Chrome File Edit View History Bookmarks Profiles Tab Window Help 21:07

M W F

Constrained Optimization

& SGD

3 : 15 min

Request → 11:30 PM sharp

Q&A → 11:30 +

plot(x^2+y^2) - Google Search

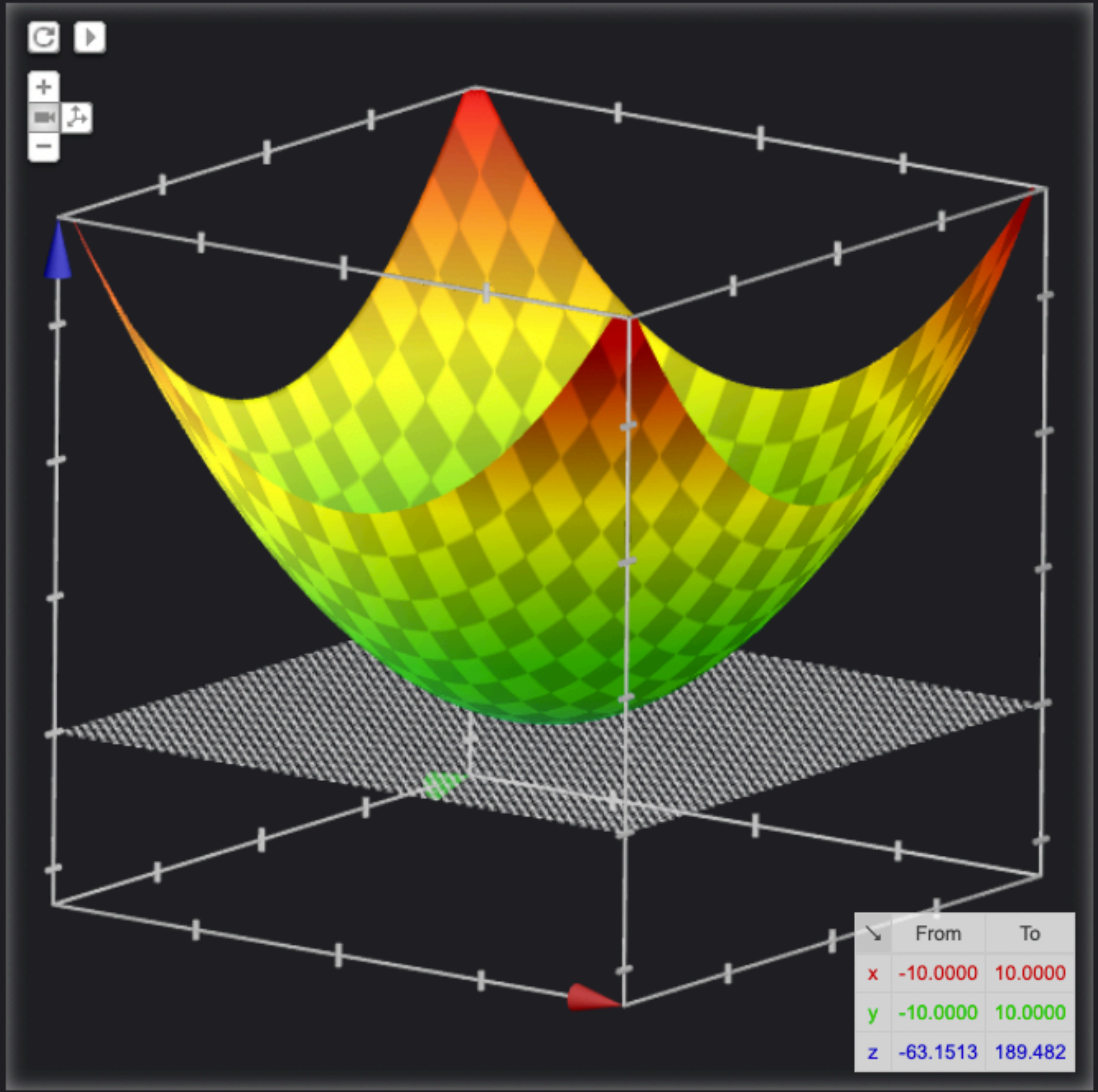
google.com/search?q=plot(x%5E2%2By%5E2)&rlz=1C5CHFA_enIN958IN958&oq=plot(x%5E2%2By%5E2&aqs=chrome.0.69i59j69i57j69i59j69i59j0i30l7.4707j0j7&sourceid=chrome&ie=UTF-8

plot(x^2+y^2)

All Images News Maps Videos More Tools

About 81,20,00,000 results (0.55 seconds)

Graph for x^2+y^2



	From	To
x	-10.0000	10.0000
y	-10.0000	10.0000
z	-63.1513	189.482

More info

2 / 67

Maxima & Minima

η : learning rate

Prev:

- $\boxed{GD}$ for $f(x) \rightarrow$ one-variable

Today

→ GD for $f(w, w_0) \rightarrow d+1$ variables → Binary classification

→ Constrained Optmzn → Binary classification, PCA (next class)

→ SGD & minibatch GD

Contour

Ops:

Topic → Chat (end of topic)

off topic → Ques (end of session)

Audio

4 / 4

Overshooting:

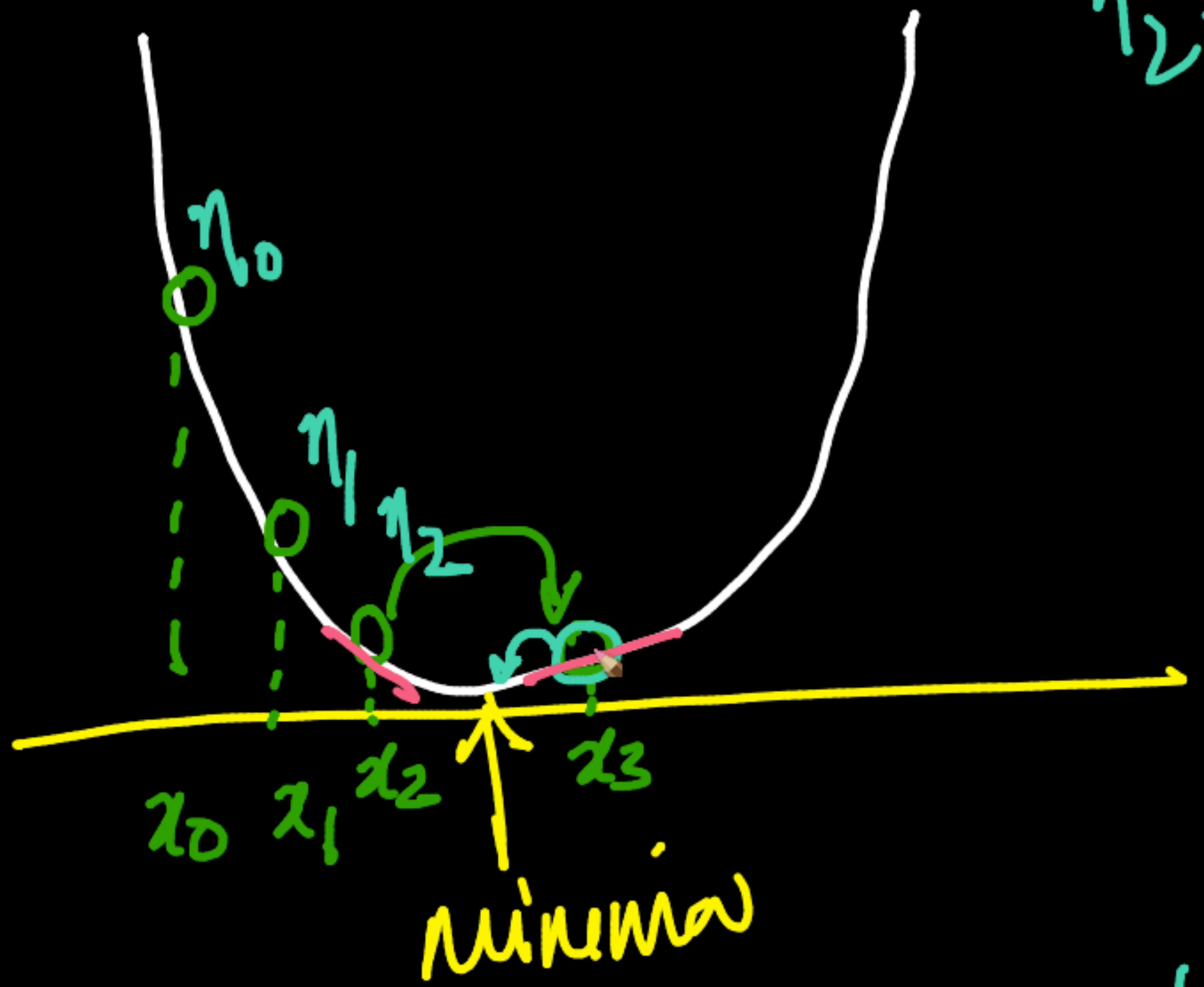
→ code Snippet

$\eta_0 = 0.1$; $\eta_1 = 0.09$

$\eta_2 = 0.081$

0.1 (const)

η : learning rate



i : iteration number

$\eta \downarrow$ as $i \uparrow$

$$x_{i+1} = x_i - \eta \left. \frac{\partial f}{\partial x} \right|_{x_i}$$

$$\eta = f(i)$$

$\eta = 0.1$

$\eta_{new} = 0.9 \times \eta_{old}$

$\eta = 0.01$; $c = 0.9$

$i = 0$

0.001

while $|df\text{-}dx - c| < 0$

{

update eqn

$\boxed{\eta = \eta \times 0.9}$

$\to i = i+1$

}

$i \uparrow \quad \eta \downarrow$

Chrome File Edit View History Bookmarks Profiles Tab Window Help 21:19

8 / 8

Chrome File Edit View History Bookmarks Profiles Tab Window Help 21:32

$z = f(x, y)$ 2 variables

→ initialize x_0 & y_0 randomly

while $\frac{\partial f}{\partial x} \,!\approx 0.0$ and $\frac{\partial f}{\partial y} \,!\approx 0$

✓ $x_1 = x_0 - \boxed{\eta} \left.\frac{\partial f}{\partial x}\right|_{x=x_0, y=y_0}$

MV - GD

adv - variations of GD → DL

$y_1 = y_0 - \boxed{\eta} \left.\frac{\partial f}{\partial y}\right|_{y=y_0,\, x=x_0}$

8 / 13

randomly

$$(x_0, y_0) \rightarrow (x_1, y_1) \rightarrow (x_2, y_2)$$

$$\left.\frac{\partial f}{\partial x}\right|_{x_0}, \left.\frac{\partial f}{\partial y}\right|_{y_0}$$

Single-var GD $\rightarrow$ ($x_0 \rightarrow x_1 \rightarrow x_2 \cdots\cdots x^*$

best

$$\left.\frac{df}{dx}\right|_{x^*} \approx 0$$

MV - GD

$\downarrow$

$$(x_0, y_0) \rightarrow (x_1, y_1) \rightarrow (x_2, y_2) \rightarrow \cdots (x^*, y^*)$$

$$\left.\frac{\partial f}{\partial x}\right|_{x^*} \approx 0 \text{ and } \left.\frac{\partial f}{\partial y}\right|_{y^*} \approx 0$$

minima exists (Test)

