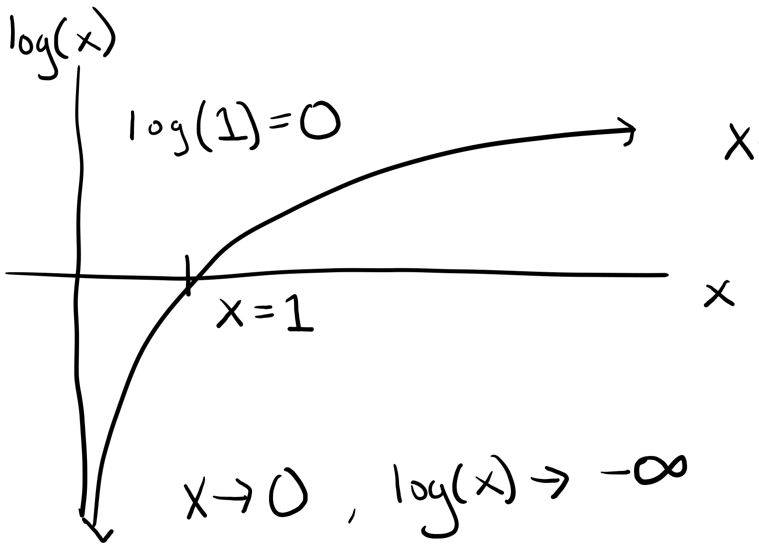
Math Review

$\boxed{\log(x)} \rightarrow$ natural log $\rightarrow \ln(x)$

`log(x)` $\rightarrow$ `log10()`, $\log_{10}(x)$



$x \to \infty$, $\log(x) \to \infty$

$x \to 0$, $\log(x) \to -\infty$

$x > 0$, $\log(x) \to$ negative or positive

$$\log(x \cdot y) = \log(x) + \log(y)$$

$$\log(x^a) = a\log(x)$$

$$\log(x/y) = \log(x \cdot y^{-1}) = \log(x) + \underbrace{\log(y^{-1})}$$

$$\boxed{\log(x/y) = \log(x) - \log(y)}$$

multiplication $\longrightarrow$ addition

division $\longrightarrow$ subtraction

$$\log(x+y) = ? \rightarrow \boxed{\begin{array}{c}\log(x+y) \\ \| \\ \log(x+y)\end{array}} \neq \cancel{\log(x) + \log(y)}$$

Wrong

Inverse of $\log(x)$ is the $\boxed{\exp(x)} = e^x$

`exp(x)` , $\exp\left(-\frac{1}{2}x^2\right)$

$\log(\exp(x)) = x$; $\exp(\log(x)) = x$

$$\log\left(\exp\left(\frac{x^2}{y}\right)\right) = \frac{x^2}{y}$$

Functions & derivatives

$$f(x) = a \cdot x^k$$

multiplication

$\downarrow$

$\times$

First derivative: $\frac{df}{dx} = f' = \frac{d}{dx}\left(ax^k\right) = a \cdot k \cdot x^{(k-1)}$

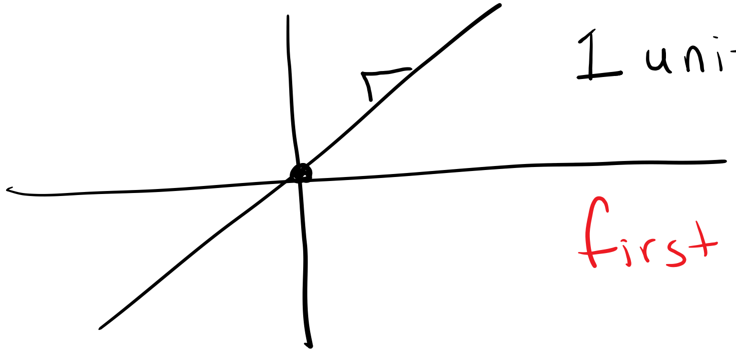
Second derivative: $\frac{d^2f}{dx^2} = f'' = \frac{d}{dx}\left(\frac{df}{dx}\right) = \frac{d}{dx}\left(ak\,x^{(k-1)}\right)$

