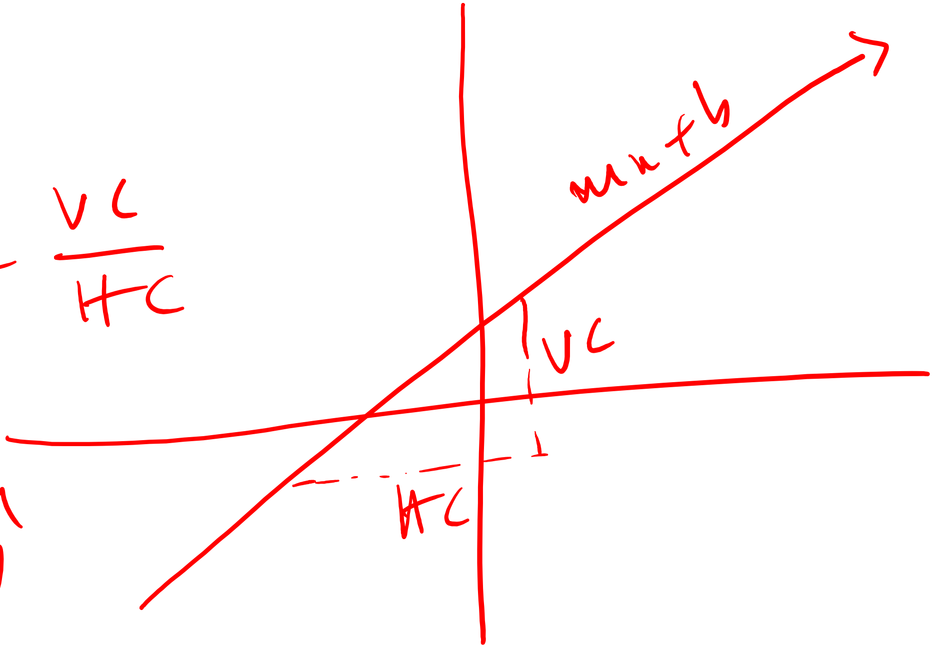
Summary

* $y = mx + b$
  * slope $\longrightarrow$ m
  * intercept $\longrightarrow$ b

$$m = \frac{VC}{HC}$$



* Least squares

x	y	$\hat{y}$
x_1	y_1	$\hat{y}_1$
$\vdots$	$\vdots$	$\vdots$
x_n	y_n	$\hat{y}_n$

$$SSE(m,b) = SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - (m x_i + b)\right)^2$$

$$\min_{(m,b)} SSE(m,b) \longrightarrow \hat{m}, \hat{b}$$

Question

$f(x) = x^2 - 2 + 2x$ on $\mathbb{R}$ $\longrightarrow$ $f'(x) = 2x + 2 == 0$

$+ x^3$

$\hookrightarrow x = -1$

1. What is the minimum of f on $\mathbb{R}$

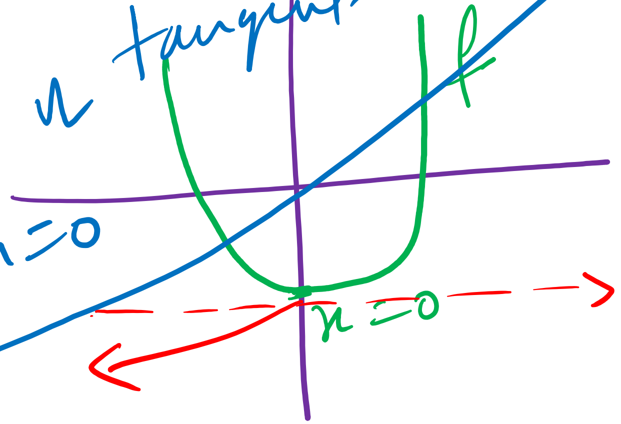
$\hookrightarrow -2$

2. Where does that minimum occur in $\mathbb{R}$?

$f(x) = x^2 - 2$

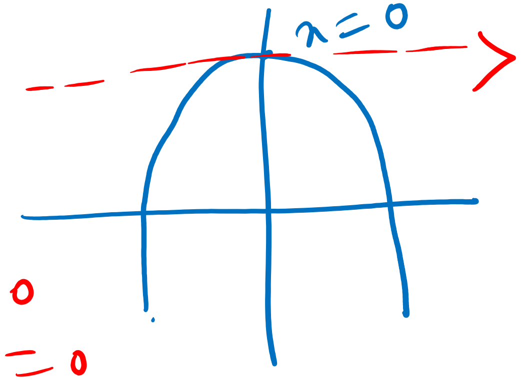
$g(x) = 2 - x^2 \longleftarrow$ max of g