min $f(x, y)$

$\downarrow$

MV-GD

x^*, y^* $\rightarrow$ local minima ✓ maxima

higher order derv. test

$\downarrow$

expensive

$f(x) \rightarrow$

$f(w, w_0) \rightarrow$

Chrome File Edit View History Bookmarks Profiles Tab Window Help 21:29

Comp hack

$\nabla f_{x^*,y^*} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

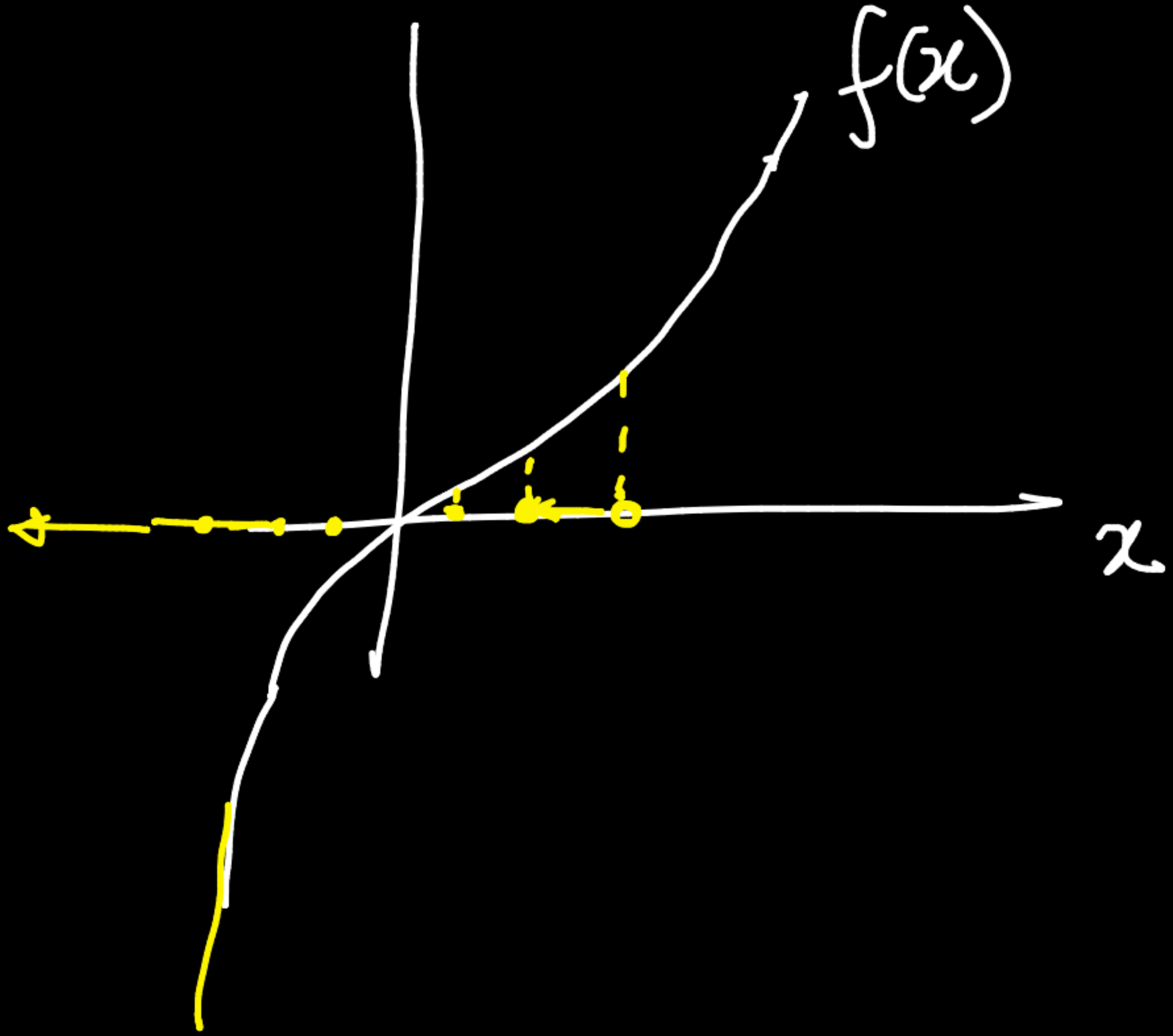
$f(x^*, y^*) = 10$ (let)

$f(x^* + 0.1, y^* + 0.1) = 9 \Rightarrow x^*, y^*$ cannot be minima $\rightarrow$ Maxima

$f(x^* + 0.1, y^* + 0.1) = 12 \Rightarrow x^*, y^*$ cannot be the Maxima $\Downarrow$ minima

$\boxed{\text{GD}} \rightarrow$ ignore saddle points

12 / 12



∞ loop

10000-iterations

$$\left.\frac{df}{dx}\right|_{x_{10000}} \approx 0$$

$f(x, y)$

$x_0, y_0 \longrightarrow x_1, y_1 \rightarrow \cdots \rightarrow x_{10}, y_{10}$

$\eta_x =$ v. v. small

$\eta_y =$

✓

$\left.\frac{\partial f}{\partial x}\right|_{x_{10}, y_{10}} \approx 0 \rightarrow$ Stop updating

$\left.\frac{\partial f}{\partial y}\right|_{x_{10}, y_{10}} \not\approx 0 \rightarrow$ updating

$$x_{10} \rightarrow \eta \boxed{\frac{\partial f}{\partial x}}\Bigg|_{x_0, y_0} \rightarrow 0$$

$$\left.\frac{\partial f}{\partial x}\right|_{x_0, y_0} \approx 0$$

$$\underbrace{x_{10}, y_0} \rightarrow \underbrace{x_{10}, y_{11}}$$

$$\left.\frac{\partial f}{\partial x}\right|_{x_{10}, y_{11}} \; ! \approx 0 \quad \checkmark$$