$$\frac{d^2f}{dx^2} = ak \cdot (k-1) \cdot x^{(k-1)-1}$$

$$\boxed{\frac{d^2f}{dx^2} = ak(k-1)\,x^{(k-2)}}$$

$\boxed{f(x) = x^2}$, $\boxed{f' = 2x}$, $f'' = 2$

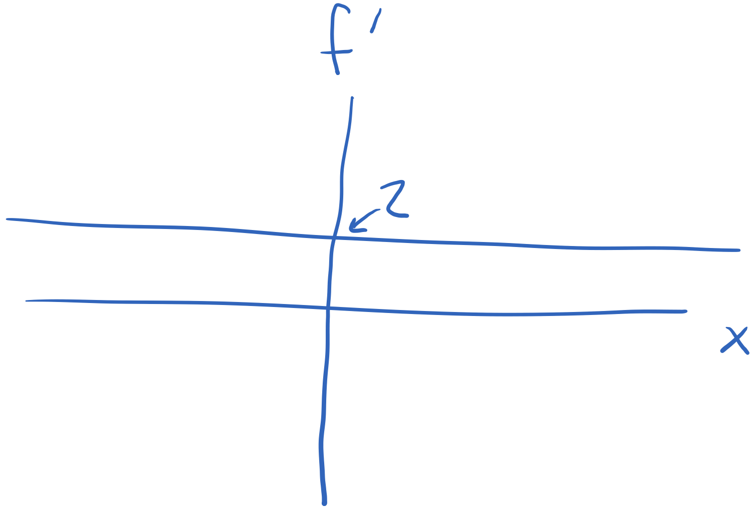
$\boxed{f(x) = 2x}$, $f' = 2$

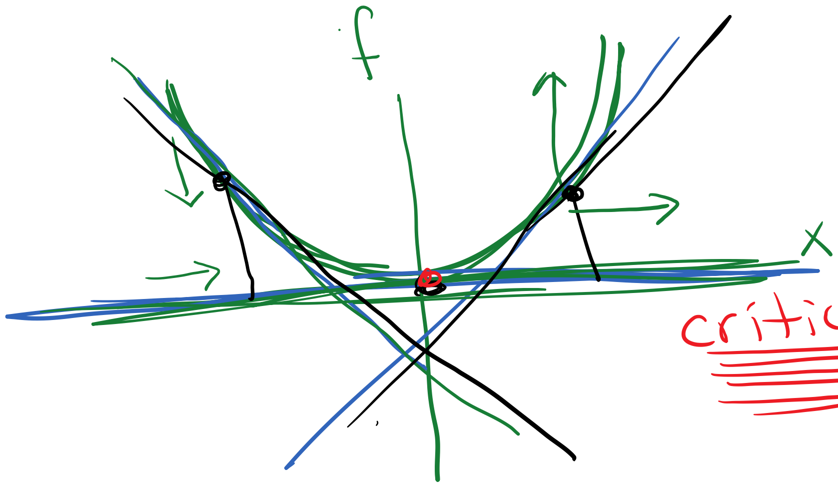


1 unit increase in x

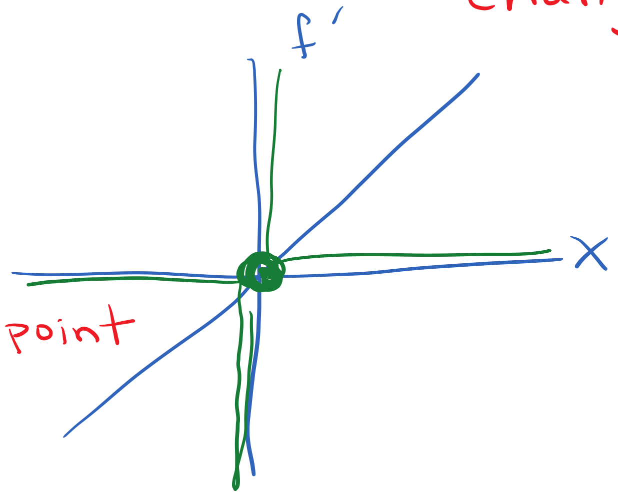
f(x) increases 2 units

first derivative is the **slope** or rate of change





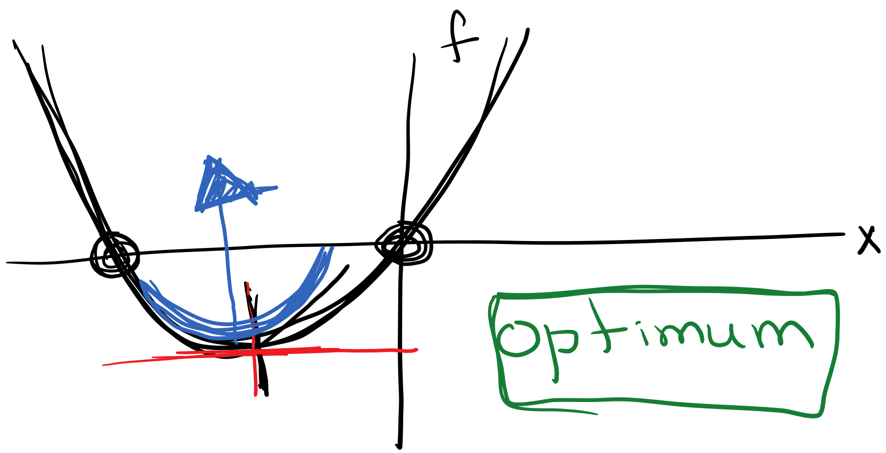
critical point



first derivative is the instantaneous rate of change

tangent

$f(x) = ax^2 + bx$, $\boxed{a > 0}$



$\boxed{\text{optimum}}$

minimum

$f'(x_*) = 0$

$f'(x) = 2ax_* + b = 0$

$2ax_* = -b$

$$\boxed{x_* = \frac{-b}{2a}}$$

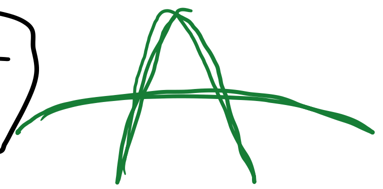
$$\boxed{f''(x) = 2a}$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-b \pm \sqrt{b^2 - \cancel{4a0}}}{2a}$$

$$x = \frac{-b \pm \sqrt{b^2}}{2a} = \frac{-b \pm b}{2a}$$

$$x = \frac{-b+b}{2a} \quad \text{and} \quad x = \frac{-b-b}{2a} = \frac{-2b}{2a}$$

