}$$

$$w_0 = w_0 - \eta \cdot \boxed{\frac{\partial L}{\partial w_0}}$$

$$\lambda = \lambda - \eta \cdot \boxed{\frac{\partial L}{\partial \lambda}} \rightarrow \sqrt{w^T w} - 1$$

}

$$L = -\frac{1}{n} \sum_{i=1}^{n} (w^T x_i + w_0) \cdot y^i + \lambda \left(\sqrt{w^T w} - 1 \right)$$

$$\frac{\partial L}{\partial w_0} = -\sum \left[\cancel{\frac{\partial}{\partial w_0} w^T x_i \cdot y_i} + \frac{\partial}{\partial w_0} w_0 \cdot y_i \right] + \lambda \cancel{\frac{\partial \left(\sqrt{w^T w} - 1 \right)}{\partial w_0}}$$

$$= -\frac{1}{n} \sum_{i=1}^{n} y_i$$

$$L = -\frac{1}{n} \sum_{i=1}^{n} (w^T x_i + w_0) \cdot y^i + \lambda \left(\sqrt{w^T w} - 1 \right)$$

$w_1 x_1 + w_2 x_2$

$$\frac{\partial L}{\partial w} = -\sum \left[\frac{\partial w^T x}{\partial w} \cdot y + \cancel{\frac{\partial w_0 y}{\partial w}} \right] + \lambda \frac{\partial \sqrt{w^T w}}{\partial w} - \cancel{\frac{\partial \lambda}{\partial w}}$$

$$= -\frac{1}{n} \sum_{i=1}^{n} y^i \bar{x}^i + \lambda \cdot \frac{1}{2} \cdot \frac{1}{\sqrt{w^T w}} \cdot \frac{\partial w^T w}{\partial w}$$

$$+ \lambda \cdot \frac{1}{\cancel{2}} \frac{1}{\sqrt{w^T w}} \cdot \cancel{2} \bar{w}$$

$$\left[\frac{\partial L}{\partial \bar{w}} = -\frac{1}{n} \sum_{i=1}^{n} y^i \cdot \bar{x}^i + \frac{\lambda \bar{w}}{\|\bar{w}\|} \right]$$

$$\frac{\partial}{\partial w}(w^T w)^{1/2}$$

$$\frac{\partial (w^T w)^{\frac{1}{2}}}{\partial w^T w} \cdot \frac{\partial w^T w}{\partial w}$$

$$= \frac{1}{2\sqrt{w^T w}} \cdot 2w$$

$$\frac{d p^{1/2}}{d p} = \frac{1}{2} p^{-1/2}$$

$$= \frac{1}{2\sqrt{p}}$$