Gradient Descent

29 July 2020 03:48 PM

Starting @ 4:10 PM

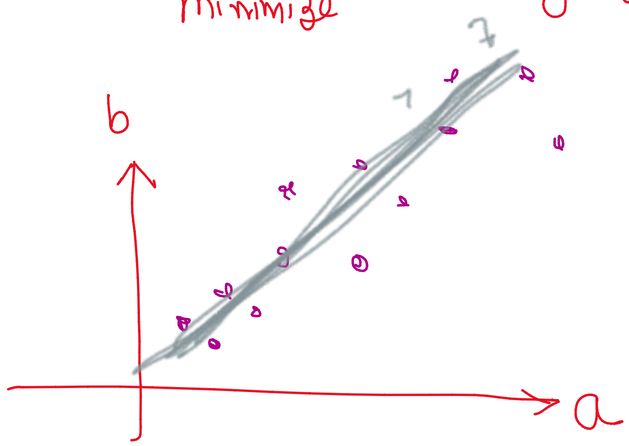
Topic ↓

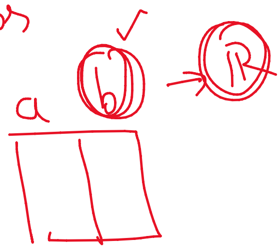
- [Gradient Descent]

Linear Reg ↓

$$\underbrace{\min}_{\text{minimize}} \left(\frac{1}{N} * \sum_{i=1}^{N} (y_i - \hat{y}_i)^{2} \right)$$

avg squared errors

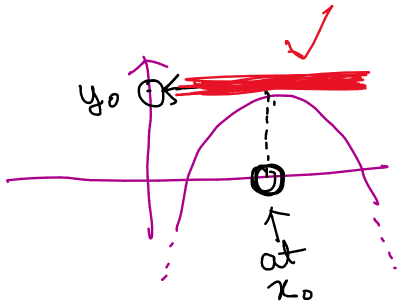




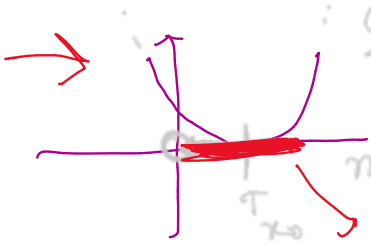
Task → Find the BEST FIT LINE

differ

$$\underbrace{m^*, c^* = \underset{m,c}{\arg\min} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left(y_i - (mx_i + c)\right)^{2} \right\}}_{\text{Optimization formula for Linear Reg}}$$