$$\boxed{x = 0} \qquad \boxed{x = \frac{-b}{a}}$$



to know it's a minimum we need the 2nd derivative test

$f'' < 0 \rightarrow$ local max

$f'' > 0 \rightarrow$ local min

$f'' = 0 \rightarrow$ inconclusive

curvature of the function

$$\frac{d}{dx}\left(\log(x)\right) = ? \rightarrow \boxed{\frac{d}{dx}\left(\log(x)\right) = \frac{1}{x}}$$ Remember for machine learning

$$\frac{d}{dx}\left(\log(1-x)\right)$$

calculate the derivative through a function } substitution & the Chain Rule

$$u = 1 - x \Rightarrow \log(1-x) = \log(u) = f$$

$$\frac{df}{dx} = \frac{df}{du} \cdot \frac{du}{dx} = \frac{1}{u} \cdot (-1) = \frac{-1}{u} = \boxed{\frac{-1}{1-x} = \frac{df}{dx}}$$

$$\frac{df}{du} = \frac{d}{du}\left(\log(u)\right) = \frac{1}{u}$$

$$\frac{du}{dx} = \frac{d}{dx}\left(1-x\right) = -1$$

$$f(x) = (1-x)^q \quad ; \quad u = 1-x \quad , \quad f = u^q$$

$$\frac{df}{dx} = \frac{df}{du} \cdot \frac{du}{dx} = q \cdot u^{q-1} \cdot (-1) = -qu^{q-1}$$

