Artificial Intelligence/Machine Learning/Deep Learning: 'Bridging the Skills Gap'

Optional: Calculus Refresher

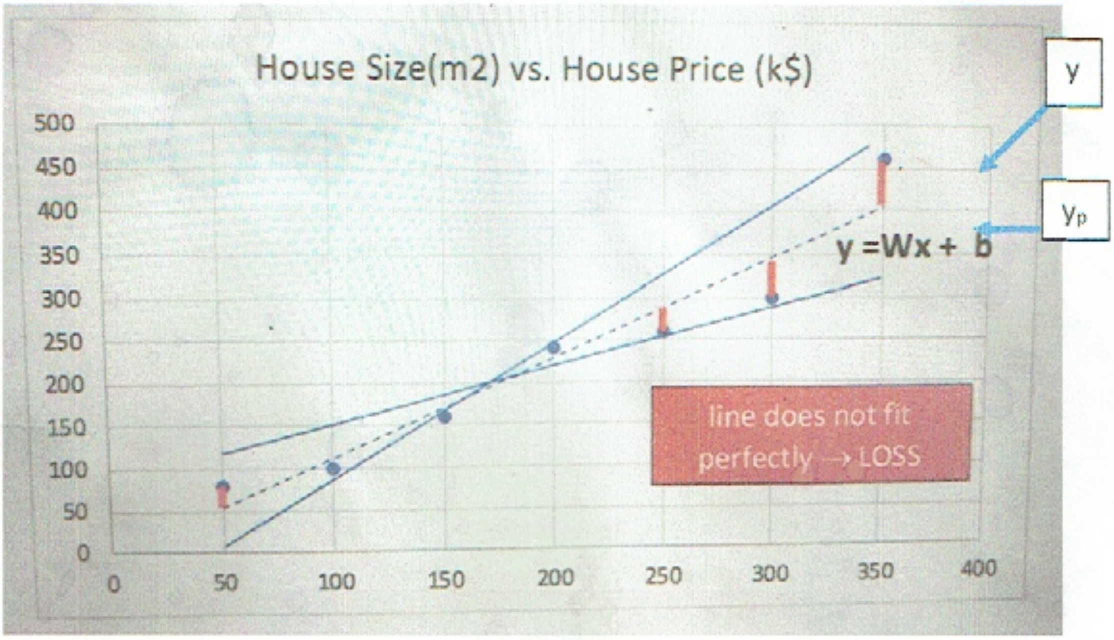
Calculus is very important if you want to understand ML. Not only calculus with respect to functions but also with respect to vectors and matrices (matrix calculus).

Topics of this session include:

1. Derivatives
2. Partial Derivatives
3. Gradient/Hessian
4. Minimizing a function
5. Minimizing a function with constraints
6. Taylor Series

The chain rule will not be discussed at this point. We will have a dedicated session once we look at Deep Learning.

In lesson 1 we discussed a linear regression problem: *housing size vs. housing price*.



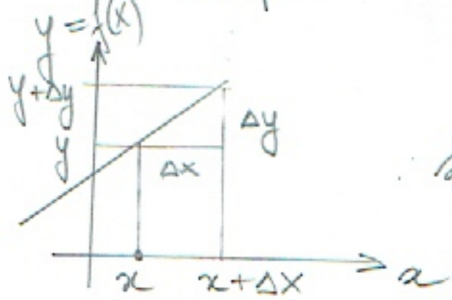
In order to measure the performance of the model we introduced a **cost function C** to reflect the error ($y-y_p$) of the model. We used the MSE formula:

C(w,b) = $\frac{1}{6}\sum_{i=1}^{6}(y_p^i - y^i)$

ML is all about selecting the model (line in this case) that minimizes the error → **minimization problem → derivative problem**

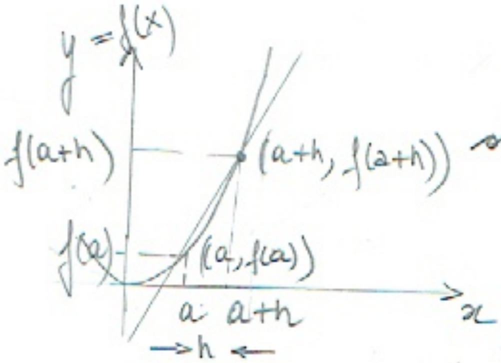
Derivatives

↳ rate of change of a function
slope



$$\text{slope} = \frac{\Delta y}{\Delta x}$$

↳ constant for line



slope changes

$$\boxed{\text{slope} = \frac{\Delta y}{\Delta x} = \lim_{h \to 0} \frac{f(a+h) - f(a)}{\cancel{a}+h-\cancel{a}}}$$

↳ derivative in point $(a, f(a))$

$\longrightarrow \frac{dy}{dx}$ instantenious rate of change at a given point

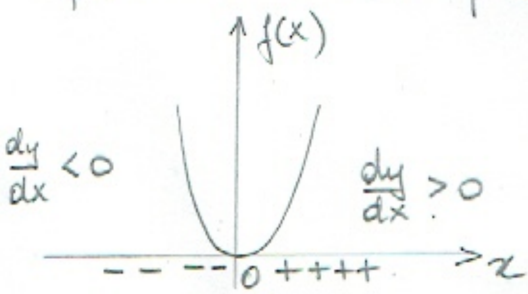
Minimum of maximum of a function $y = f(x)$

example : $f(x) = x^2$

$$\frac{dy}{dx} = f'(x) = 2x$$

1. figure out where $f'(x) = 0 \Rightarrow 2x = 0 \Rightarrow x = 0$

↓
MINIMUM !