3Blue 1Brown

Khan Academy Gradient Descent 1hr

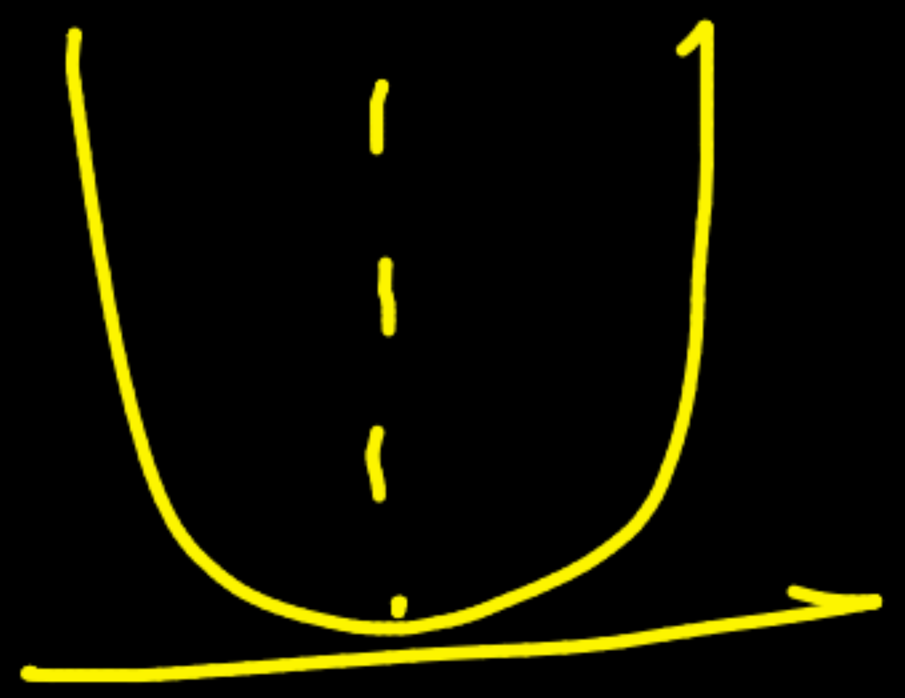
$f(\cdots)$

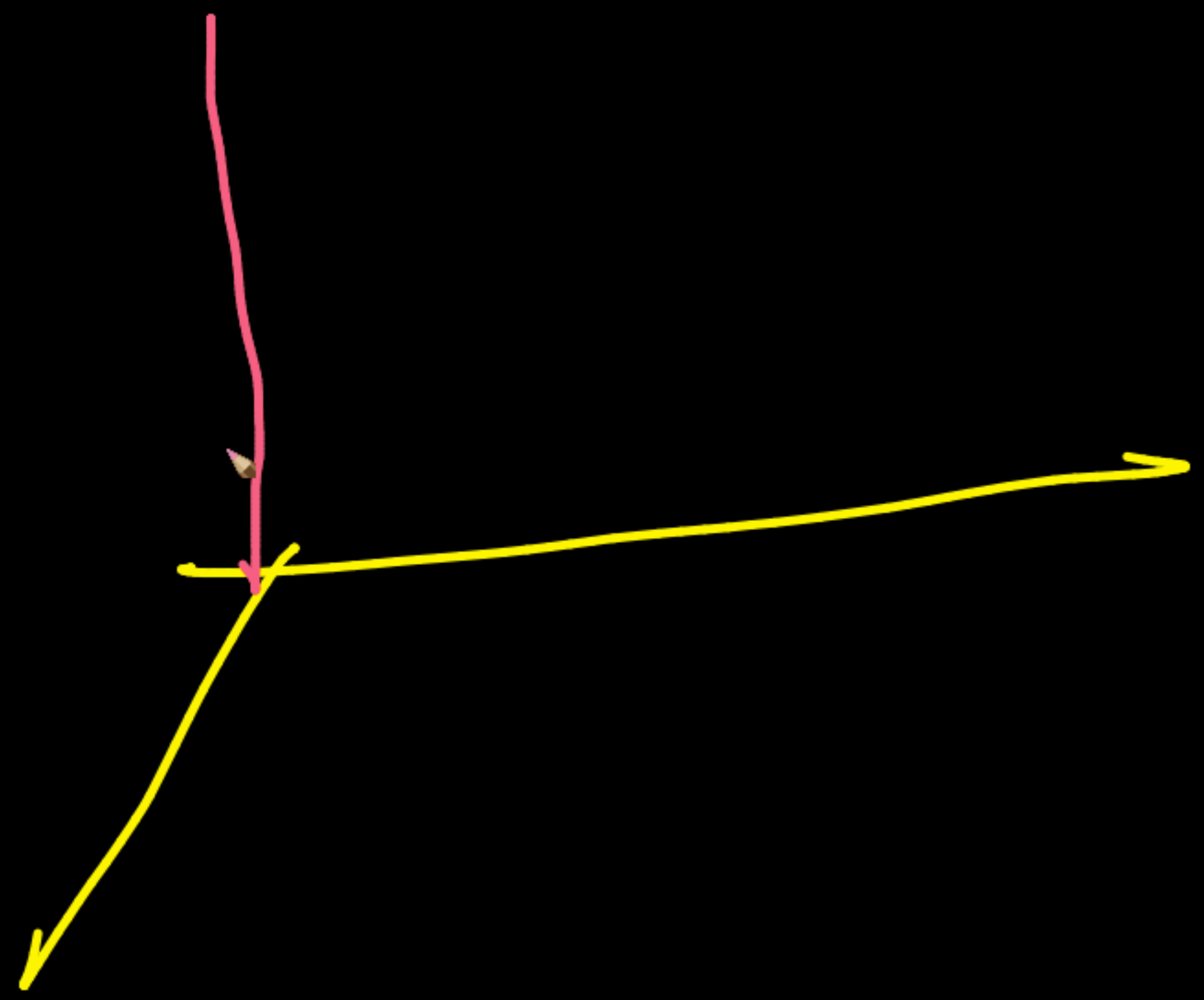
x^*, y^*

$f(x)$

popular interview ques

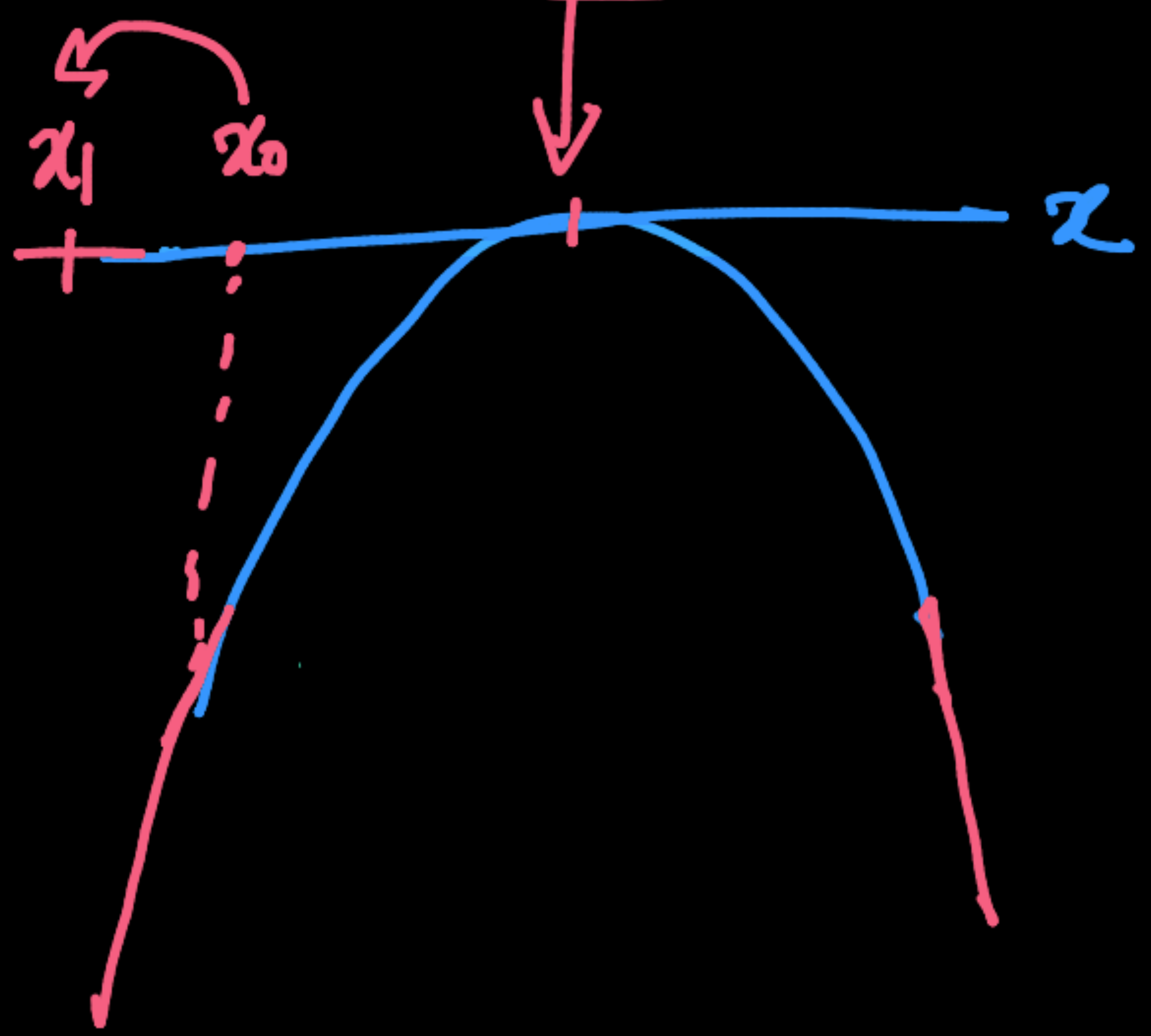
Code
Math & intuition
boundary cases
⋮





Maxima

GD update → ~~for~~ minimization only



$$x_{i+1} = x_i - \eta \left.\frac{\partial f}{\partial x}\right|_{x_i}$$

$$x^* = \max_x f(x) \;\; \simeq \;\; x^* = \min_x \left[-f(x) \right]$$

ML: min only

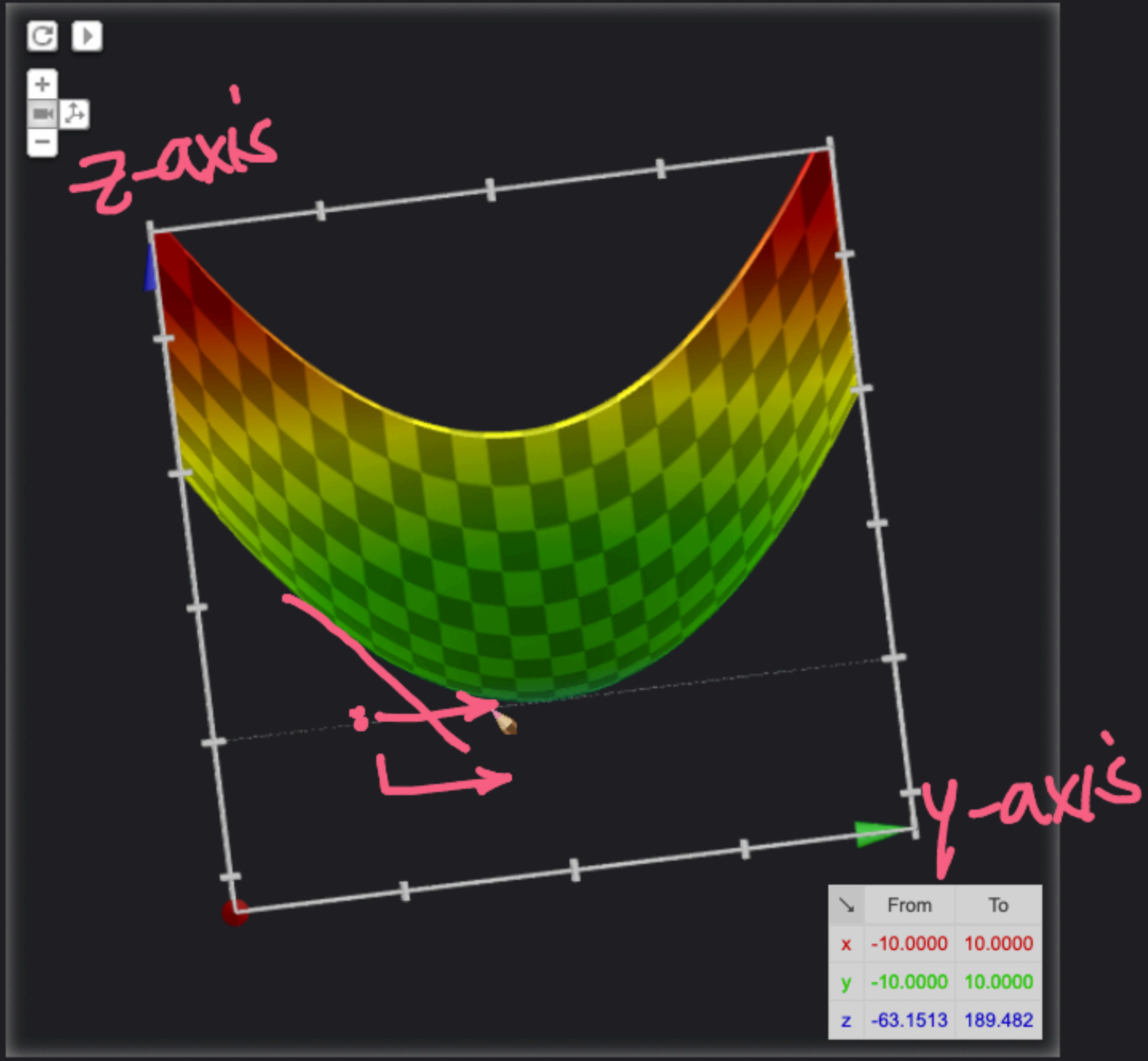
GD-update

plot(x^2+y^2)

All | Images | News | Maps | Videos | More | Tools

About 81,20,00,000 results (0.55 seconds)

Graph for x^2+y^2



More info

x_0, y_0

$\downarrow \quad \downarrow$

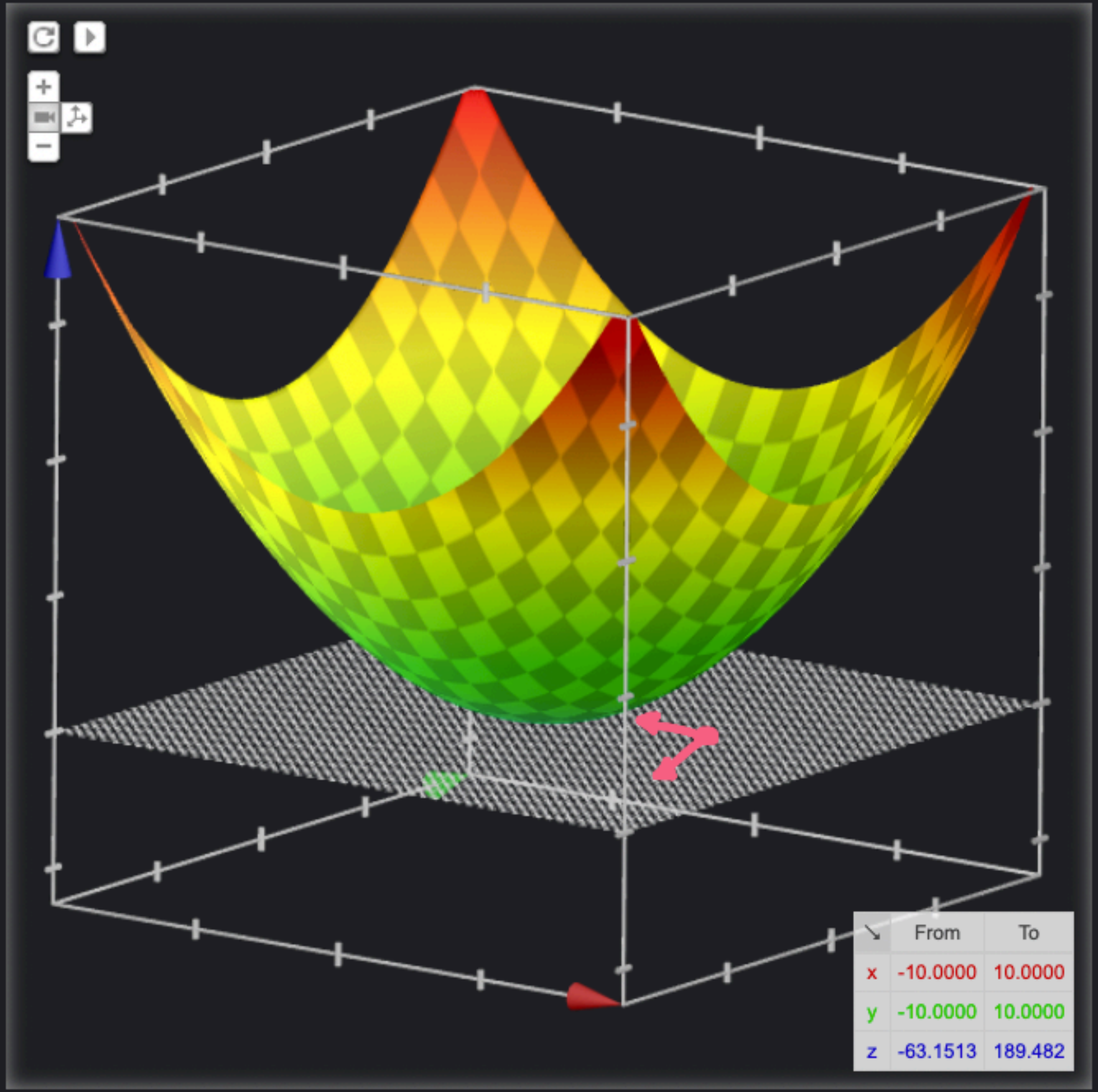
(reduce) $x_1 \quad y_1$ gr

plot(x^2+y^2)

All | Images | News | Maps | Videos | More | Tools

About 81,20,00,000 results (0.55 seconds)

Graph for x^2+y^2



More info

$$f(\underline{x},\underline{y}) = z$$

Chrome File Edit View History Bookmarks Profiles Tab Window Help 21:43

plot(x^2+y^2) - Google Search

google.com/search?q=plot(x%5E2%2By%5E2)&rlz=1C5CHFA_enIN958IN958&oq=plot(x%5E2%2By%5E2&aqs=chrome.0.69i59j69i57j69i59j0i30l7.4707j0j7&sourceid=chrome&ie=UTF-8

Update

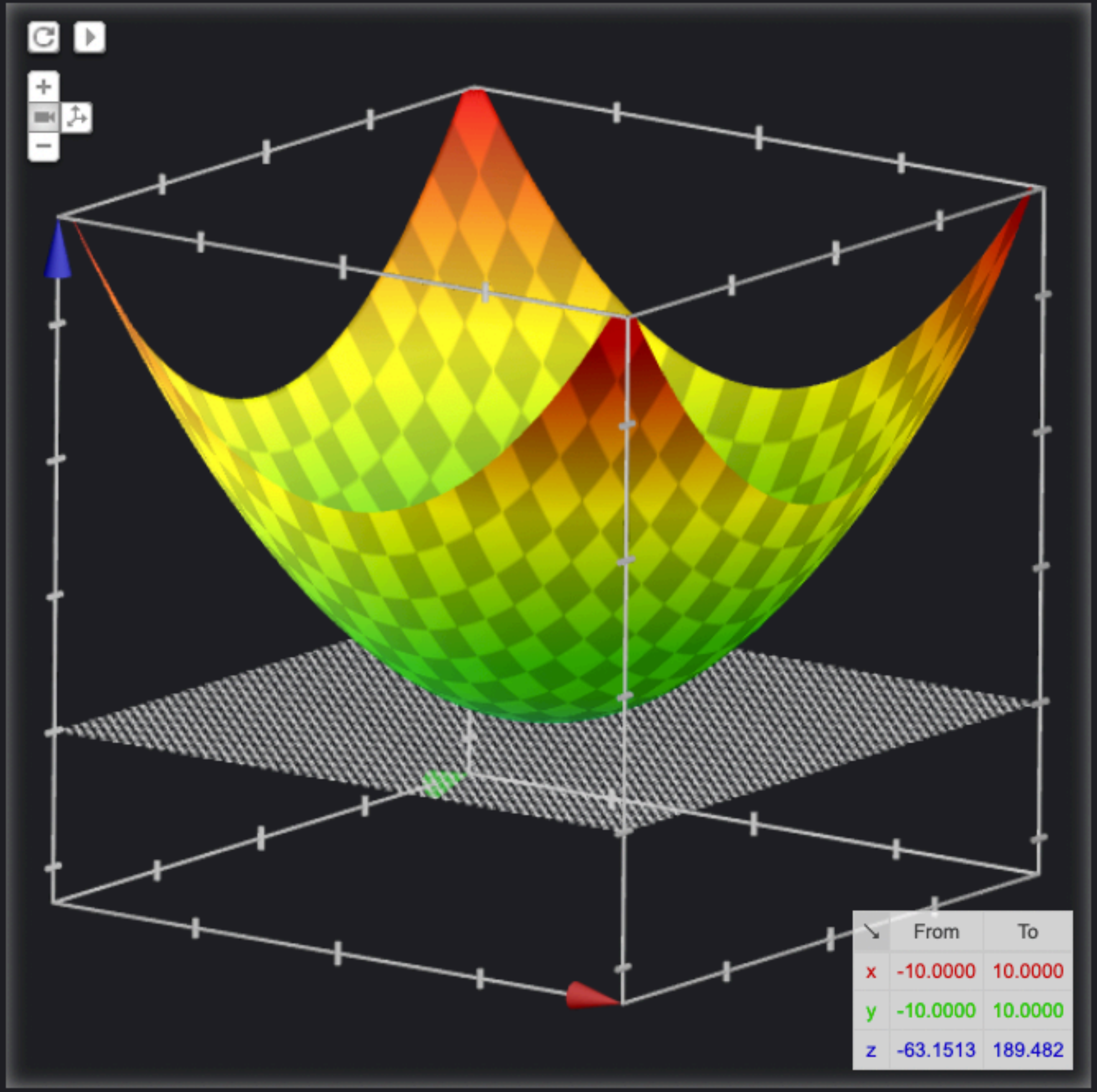
Google

plot(x^2+y^2)