Slope of tangent to f at $x=0$



$f'(x_0) = 0$

$f'(0) = 0$



Theorem: Let f be a diff. function on $\mathbb{R}$. If f attains its min/max at x_0, then $f'(x_0) = 0$

$f'(x) = 2x == 0$

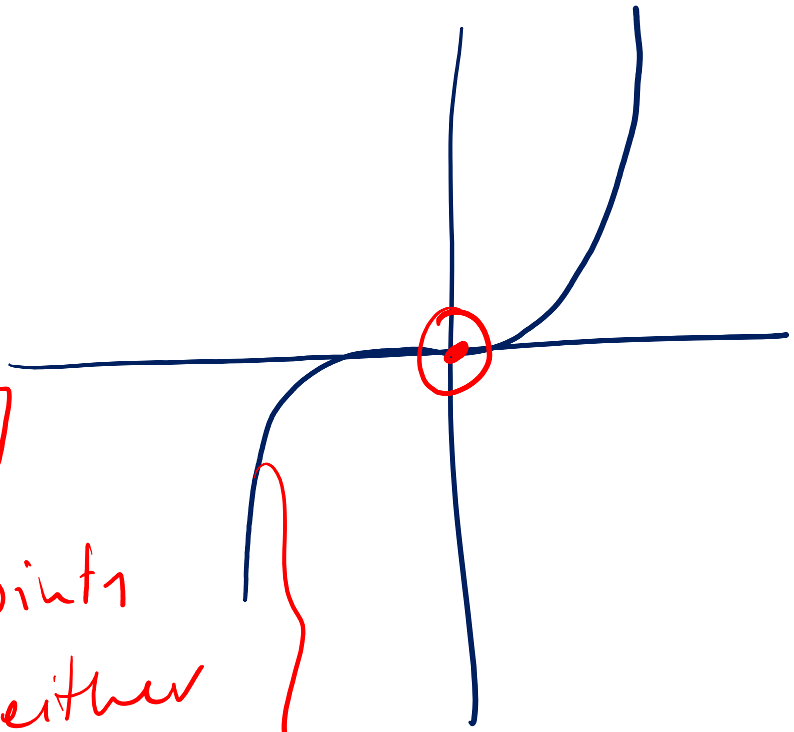
$\hookrightarrow x = 0$

Theorem: f be a diff on $\mathbb{R}$

If $f'(x_0) = 0$ for some $x \in \mathbb{R}$, this does not mean

f has minimum/max at "x_0"

$$f(x) = x^3$$



Stationary Points $= \{ x : f'(x) = 0 \}$

Saddle Points $= \{ x \in$ Stationary Points but x is neither min/max $\}$

Let ~~θ~~ x_0 be a stationary point & not a saddle point of f.

~~Solution~~ $f'(x_0) = 0$.

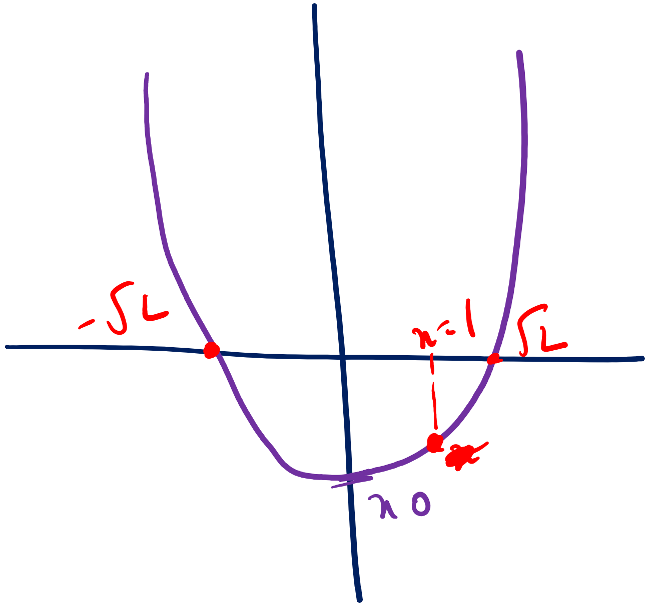
Suppose $f''(x_0) \geq 0$ then f has minimum at x_0

Suppose $f''(x_0) \leq 0$, then f has maximum ~~minimum~~ at x_0.