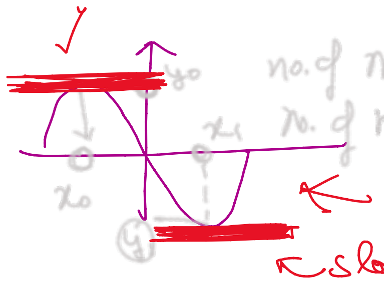
maxima $= y_0$
No minima



$\begin{cases} \text{minima} = 0 \\ \text{at } x_0 \end{cases}$

no maxima

slope $= 0$

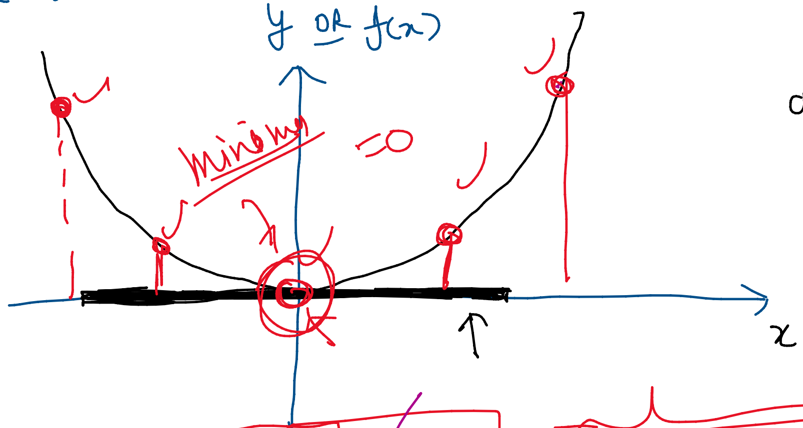


no. of minima $= 1$
no. of maxima $= 1$

slope $= 0$

at minima OR maxima
if we try to plot
a tangent, its slope
is going to be 0

$f(x) = x^2$ OR $y = x^2$



No maxima
at $x = 0$
minima $= 0 \leftarrow y$

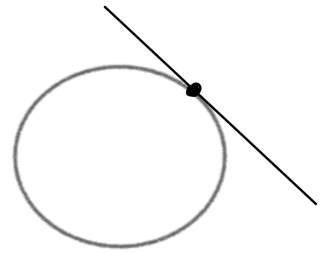
dif

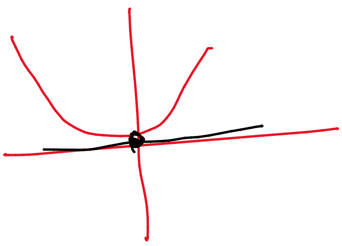
at x_y — slope of tangent $= 0$

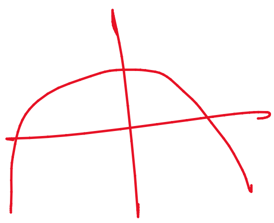
minima pt. exists

$\left[\frac{df}{dx}\right]_{x_y}$

Curve







slope of tangent $= 0$ at minima or.

✓ Slope of tangent = 0 at minima or maxima

$\boxed{f(x)}$

$$\boxed{\frac{d f(x)}{dx} = 0}$$

Ex: $f(x) = x^2 - 3x + 2$

$\boxed{\text{minima OR maxima}}$ ✓

$$\frac{d(x^2 - 3x + 2)}{dx} = 0$$

$$\boxed{\frac{dx^2}{dx}} - \boxed{\frac{d\,3x}{dx}} + \boxed{\frac{d2}{dx}} = 0$$

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

$$2x - 3 = 0$$

$$x = 3/2 = \boxed{1.5}$$

$f(x) = \boxed{x^2 - 3x + 2}$ → { Slope of tangent = 0 at $x = 1.5$ }

$\boxed{2}$

→ $\boxed{\text{minima OR max}}$

Double Differentiation → ① $\boxed{\frac{d^2 f}{dx} < 0}$ → $\boxed{\text{maxima}}$

② $\boxed{\frac{d^2 f}{dx} > 0}$ → $\boxed{\text{minima}}$ ✓

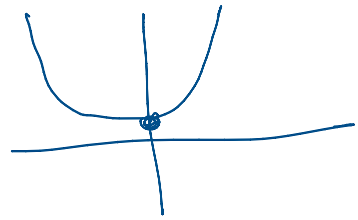
2nd

$$\frac{d}{dx}\left(x^2 - 3x + 2\right) \cancel{\neq 0}$$

$$\frac{d}{dx}\left(2x - 3\right) = \boxed{2 * \frac{dx}{dx}} - \boxed{\frac{d3}{dx}}$$