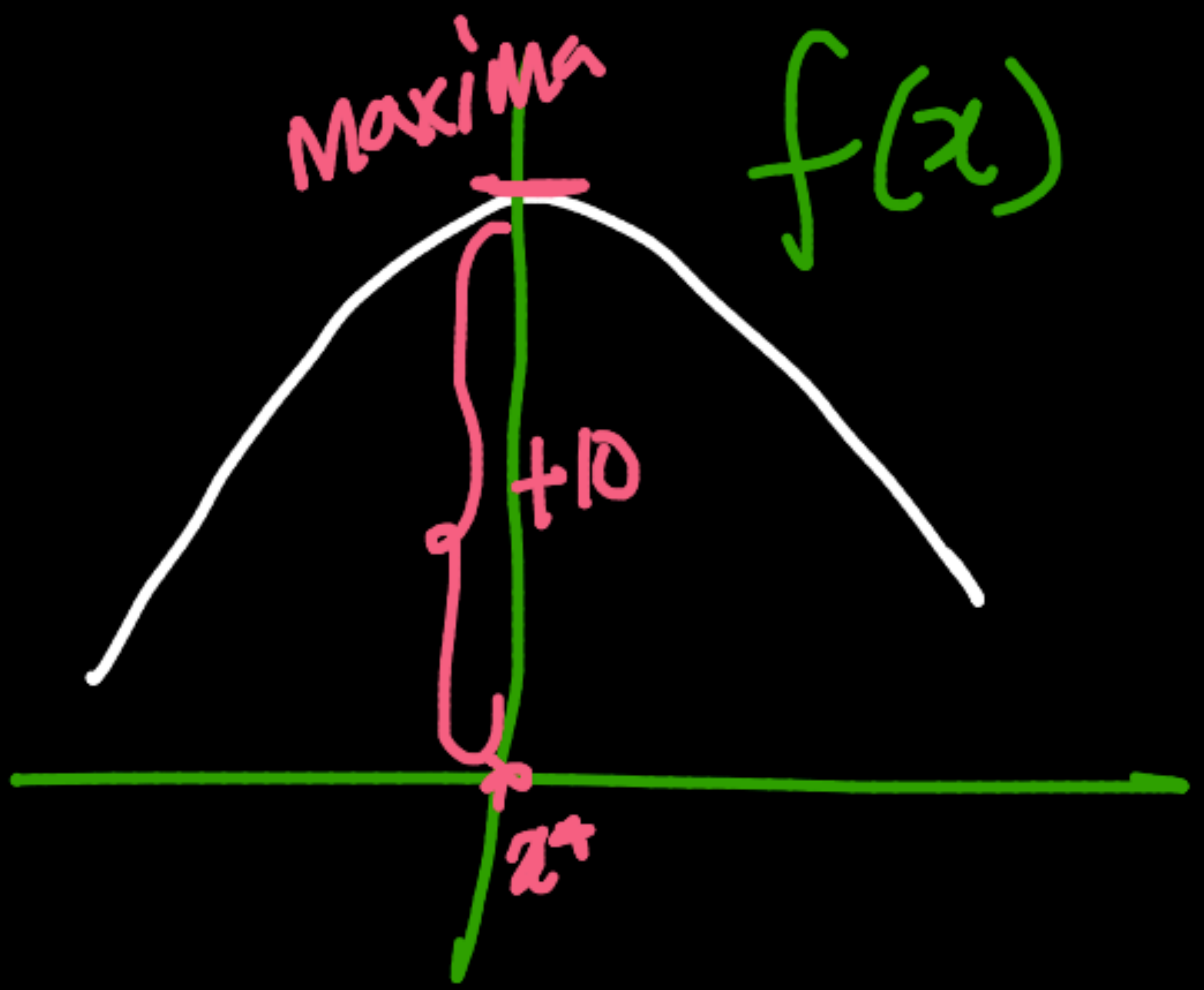
All Images News Maps Videos More Tools

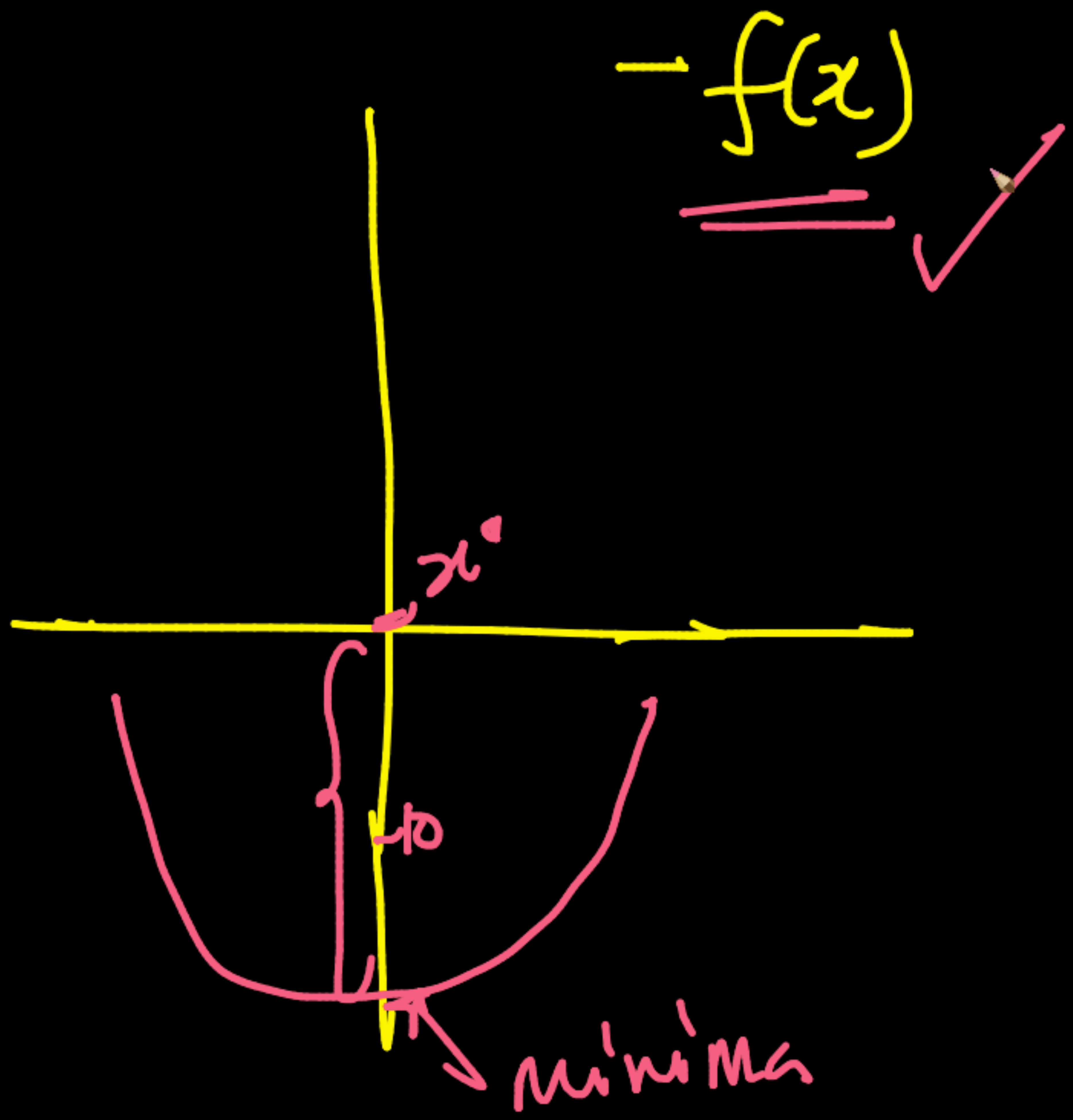
About 81,20,00,000 results (0.55 seconds)

Graph for x^2+y^2



More info

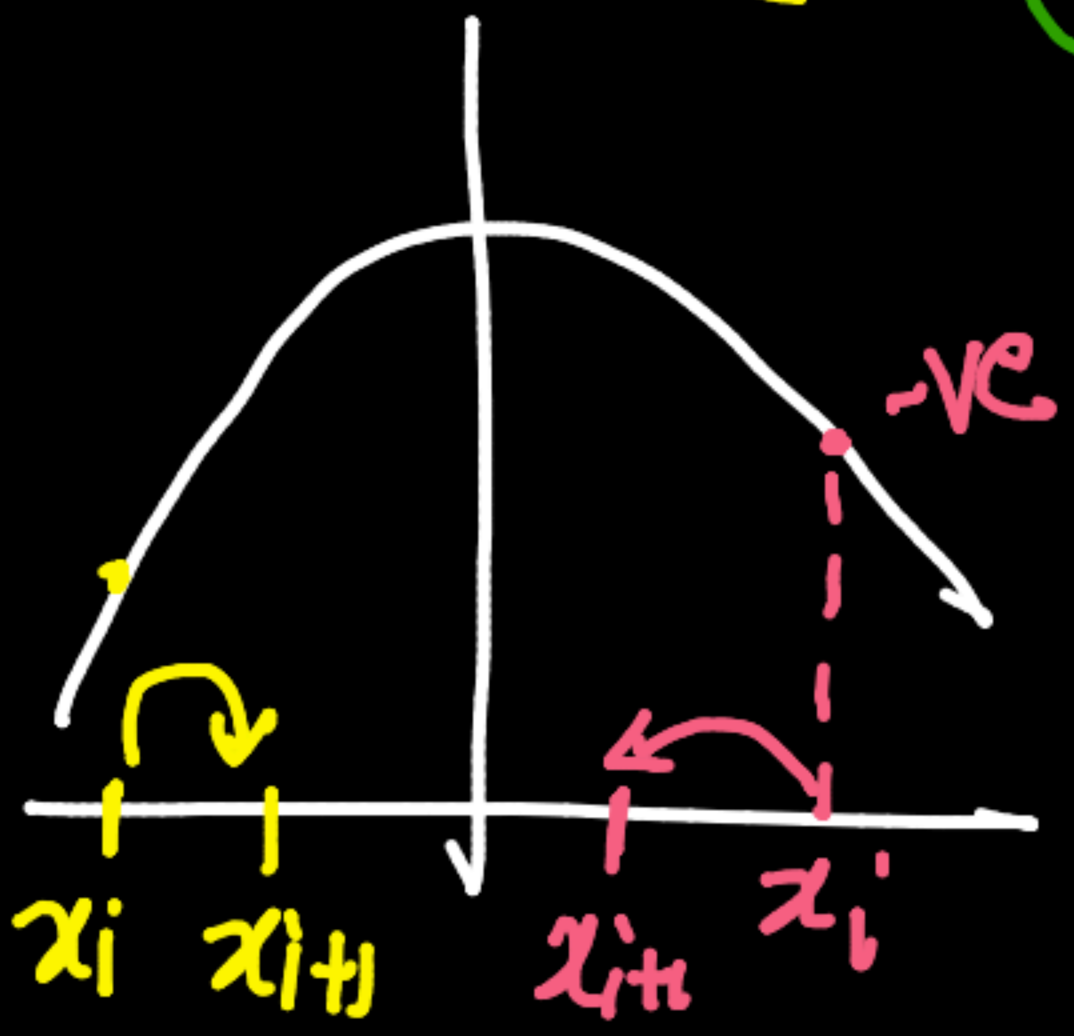

Maxima
f(x)
+10
x*


-f(x)
x*
-10
minima

$\boxed{Q}$

GD-min: $x_{i+1} = x_i - \eta \frac{\partial f}{\partial x}\Big|_{x_i}$

GD-max: $x_{i+1} = x_i + \eta \overbrace{\frac{\partial f}{\partial x}}^{+ve}\Big|_{x_i}$



Chrome File Edit View History Bookmarks Profiles Tab Window Help 22:08

ML: $y_i \in \{0,1\}$

Binary Classfn. Task

$$\min_{w, w_0} \cancel{\max} \sum_{i=1}^{n} -y_i \frac{(w^T x_i + w_0)}{\|w\|}$$