$q \cdot u^{q-1}$

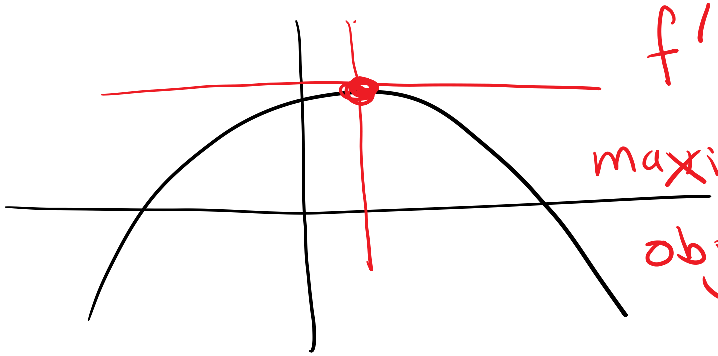
-1

$$\boxed{\frac{df}{dx} = -q(1-x)^{q-1}}$$

$$f(x) = (1-x)^q \Rightarrow q(1-x)^{q-1} \cdot (-1) = -q(1-x)^{q-1}$$

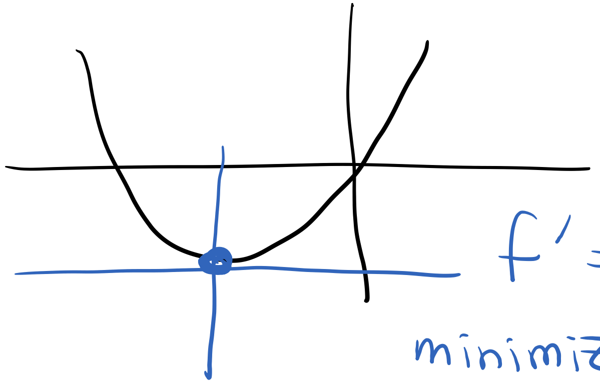
$$\frac{d}{dx}\left(\log(\exp(-x^2))\right) = \frac{d}{dx}(-x^2) = -2x$$