$$= \boxed{2} \checkmark$$

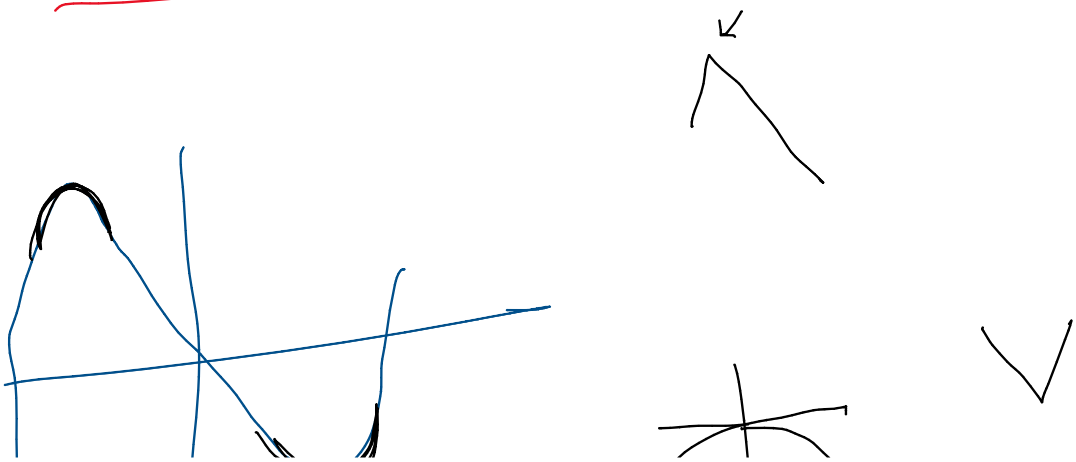
$$x^2 - 3x + 2$$

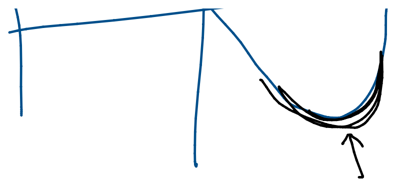


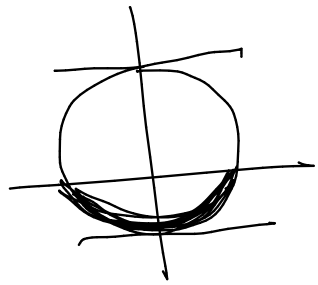
Linear Reg

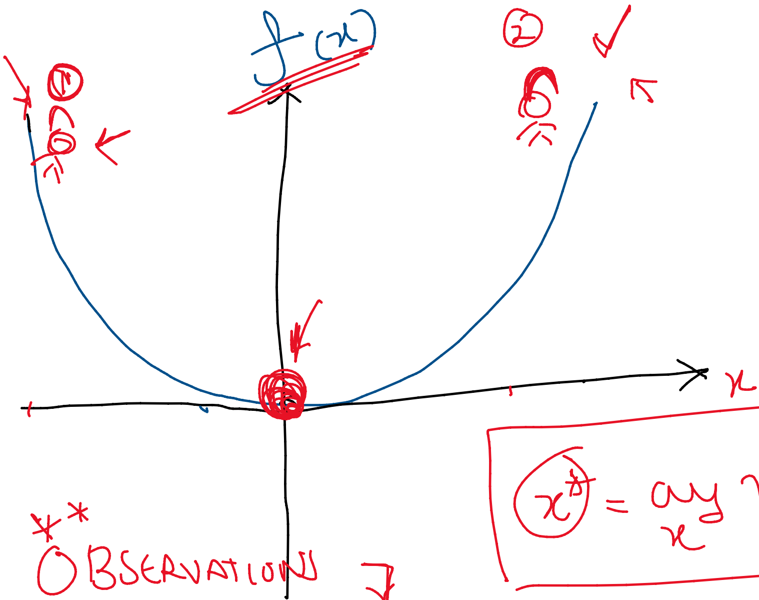
$$m^*, c^* = \arg\min_{m,c} \left\{ \frac{1}{N} \sum \left(y_i - (mx + c) \right)^2 \right\}$$