$-1, +1$

$d+1$-var

$f(w_0, w_1 \ldots w_d)$

$\mathcal{L}_i$

$$\min_{w, w_0} \left\{ \sum_{i=1}^{n} \left[-y_i \frac{(w_0 + w_1 x_{i1} + w_2 x_{i2} + \cdots + w_d x_{id})}{\sqrt{w_1^2 + w_2^2 + w_3^2 + \cdots + w_d^2}} \right] \right.$$

Loss-fn:

$(x_i, y_i) \leftarrow D$

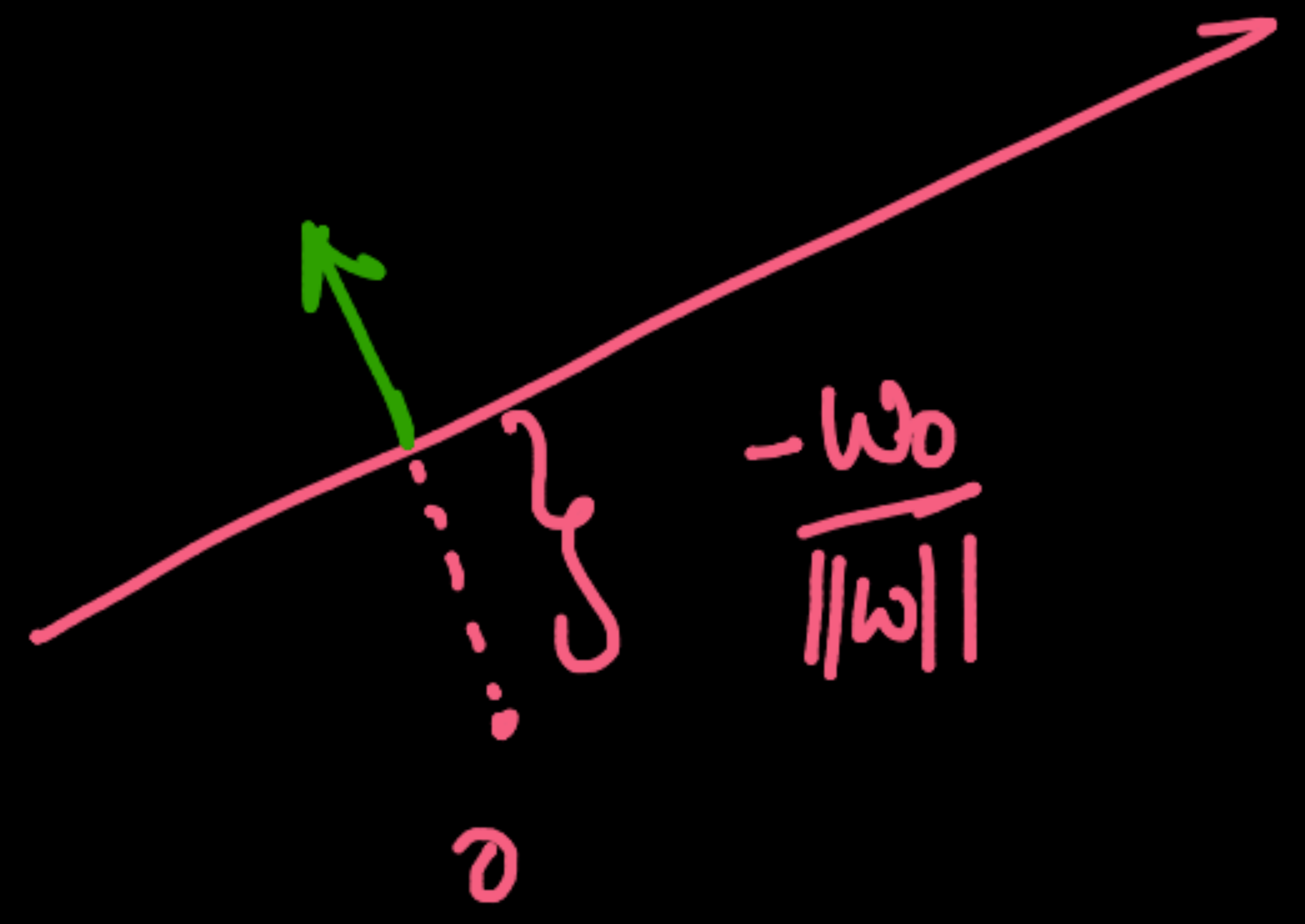
25 / 29

Eqn. of plane:

$$\underbrace{w^T}_{d} x + \underbrace{w_0}_{1}$$

$$y = mx + c$$

$$\pi_d : w^T x + w_0 = 0$$



Binary-classfn:

(i) $w_0, w_1 \ldots w_d \leftarrow$ random-values

Chrome File Edit View History Bookmarks Profiles Tab Window Help 22:07

Binary-classfn:

CODE

params

① $\underbrace{w_0, w_1 \ldots w_d}$ ← random-values

while loop

② for each j : 0, 1, 2, ... d

(index on the d+1 var)

update eqns

$$w_j^{new} = w_j^{old} - \eta \left. \frac{\partial L}{\partial w_j} \right|_{w_0^{old}, w_1^{old} \ldots}$$

all w_j's are updated independently

28 / 29

for all

$\forall j = 0, 1, 2 \ldots d$; $\boxed{\dfrac{\partial L}{\partial w_j}} \approx 0$

easily

$\{d$-~~variables~~ features

$d+1 \rightarrow$ variables

Chrome File Edit View History Bookmarks Profiles Tab Window Help 22:14

$\boxed{Q}$

$\frac{d}{dx}\frac{1}{x}$

$\frac{\partial L}{\partial w_1} = ?\ \sum_{i=1}^{n}$

{ Quoitent - rule

V-complex

$L = \sum_{i=1}^{n} -y_i\left(w_0 + w_1 \boxed{x_{i1}} + w_2 x_{i2} + \dots + w_d x_{id}\right)$

messy

$\left\{\sqrt{\boxed{w_1^2} + w_2^2 + \dots + w_d^2}\right.$

$\frac{d\, f+g}{dx}$

{ Chain rule

$\boxed{\partial \frac{f}{g} = \frac{f'g - g'f}{g^2}}$

30 / 31

Hack of geometry
& optimization

$$L = -\sum_{i=1}^{n} y_i \boxed{\frac{(w^T x_i + w_0)}{\cancel{\|w\|}}}$$

enforce $\|w\| = 1$

↑
unit-vec

Chrome File Edit View History Bookmarks Profiles Tab Window Help 22:31

Unconstrained

$$\min_{w, w_0} L = -\sum_{i=1}^{n} y_i \frac{(w^T x_i + w_0)}{\|w\|}$$

$\|w\| \rightarrow$ headache

Πd

$\approx$

Constrained opt

$$\min_{w, w_0} L = -\sum_{i=1}^{n} y_i (w^T x_i + w_0)$$

such that $\boxed{\|w\| = 1}$

Constraint

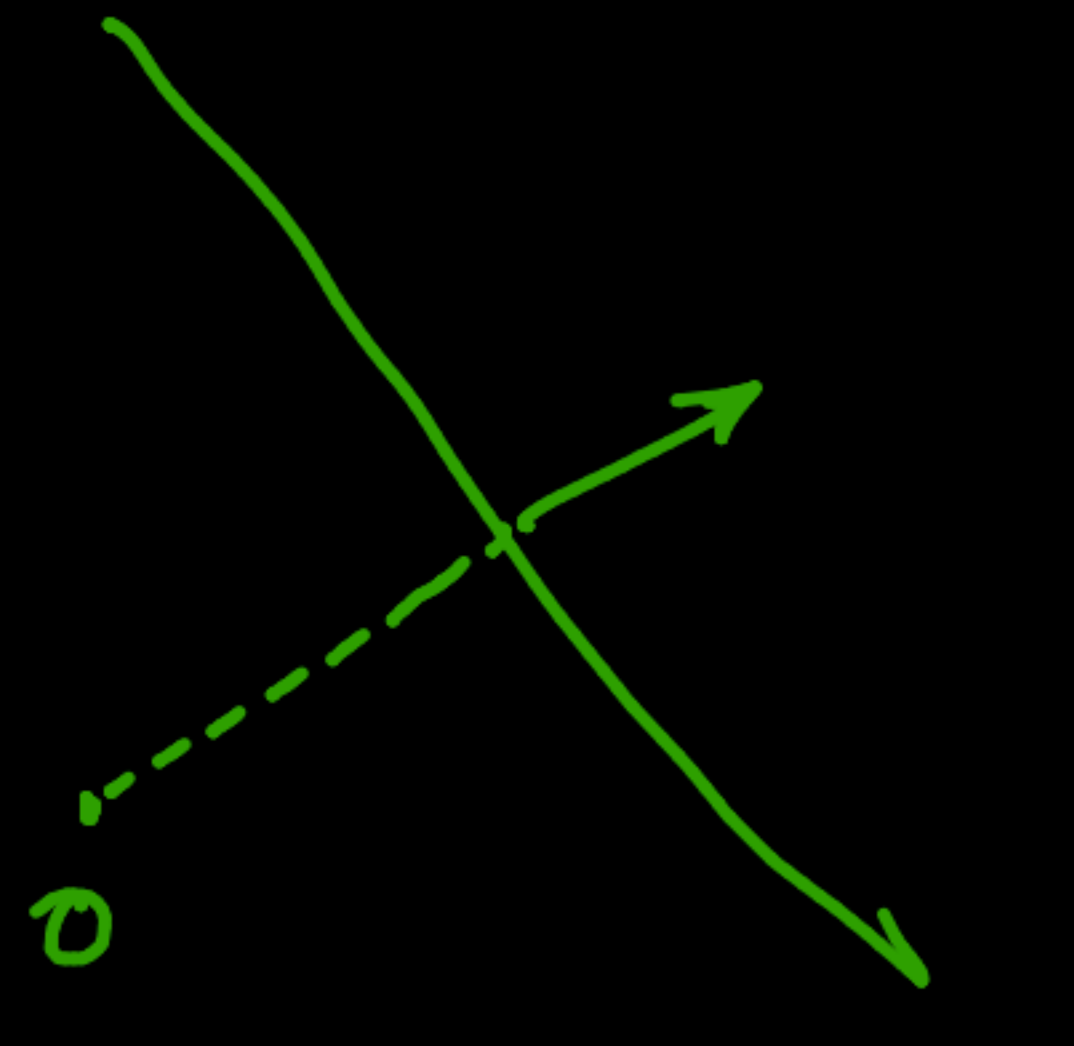
32 / 36

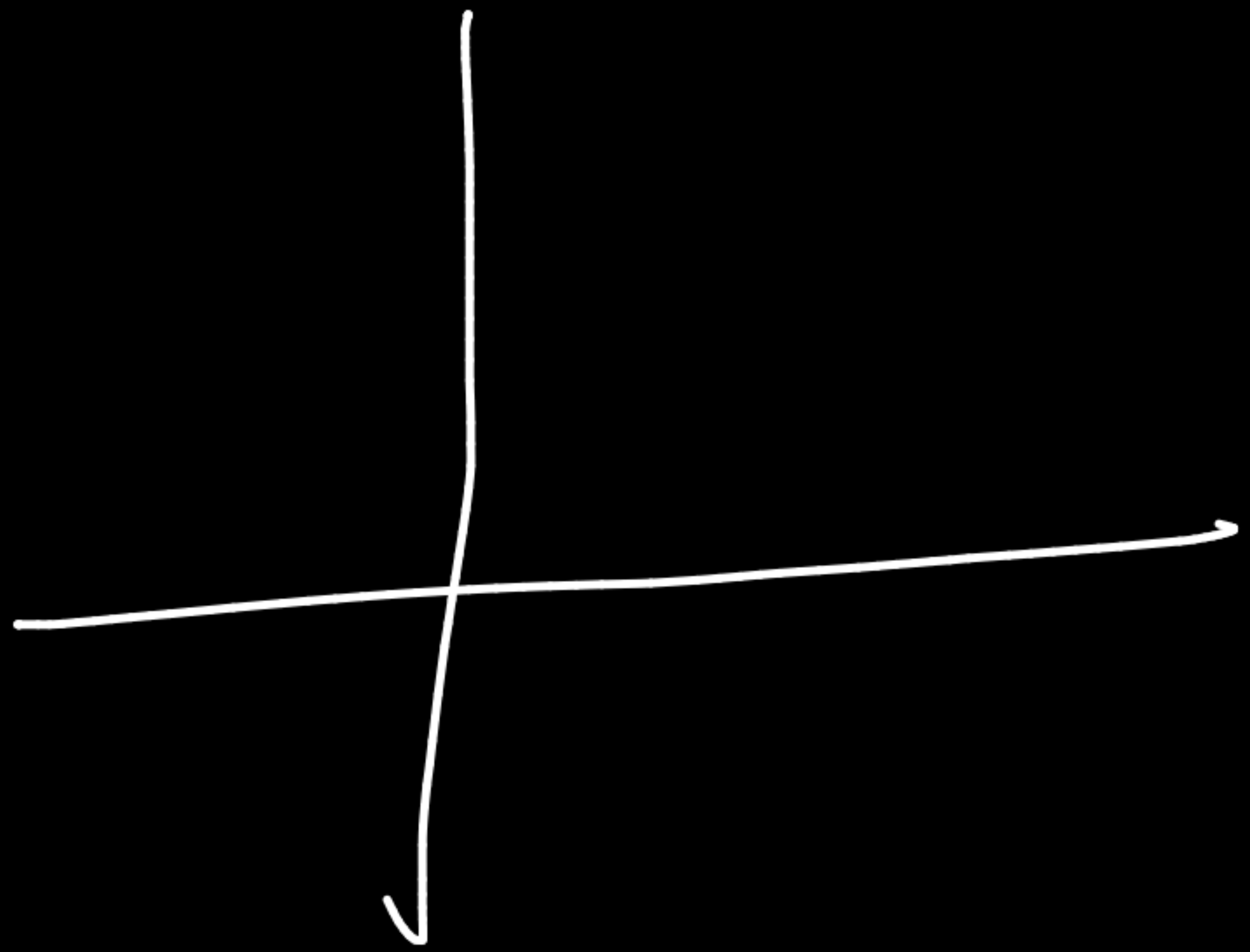
$\rightarrow$ dist : $\dfrac{w_0}{\|w\|}$ ✓