Gradient Descent $f(x) = x^2 - 2$

Step 1: $x_0 = 1$

$f'(0) = 0$



$f'(x_0) = 2x_0 = 2\cdot 1 = 2$

$x_1 = x_0 - \alpha f'(x_0)$

$= 1 - 0.1 \times 2$

$= 0.8$

$\boxed{\alpha = 0.1}$ $\downarrow$ Learning Rate

$f'(x_1) = 1.6$

$x_2 = x_1 - \alpha \cdot f'(x_1)$

$= 0.8 - 0.1 \times 1.6$

$= 0.8 - 0.16 = 0.64$

$f'(x_2)$

$= 1.28$

$x_{i+1} < x_i < \dots < x_2 < x_1 < x_0$

$f'(x_2) < f'(x_1) < f'(x_0)$

If $|x_i - x_{i+1}| < \varepsilon$

Then stop iteration.

Gradient Descent Algorithm

Initializers: $x_{old} = 1$, $\alpha = 0.1$, $\varepsilon = 10^{-8}$, $x_{new} = x_{old} + 2\varepsilon$

$\delta = 10^{-4}$

$$f'(x_{old}) \approx \frac{f(x_{old}+\delta) - f(x_{old})}{\delta}$$

while $(|x_{new} - x_{old}| > \varepsilon)$

{

$x_{old} = x_{new}$

$x_{new} = x_{old} - \alpha \cdot f'(x_{old})$

}

$x_n \longrightarrow x^*$

If f is continuous, & $x_n \longrightarrow x^*$
then $f(x_n) \longrightarrow f(x^*)$

Calculation of parameters in $y = mx + b$

X Year	y [illegible]	$\hat{y}$	error	$(\text{error})^2$
x_1	y_1	$\hat{y}_1$	$y_1 - \hat{y}_1$	
x_2	y_2	$\hat{y}_2$		
$\vdots$				
x_n	y_n	$\hat{y}_n$	$y_n - \hat{y}_n$	

$$x \cdot y = \sum_{i=1}^{n} x_i y_i$$

$$ASSE = ASSE(m, b) = \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - (m x_i + b) \right)^2$$

$$\frac{\partial ASSE}{\partial m} = \frac{1}{\cancel{2}n} \sum_{i=1}^{n} \cancel{2} \cdot \left(y_i - (m x_i + b) \right) (-x_i) = -\sum_{i=1}^{n} \left(y_i - (m x_i + b) \right) x_i$$

$$= -\frac{1}{n} (Y - \hat{Y}) \cdot X$$

$$-\frac{1}{n} (Y - \hat{Y}) \cdot X = 0$$

$$\frac{\partial\, ASSE}{\partial b} = \frac{1}{\cancel{2}n} \sum_{i=1}^{n} \cancel{2}\left(y_i - (m x_i + b) \right) \cdot (-1)$$

Least Square Method

$$= -\frac{1}{n} \sum_{i=1}^{n} \left(y_i - (m x_i + b) \right) = 0$$

$$\Rightarrow -\frac{1}{n} \left(\sum_{i=1}^{n} \left(y_i - (m x_i + b) \right) \right) = 0$$

$$\hookrightarrow \sum_{i=1}^{n} y_i - m \sum_{i=1}^{n} x_i - n \cdot b = 0$$

$$b = \left(\frac{\sum y_i - m \sum x_i}{n} \right)$$

$$m = \frac{n \sum_{i=1}^{n} y_i x_i - \left(\sum_{1}^{n} x_i\right)\left(\sum_{1}^{n} y_i\right)}{n \sum x_i^2 - \left(\sum x_i\right)^2}$$

$$\frac{\partial\, ASSE}{\partial m} = 0$$

$$(Y - \hat{Y}) \cdot X = 0$$

$$\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \hat{y}_i x_i = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} (m x_i + b) x_i = 0$$

Agenda

1 * ~~Leat~~ Find m & b for a given data set in Python

2 * Gradient descent algo in Python

Summary

* Min & Max of a function
* Derivative of f
* Stationary Points & Saddle Points
* Gradient Descent Algo
* Least Square Method → m & b

* Python → m & b
→ Gradient Descent

$$x_{new} = x_{old} - \alpha \cdot f'(x_{old})$$