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



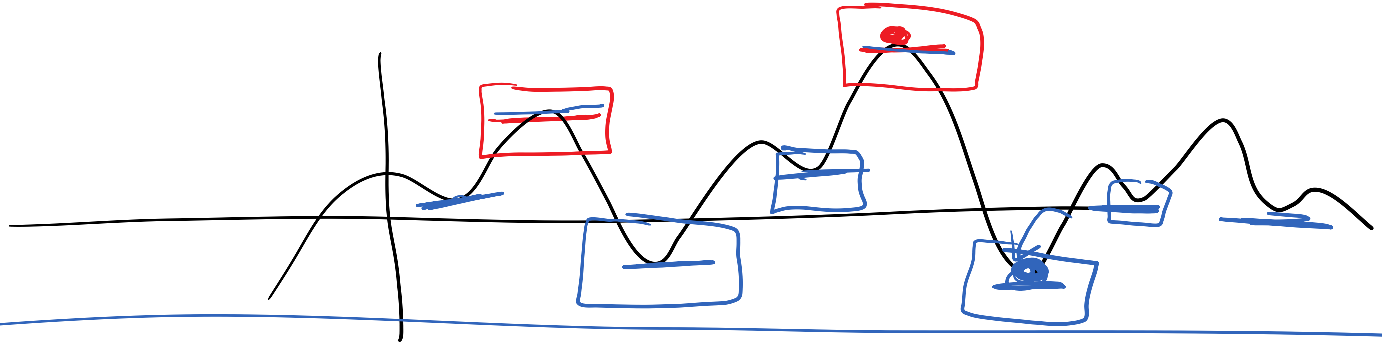
$f' = 0$

maximize an objective



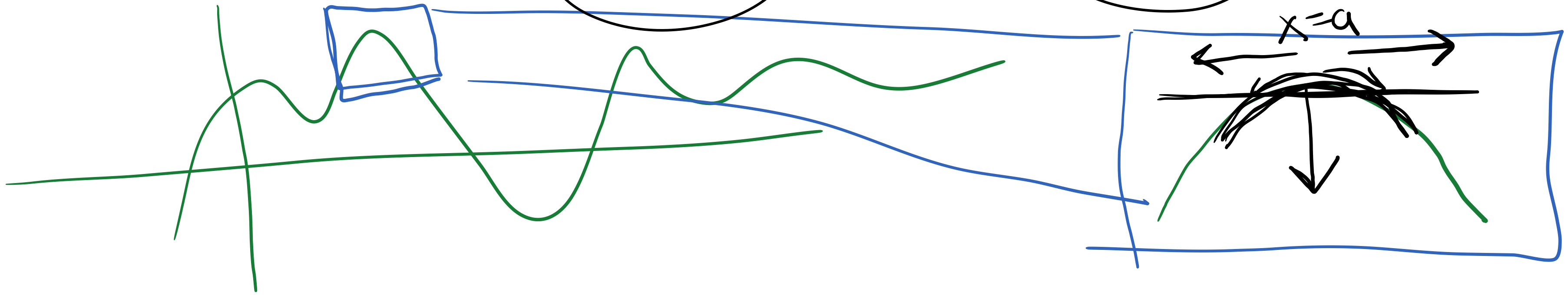
$f' = 0$

minimize a loss



Taylor Series Expansion

$$f(x) \approx f(x=a) + (x-a) \cdot \left.\frac{df}{dx}\right|_{x=a} + \frac{1}{2}(x-a)^2 \left.\frac{d^2f}{dx^2}\right|_{x=a} + \cdots$$



Derivatives are Linear Operators

$$f(x, \theta)$$

$$\underbrace{f(x_1, \theta) + f(x_2, \theta) + f(x_3, \theta) + \cdots + f(x_N, \theta)}_{\sum_{n=1}^{N} \left\{ f(x_n, \theta) \right\}}$$

$$\frac{d}{d\theta}\left(\sum_{n=1}^{N}\left\{f(x_n, \theta)\right\}\right) = \sum_{n=1}^{N}\left\{\frac{d}{d\theta}\left(f(x_n, \theta)\right)\right\}$$

Vectors & Matrices

$$\underline{x} = [x_1, x_2, \dots, x_n, \dots, x_N] \rightarrow \text{row vector} \rightarrow (\overset{\text{rows}}{1} \times \overset{\text{columns}}{N})$$

$\vec{x}$

\mathbf{x}

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ \vdots \\ x_N \end{bmatrix} \rightarrow \text{column vector} \rightarrow (N \times 1)$$

2 elements: x_1 & x_2

$$\underline{x} = [x_1, x_2]^T \quad \text{transpose}$$

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow (2 \times 1)$$

Matrix: $\underline{\underline{X}} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{23} \end{bmatrix}$ 3rows x 2cols

\mathbf{X}

$$\underline{\underline{X}}^T = ?$$

Matrix Multiplication

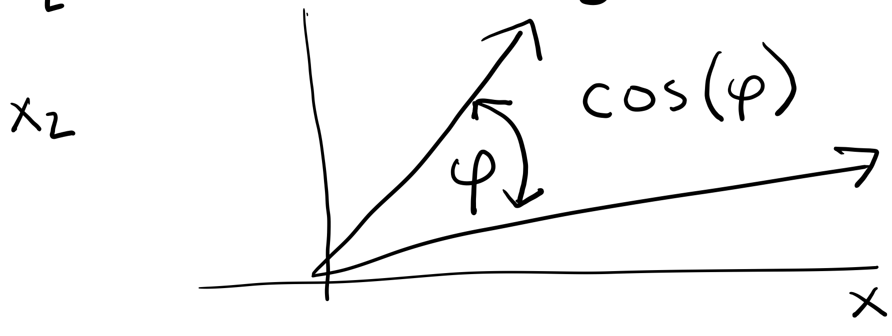
$$\underline{x}^T \underline{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \cdot x_1 + x_2 \cdot x_2 = x_1^2 + x_2^2$$

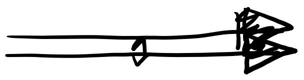
$$\underline{x} = [x_1 \;\; x_2]^T$$

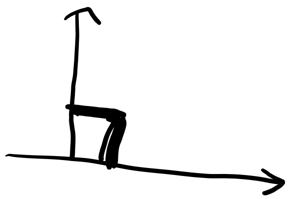
(2×1)

$$\boxed{\underline{x}^T \underline{x} = \sum_{d=1}^{D} \{ x_d^2 \}}$$ Inner Product

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} ; \; \underline{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \rightarrow \underline{x}^T \underline{y} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = x_1 y_1 + x_2 y_2$$