$\pi_1 \rightarrow \left\{ w_0 + w_1 x_1 + w_2 x_2 + \dots + w_d x_d = 0 \right.$

$\pi_2 \rightarrow \left\{ \overbrace{\boxed{\frac{w_0}{c}}}^{w_0'} + \overbrace{\boxed{\frac{w_1}{c}}}^{w_1'} x_1 + \overbrace{\boxed{\frac{w_2}{c}}}^{w_2'} x_2 + \dots + \overbrace{\boxed{\frac{w_d}{c}}}^{w_d'} x_d = 0 \right.$

dist: $\left\{ \dfrac{w_0/\cancel{c}}{\sqrt{\left(w_1/\cancel{c}\right)^2 + \left(\frac{w_2}{\cancel{c}}\right)^2 + \dots + \left(\frac{w_d}{\cancel{c}}\right)^2}} \right.$





$$L_1: \quad \overbrace{2x + 4y}^{w^T x} + \overbrace{6}^{w_0} = 0$$

$$L_2: \quad x + 2y + 3 = 0$$

$$x^*, y^* = \min_{x,y} f(x,y)$$

s.t. $g(x,y) = c$

$$\Downarrow$$

$$g(x,y) - c$$

(x_{10}, y_{10})

$$\approx \; x^*, y^* \left[\min_{x,y} f(x,y) + \lambda \big(g(x,y) - c \big) \right]$$

$\lambda \geq 0$

Lagrange Multiplier

$$\mathcal{L}(x,y,\lambda) = f(x,y) + \lambda\,(g(x,y) - c)$$

Lagrangian fn

eq. constraint
opt. problem

$$\min_{w,w_0} \overbrace{\sum_{i=1}^{n} -y_i(w^T x_i + w_0)}^{f(w,w_0)}$$

$$\text{s.t } \|w\|^2 = 1$$

$$\underbrace{\|w\|^2 - 1}_{g(w,w_0)} = 0 \rightarrow \underbrace{0 \cdot w_0 + w_1^2 + w_2^2 + \dots w_d^2}_{} - 1 = 0$$

w_0^*, w^*

Unconstrained opt :

$$\min_{w, w_0} \sum_{i=1}^{n} -y_i (w^T x_i + w_0) + \boxed{\lambda} (\|w\|^2 - 1)$$

$\lambda > 0$

w^*, w_0^*

x^*, y^*

$$\min_{x,y} \boxed{f(x,y)}$$

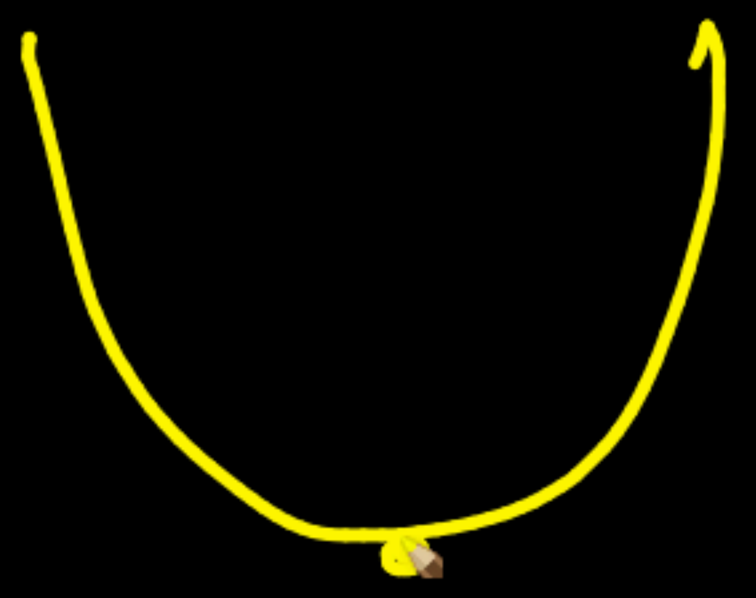
$$\text{s.t. } g(x,y) = c$$

Const. opt

$$\min_{x,y} \boxed{f(x,y)} + \lambda \underbrace{(g(x,y) - c)}_{=0}, \quad \lambda \geq 0$$

Unconst. opt. prob

$$x^2 + y^2$$



$(0,0)$

Chrome File Edit View History Bookmarks Profiles Tab Window Help 22:49

GD:

$$\min_{x,y} \; f(x,y) + \boxed{\lambda}\,(g(x,y) - c)$$

optmzn

$\lambda \geq 0$

hyper-param

overfitting & underfitting

x_0, y_0 ← randomly

x_1, y_1

⋮

x_k, y_k

42 / 42

ML

$$\min_{w, w_0} \sum_{i=1}^{n} -y_i (w^T x_i + w_0)$$

s.t $\|w\|^2 = 1$

$\lambda = 10 \geq 0$

$\lambda \geq 0$

how to solve

$$\min_{w, w_0} \sum_{i=1}^{n} -y_i (w^T x_i + w_0) + \boxed{\lambda} (\|w\|^2 - 1)$$

GD → best w^*, w_0

43 / 48

hyper-param Search

$\lambda = 0 \rightarrow w^0, w_0^0 \rightarrow \pi^0 \rightarrow$ Mis-classfn rate

$\lambda = 1 \rightarrow w^1, w_0^1 \rightarrow \boxed{\pi^1} \rightarrow 1\%$

$\lambda = 10 \rightarrow w^{10}, w_0^{10} \rightarrow \pi^{10}$

$\lambda = 20 \qquad \vdots \qquad \rightarrow \boxed{\pi^{20}} \rightarrow 1\%$

$\vdots \qquad \vdots \qquad \vdots$

Chrome File Edit View History Bookmarks Profiles Tab Window Help 23:04

Recap:

$\boxed{w, w_0}$

den $||w||$ → cumbersome $\frac{\partial L}{\partial w_j}$

Constr. opt. problem $||w||^2 = 1$

Lagrange Multiplier

Unconstr. opt. problem $\boxed{\lambda}$ $\lambda \geq 0$

$\lambda = 0$

$\lambda = 1$

$\lambda = 10$

$\lambda = 20$

⋮

1000's of var

$\boxed{GD}$

45 / 47

baulé free $\rightarrow$ $\underbrace{w_0, w_1 \, \ldots \, w_d}_{d+1\text{-var}}$

$\boxed{\lambda}$

$\downarrow$

100

$\downarrow$

random

$$\begin{cases} \lambda = 0.1 \\ \lambda = 1 \\ \lambda = 10 \\ \lambda = 100 \end{cases}$$

→ hyper-param tuning

hyper-param Search (λ)

Lagrange fn $\longrightarrow \mathcal{L}(x,y,\lambda)$

Constr opt $\longrightarrow$

$$\min_{x,y}\ \overset{\log}{f(x,y)} + \underset{\lambda \geq 0}{\boxed{\lambda}}\left(\overset{e\ldots}{g(x,y)} - c\right)$$

Unconstr opt

$$\frac{\partial \mathcal{L}}{\partial x} = 0 \longrightarrow \text{one eqn} \quad x, y, \lambda$$

$$\frac{\partial \mathcal{L}}{\partial y} = 0 \longrightarrow \text{2nd eqn} \quad x, y, \lambda$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \longrightarrow \text{3rd eqn} \quad x, y, \lambda$$