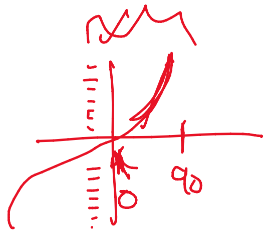
{ 5:11 — 5:15 → BREAK }











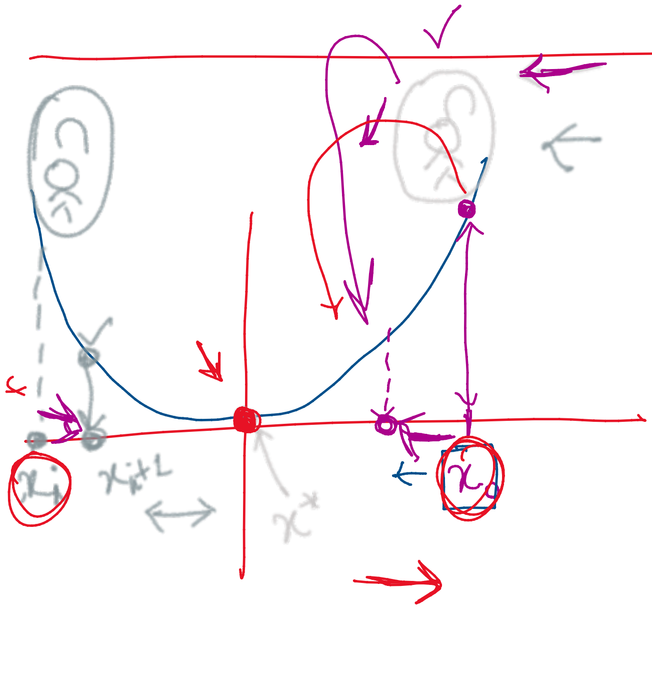
$$\left(\frac{df}{dx}\right)_{x_0}$$

$$\boxed{x^* = \underset{x}{\arg\min}\ f(x)}$$

** OBSERVATIONS ⇒

Minima: ⇒

- Towards one side → slope is +ve | Towards another side slope is negative
- As you move towards minima → slope will reduce
- Slope changes it's sign at minima



Task → Reach the minima pt. x^*

1. Pick any initial pt x_i at Random
2. Pick x_{i+1}, such that x_{i+1} is closer to x^*

x_{i+1} is closer to x^* ✓

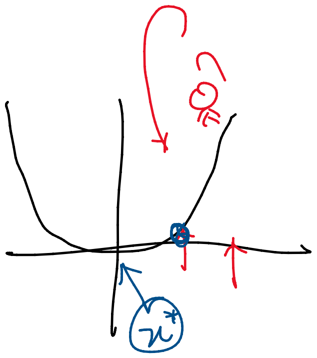
To travel right d x^*

Left

$x_{i+1} = \boxed{x_i} - (+ve)$

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

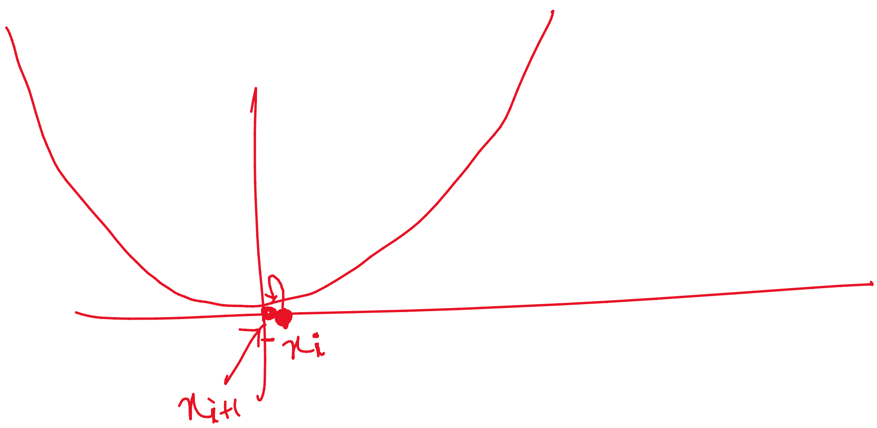
Jumping $\left\{ x_{i+1} = x_i - \eta * \frac{df}{dx} \right.$



$$x_{i+1} = x_i - \eta * \left[\frac{df}{dx}\right]_{x_i}$$

η → Step size OR Learning rate

3. If $(x_{i+1} - x_i)$ ✓ is very small then terminate. Otherwise recompute Step -2 again



GRADIENT DESCENT 7

Algo :