$\underline{x} = [x_1 \quad x_2]^T \quad (2 \times 1)$

$$\underline{x}\,\underline{x}^T = \begin{bmatrix} \boxed{x_1} \\ \boxed{x_2} \end{bmatrix} \begin{bmatrix} \boxed{x_1} & x_2 \end{bmatrix} = \begin{bmatrix} x_1 \cdot x_1 & x_1 x_2 \\ x_2 x_1 & x_2 x_2 \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1 x_2 \\ x_2 x_1 & x_2^2 \end{bmatrix}$$

$$(2 \times 1)\quad(1 \times 2) = 2 \times 2$$

OUTER Product

Vectors

~~Inner~~ $\underline{x}^T \underline{x} \longrightarrow (1)$ scalar

outer

$\underline{x}\,\underline{x}^T \longrightarrow$ Matrix

vector

$\underline{\underline{X}}$ $\underline{b}$

$(N \times D)$ $(D \times 1)$ $\rightarrow$ $\boxed{N \times 1}$

$\underline{\underline{A}}\,\underline{x} = \underline{b}$ $\rightarrow$ solve $\underline{x}$ $\rightarrow$ $\boxed{\underline{x} = \underline{\underline{A}}^{-1}\underline{b}}$

$(N \times D)(D \times 1) = (N \times 1)$

$(N \times 1)$

$a x = b$

$x = b/a$

functions of multiple variables

$$f(\underline{x}) = x_1^a + x_2^b$$

partial derivatives $\Rightarrow$ f can change as x_1 changes <u>AND</u> as x_2 changes

\partial

$$\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_1}\left(x_1^a + \cancel{x_2^b}\right) = a x_1^{a-1}$$

$$\frac{\partial f}{\partial x_2} = \frac{\partial}{\partial x_2}\left(\cancel{x_1^a} + x_2^b\right) = b x_2^{b-1}$$

<u>gradient vector</u> : $\underline{g} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right]^T = \left[f'_{x_1}, f'_{x_2}\right]^T$

\mathbf{g}

$$f(\underline{x}) = x_1^a + x_2^b + c \cdot x_1 x_2$$

$$\frac{\partial f}{\partial x_1}\left(x_1^a + \cancel{x_2^b} + c\,\underline{\underline{x_1 x_2}}\right) = a x_1^{a-1} + c\,\underline{x_2}$$

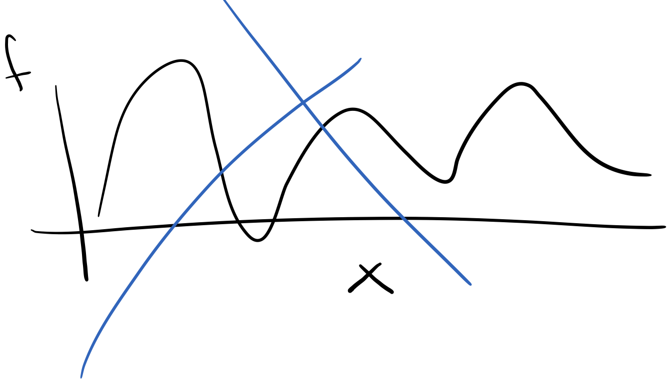
$$\frac{\partial f}{\partial x_2}\left(\cancel{x_1^a} + x_2^b + c\,\underline{\underline{x_1 x_2}}\right) = b x_2^{b-1} + c\,\underline{x_1}$$