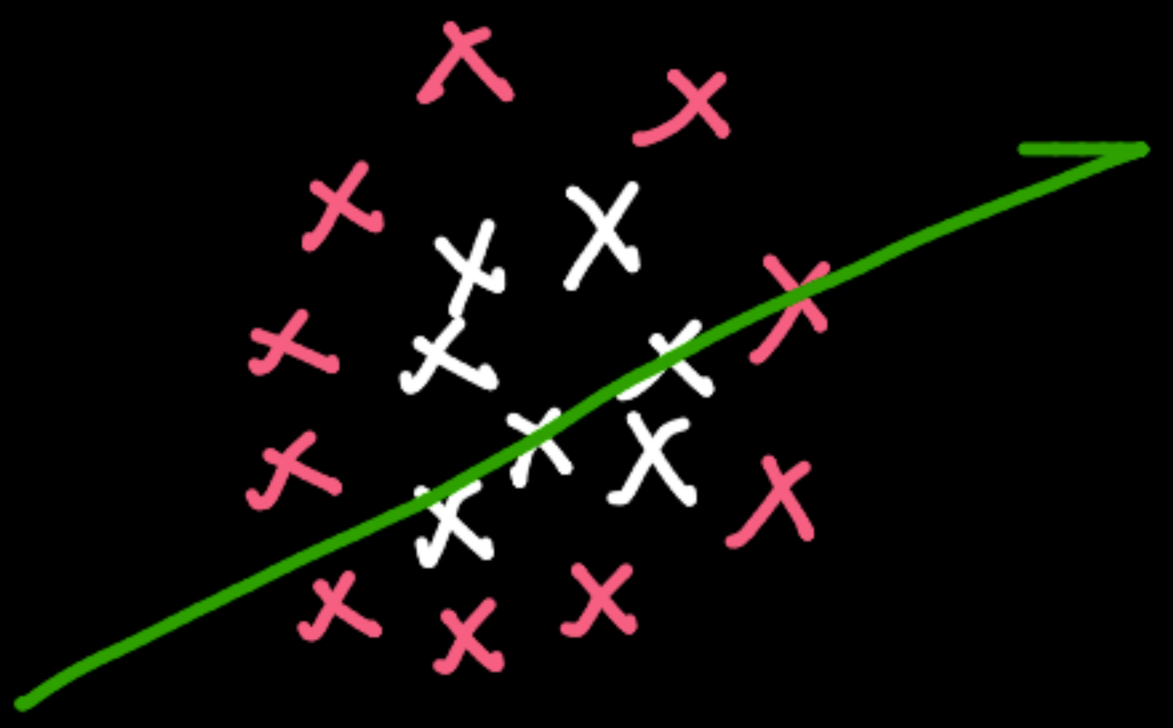
$\longrightarrow x, \lambda, y$

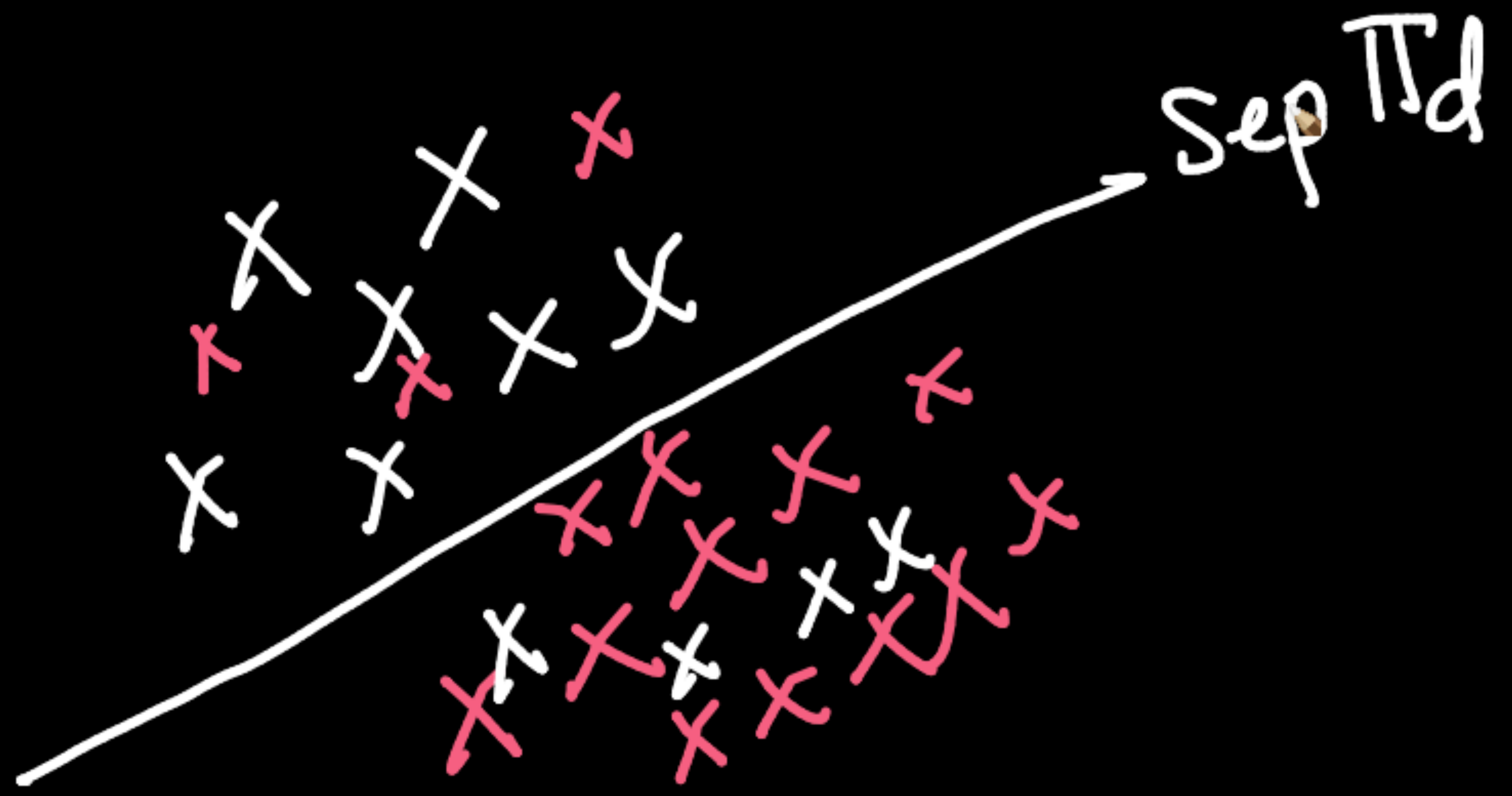
Data → ✓

Model → Sep Tld

hyper-param-tuning

(λ) ✓




Sep

Chrome File Edit View History Bookmarks Profiles Tab Window Help 23:17

$\frac{dL}{\partial w_i}$

den $||w||$ $\longrightarrow$ Constr. opt $\longrightarrow$ Uncnst. opt

Cumbersome

s.t $||w||^2 = 1$

hyper-param

$$\min_{w, w_0} \sum_{i=1}^{n} -y_i (w^T x_i + w_0) + \boxed{\lambda} (||w||^2 \cancel{-1})$$

scalar

$$\mathcal{L}(w, w_0, \lambda)$$

51 / 52

$x=0$, $y=0$ $\leftarrow$ $\min \; x^2+y^2 \boxed{+c}$

$x=0$, $y=0$ $\leftarrow$ $\min \; x^2+y^2$

Chrome File Edit View History Bookmarks Profiles Tab Window Help 23:26

$$\min_{w, w_0} \sum_{i=1}^{n} -y_i(w^T x_i + w_0) + \lambda \|w\|$$

$$w_0^*, w_1^*, \dots w_d^* = \min_{w_0, w_1 \dots w_d} \sum_{i=1}^{n} -y_i\left(w_0 + \boxed{w_1} x_{i1} + w_2 x_{i2} + w_j x_{ij} + w_d x_{id}\right) + \boxed{\lambda}\left(w_1^2 + w_2^2 + \boxed{w_j^2} + w_d^2\right)$$

$\lambda = 10$ (let)
(const)

53 / 56

$$\frac{\partial L}{\partial w_0} = \boxed{\sum_{i=1}^{n} -y_i \cdot 1} + \lambda \cdot 0 = \sum_{i=1}^{n} -y_i$$

batch GD

$\forall j = 1, 2, \dots, d$

$$\frac{\partial L}{\partial w_j} = \boxed{\sum_{i=1}^{n} \left(-y_i \, x_{ij}\right) + \boxed{\lambda \left(2 \cdot w_j\right)}}$$

n : 1 Million pts

d+1

Update eqn

$w_0 \leftarrow$

$\forall j = 1,2,\ldots d \quad w_j \leftarrow$

m-iterations

$\boxed{\lambda} \rightarrow w_0, w_1 \ldots w_d$

Chrome File Edit View History Bookmarks Profiles Tab Window Help 23:28

GD:

In each iteration (m-iter)

$\frac{\partial L}{\partial w_j}$ → go through n datapoints

$= 10^5$
100,000 pts

d+1 partial der → 10^7 = 10 Million

(100) $= 10^2$

56 / 56

In each iteration

for each update-eqn → dH-variables

$$\frac{\partial L}{\partial w_j} \leftarrow \boxed{\text{all data pts}}$$

hack/comp. tech: ?

Chrome File Edit View History Bookmarks Profiles Tab Window Help 23:33

batch GD

$$\frac{\partial L}{\partial w_j}$$

all n data points

sample m points randomly

mini batch SGD

estimate of $\frac{\partial L}{\partial w_j}$ → reasonably well

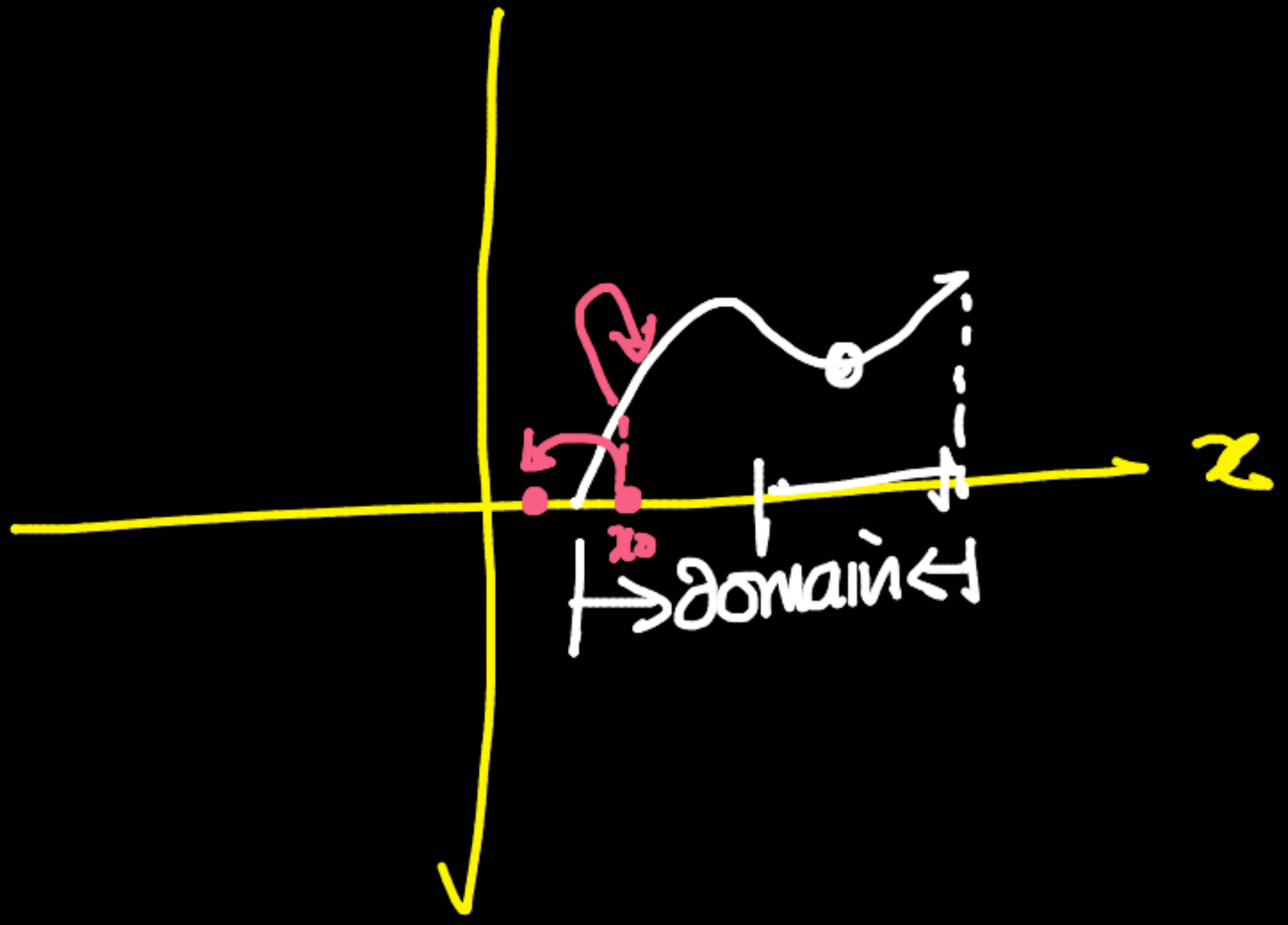
58 / 59

(1000 people) → $h_1, h_2, \ldots h_{1000}$

mean-hght-1000

100 random

mean-hght-100



$$x_1 = x_0 - \eta \left[\frac{\partial f}{\partial x}\right]_{x_0}$$

w

boundary case:

if x_1 outside the domain

→ x_1: random ✓

memory

η

smartly

adv-GD (DL)