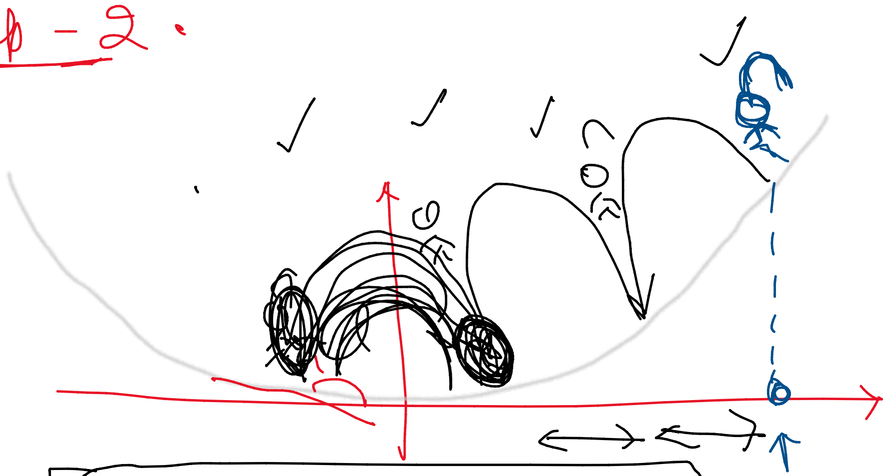
1. Pick a point x_i @ Random

2. Pick x_{i+1} , such that x_{i+1} is closer to minima x^*.

$$x_{i+1} = x_i - \eta \left[\frac{df}{dx}\right]_{x_i}$$

Converging

3. If $\boxed{(x_{i+1} - x_i)}$ is a very small value then terminate loop. Otherwise repeat step-2.



0

$$\boxed{x_{i+1} = x_i - \eta \left[\frac{df}{dx}\right]_{x_i}}$$

Logistic Reg

X_train y_train

Linear Reg ✓

Gradient descent

Logistic

from sklearn.linear_model import LinR

Regression ( )

descent

from sklearn.linear ...

reg = LinearRegression()

reg.fit(X_train, y_train)

Linear Reg → $\frac{1}{N}\sum$ Loss

Log Reg → $\sum$(Sigmoid( ))

X_train
y_train

ML algorithm

Gradient descent	Optimization formula

Linear / Logist / SVM

iteratively

min ( )
max ( )

Linear Reg → Code → wo

m, c

Gradient Descent → SGD

~~Descent~~

η → step size ✓

or

→ learning rate

ML ⇒ Maths + coding

1 month

ML → 1 month ✓

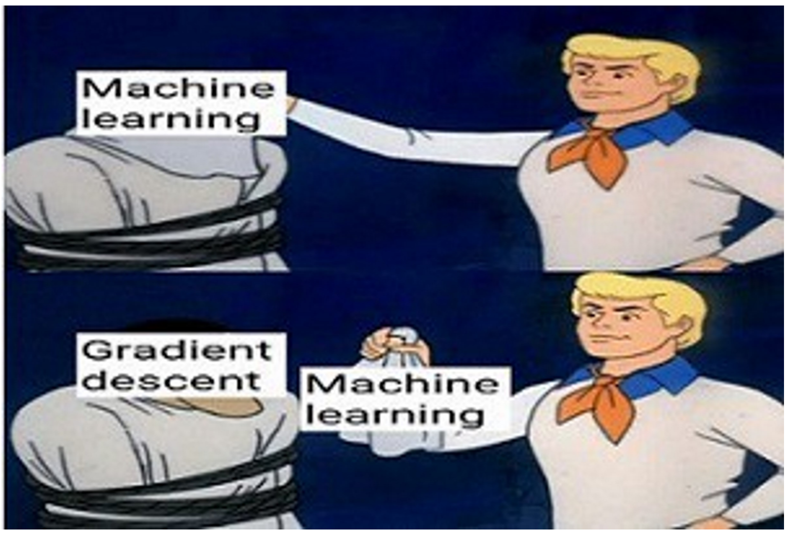
→ Code

Starting @ 4:05 PM

Topic :