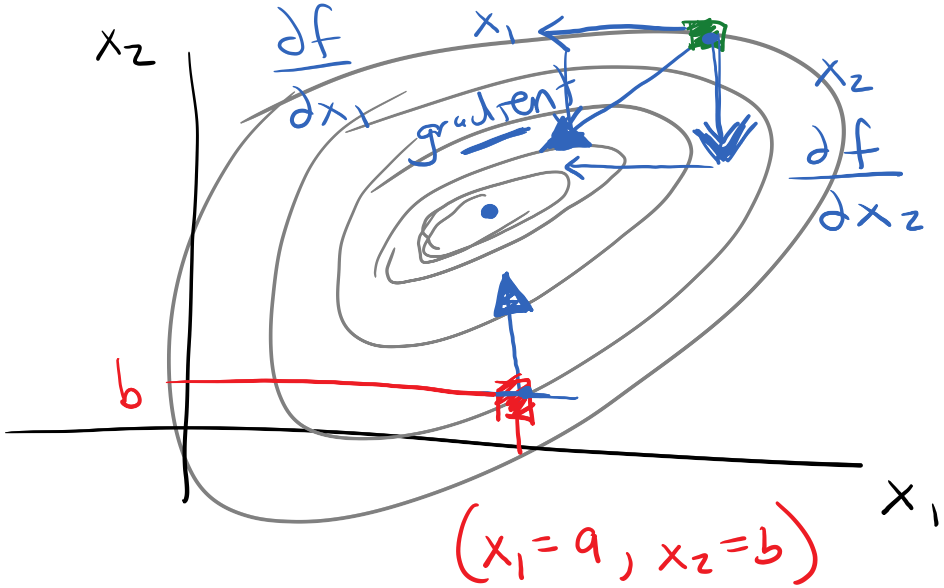
---

cross-derivatives:

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right) = \frac{\partial}{\partial x_2}\left(\cancelto{0}{a x_1^{a-1}} + c x_2\right) = c$$

$$\frac{\partial^2 f}{\partial x_2 \partial x_1} = ?$$

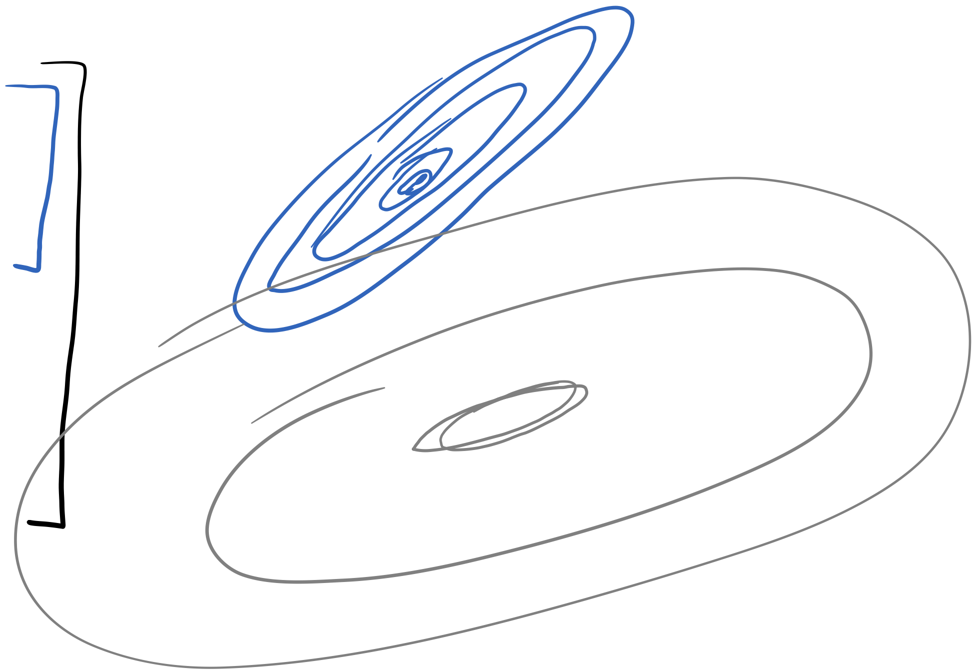




Hessian Matrix

2nd derivates
Matrix

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$



Multivariate Taylor Series

$$f(\underline{x}) \approx f(\underline{x}=\underline{a}) + (\underline{x}-\underline{a})^T \underline{g}\Big|_{\underline{x}=\underline{a}} + \frac{1}{2}(\underline{x}-\underline{a})^T \underline{\underline{H}}\Big|_{\underline{x}=\underline{a}} (\underline{x}-\underline{a}) + \cdots$$

$$\underline{x} = [x_1 \quad x_2]^T$$

$$\underline{a} = [a_1 \quad a_2]^T$$

$$(\underline{x}-\underline{a})^T(\underline{x}-\underline{a}) = \sum_{d=1}^{D} \left\{ (x_d - a_d)^2 \right\}$$