λ

100 features

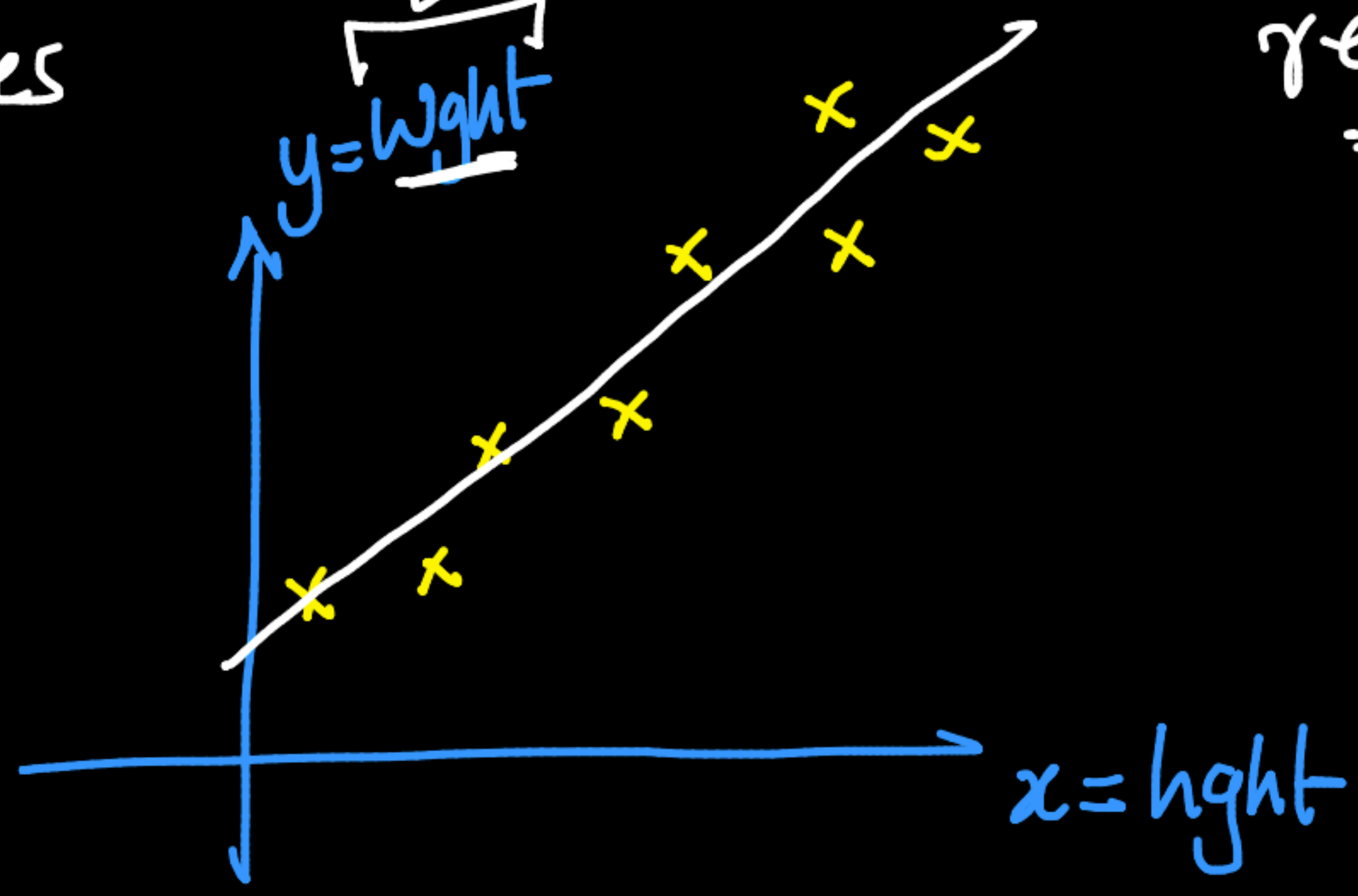
101-dim

hacks

PCA

numerical/real-valued

regression



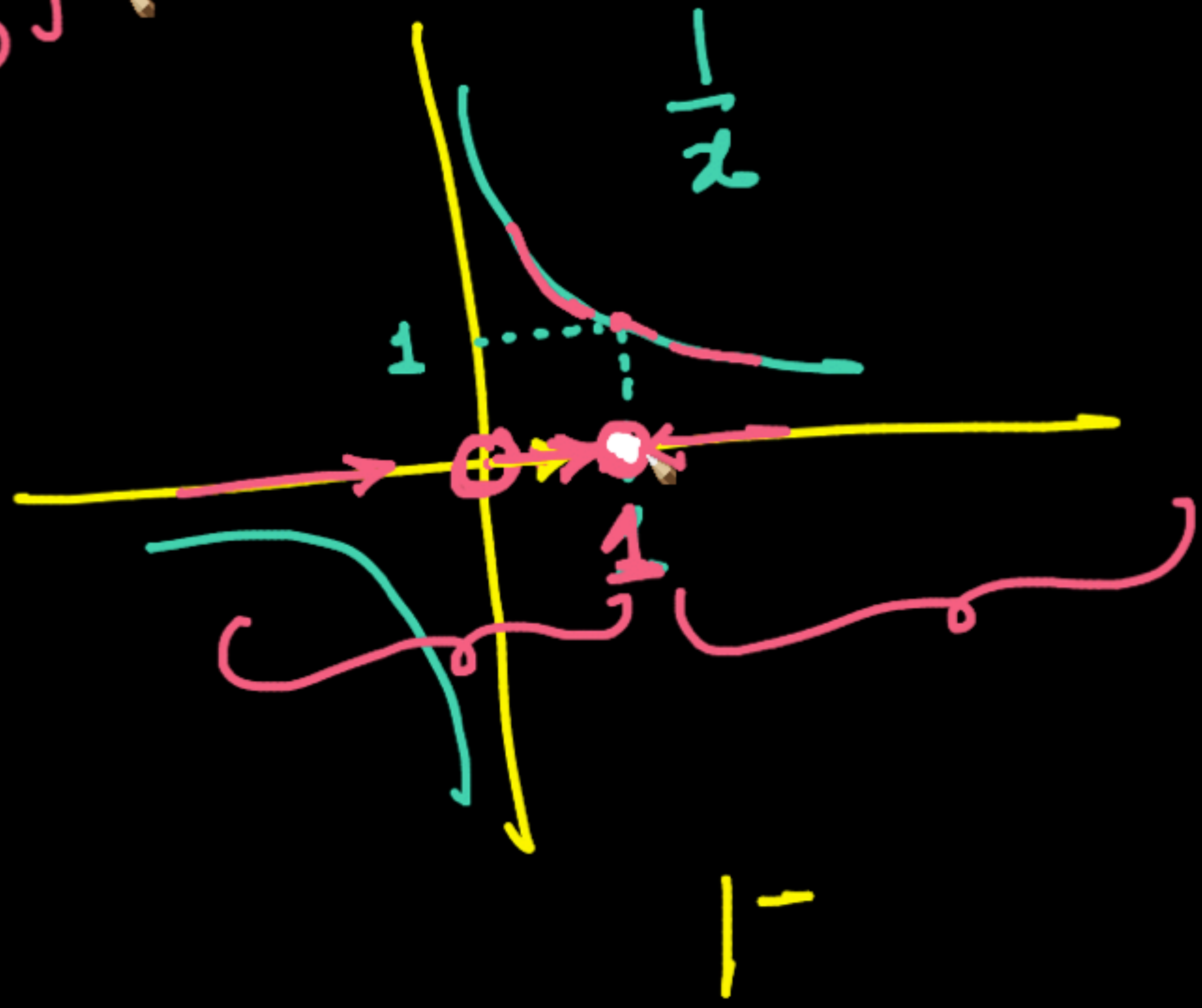
gender=Male

$f(h)=w$

one-var/feature

(Q) $f(x) = \boxed{\frac{1}{x}}$

$\lim_{x \to 0} f(x)$



Notation $\left\{ \lim_{x \to 1} f(x) = \boxed{1} \right.$

2-sided

$$\lim_{x \to 1^+} f(x) = 1$$

$$\parallel$$

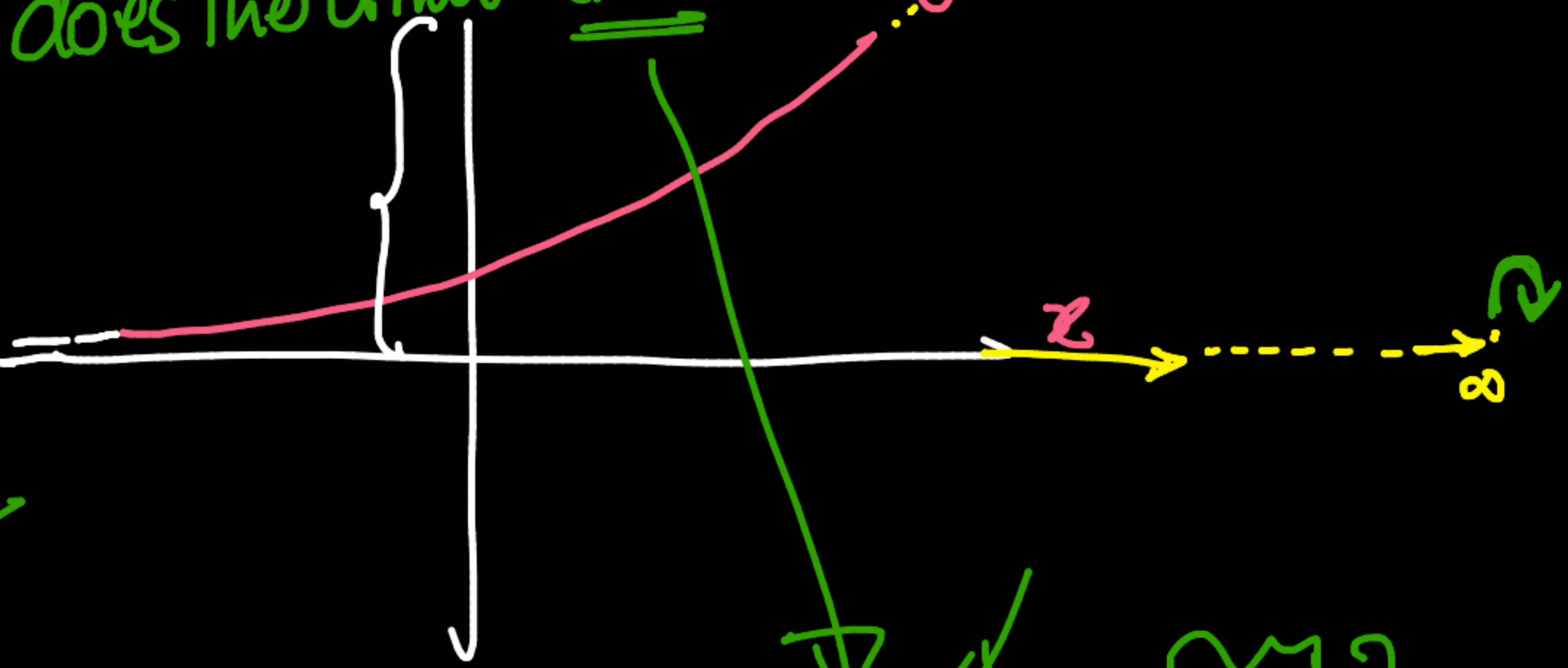
$$\lim_{x \to 1^-} f(x)$$

does the limit exists

$$\boxed{\lim_{x\to\infty} e^x = \infty}$$ ✓

RSL $-\infty$ LSL

$$\boxed{\lim_{x\to-\infty} e^x = 0}$$ ✓



dom: $(-\infty, \infty)$

range: $(0, \infty)$

LSL = RSL

numeric

features: state, hght $= \langle s_1, s_2 \ldots s_{30}, hght \rangle$

$s_1, s_2 \ldots s_{30}$

output-var: wght

numeric → regr.

Categorical → Classfn

wght

numeric

regression

Log-Reg

Softmax

Lr-Reg

SVM → constr. opt

MLP/NN

CNN

LSTM/GRU

Transformer

⋮

→ Optmzn

$\lambda ||w||$

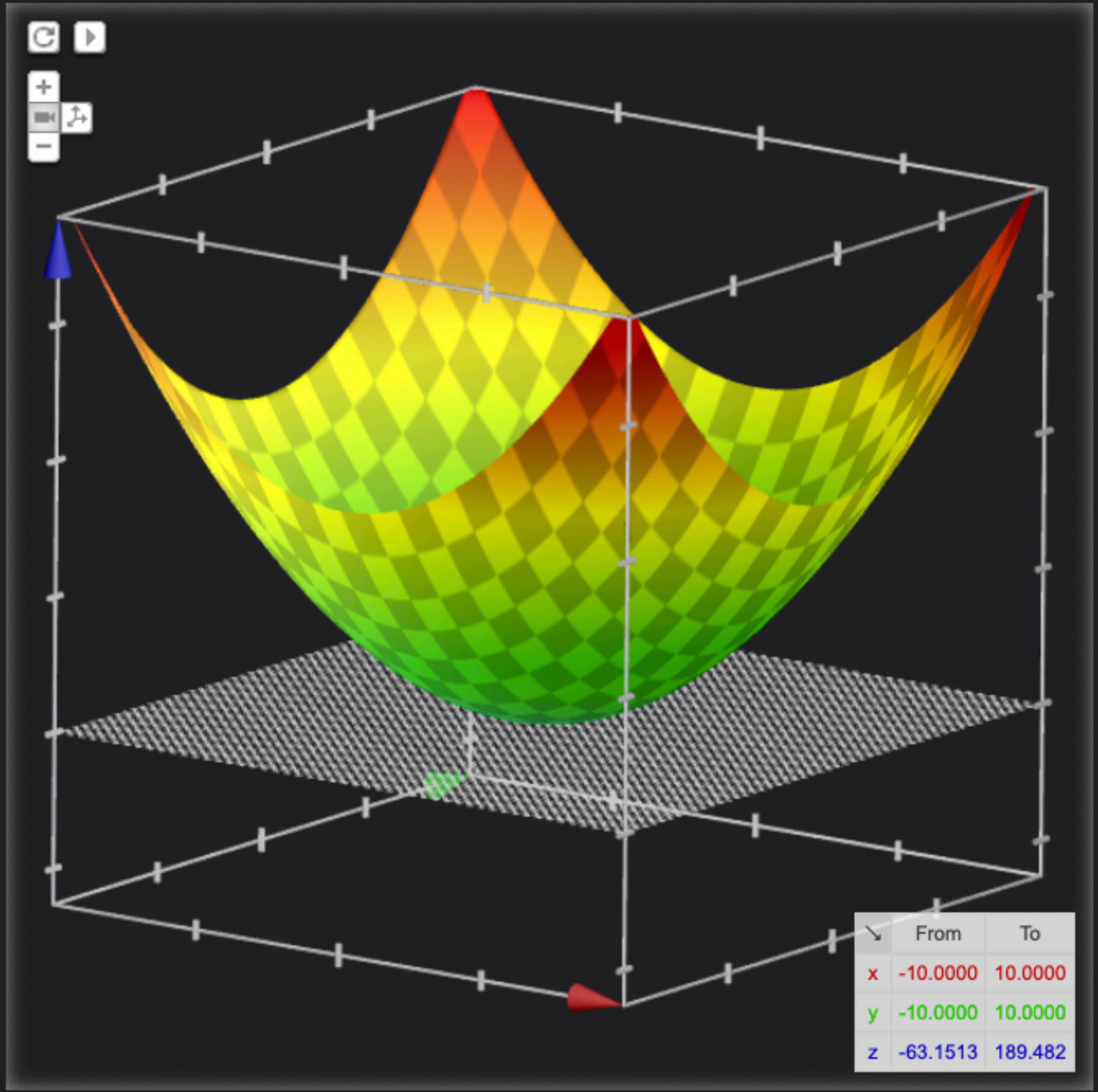
[geom/prob .....]

plot(x^2+y^2)

All Images News Maps Videos More Tools

About 81,20,00,000 results (0.55 seconds)

Graph for x^2+y^2



	From	To
x	-10.0000	10.0000
y	-10.0000	10.0000
z	-63.1513	189.482