→ Gradient Descent

$$\left\{\begin{array}{l} \text{Linear Reg} \rightarrow \\ \text{Logistic Reg} \rightarrow \end{array}\right. \left\{\begin{array}{l} m^*, c^* = \_\_\_\_\_ \\ w^*, w_0^* = \_\_\_\_\_ \end{array}\right\}$$



f

Linear Reg]

$$\hookrightarrow m^*, c^* = \arg\min_{m,c} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left(y_i - (mx_i + c) \right)^2 \right\}$$

GD]

1. Pick inital points at random $\{m_i = 0, c_i = 0\}$

2. Pick m_{i+1} & c_{i+1} such that m_{i+1} & c_{i+1} are closer to m^* & c^*

Update Equation of GD

$$\underset{\text{new}}{x_{i+1}} = \underset{\text{old}}{x_i} - \underbrace{r}_{\text{stepsize OR learning rate}} * \left[\frac{df}{dx}\right]_{x_i}$$

slope of tangent at x_i point

$$\begin{cases} m_{i+1} = m_i - r * \left[\frac{\partial f}{\partial m}\right]_{m_i} \rightarrow [\nabla_m f]_{m_i} \\ c_{i+1} = c_i - r * \left[\frac{\partial f}{\partial c}\right]_{c_i} \rightarrow [\nabla_c f]_{c_i} \end{cases}$$

Update eq of GD for Linear Reg

$$\begin{cases} m_{i+1} = m_i - r * [\nabla_m f]_{m_i} \\ c_{i+1} = c_i - r * [\nabla_c f]_{c_i} \end{cases} \quad - (1)$$

$$\boxed{f = \frac{1}{N} \sum_{i=1}^{N} \left\{ \left(y_i - (mx_i + c) \right)^2 \right\}}$$

from eqⁿ (1)

$$\hookrightarrow \nabla_m f = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \left(\boxed{y_i - (m x_i + c)}\right)^2}{\partial m}$$

$$= \frac{1}{N}\sum_{i=1}^{N} 2(y_i - (m x_i + c)) * \frac{\partial (y_i - (m x_i + c))}{\partial m}$$

$$= \frac{1}{N}\sum_{i=1}^{N} 2(y_i - (m x_i + c)) * \underbrace{\left\{ \boxed{\frac{\partial y_i}{\partial m}} - \boxed{\frac{\partial m x_i}{\partial m}} - \boxed{\frac{\partial c}{\partial m}} \right\}}$$

$0 \qquad -x_i \frac{\partial m}{\partial m} \to 1 \qquad 0$

$$= \frac{1}{N}\sum_{i=1}^{N} 2(y_i - (m x_i + c)) * \{-x_i\}$$

$$\boxed{\nabla_m f = \frac{2}{N}\sum_{i=1}^{N} (y_i - (m x_i + c)) * \{-x_i\}}$$

$$\nabla_c f = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \left(\boxed{y_i - (m x + c)}\right)^2}{\partial c}$$

$$= \frac{1}{N}\sum_{i=1}^{N} 2(y_i - (m x_i + c)) * \frac{\partial (\cancel{y_i} - \cancel{(m x_i} + c))}{\partial c}$$

$0 \qquad 0 \qquad \to -1$

$$\boxed{\nabla_c f = \frac{2}{N}\sum_{i=1}^{N} (y_i - m x_i + c) * \{-1\}}$$