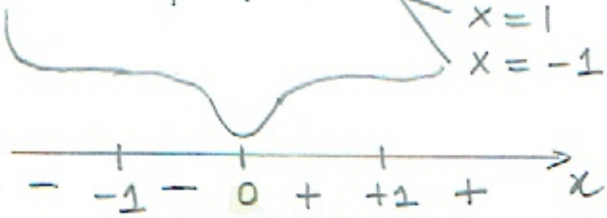


2. look at sign $f'(x)$

<u>example</u>: $f(x) = (x^2-1)^3$

$$f'(x) = 3(x^2-1)^2 2x = 6x\underbrace{(x^2-1)^2}_{\text{Allways }+}$$

$$f'(x) = 0 \quad \begin{cases} x=0 \\ x=1 \\ x=-1 \end{cases}$$



$f'(x)<0$ $f'(x)<0$ ↑ MINIMUM $f'(x)>0$ $f'(x)>0$

<u>partial derivatives</u> $z = f(x,y) = x^2+y^2-2xy+1$

$$\frac{\partial z}{\partial x} = 2x-2y \qquad \frac{\partial z}{\partial y} = 2y-2x$$

$$\nabla z = \begin{pmatrix} \frac{\partial z}{\partial x} \\ \frac{\partial z}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x-2y \\ 2y-2x \end{pmatrix}$$

GRADIENT VECTOR

↳ $\nabla z = 0$ → stationary point

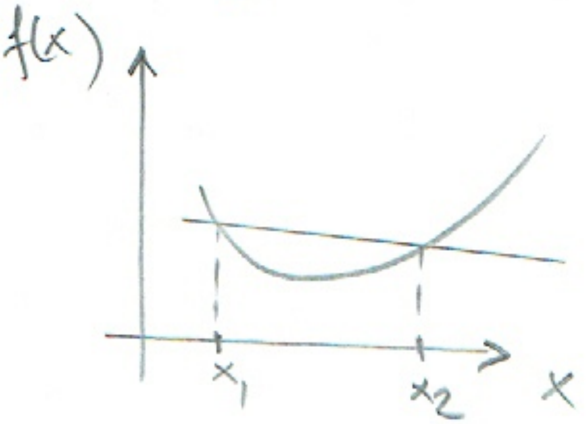
$$H = \begin{pmatrix} \frac{\partial^2 z}{\partial x^2} & \frac{\partial^2 z}{\partial x \partial y} \\ \frac{\partial^2 z}{\partial y \partial x} & \frac{\partial^2 z}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 2 & -2 \\ -2 & +2 \end{pmatrix}$$

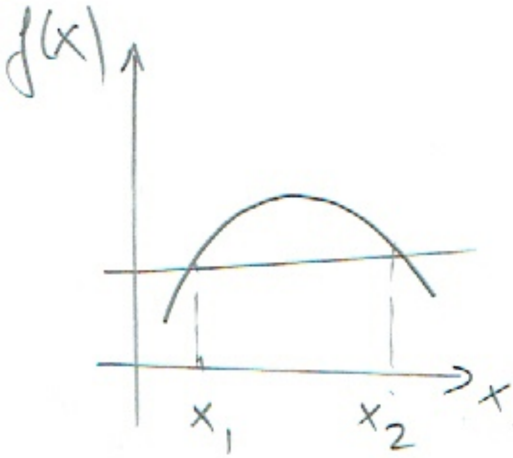
HESSIAN

→ SQUARED MATRIX of 2nd order partial derivatives of $z(x,y)$

↳ Tells us something about curvature

convexity





convex in $[x_1, x_2]$ if $f(x)$ is below the straight line.

↓

MINIMUM = GLOBAL MINIMUM!

example: MSE, logistics Regression
SVMs

non-convex in $[x_1, x_2]$ if $f(x)$ is above the straight line

↓

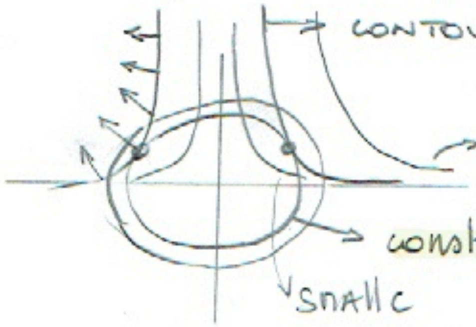
you can have several local minima

example: Neural Nets

Lagrange Multipliers

↳ MINIMIZATION / MAXIMIZATION problem WITH constraints

<u>example</u> : $f(x,y) = x^2 y$ AND $x^2+y^2=1$



→ CONTOUR line of f → $f(x,y) = c$

→ higher c

→ CONSTRAINT → $g(x,y) = x^2+y^2$

↓ CONTOUR lines are circles

→ IN OUR case only $x^2+y^2=1$ COUNTS!

small c

GOAL
→ choose c so that $f(x,y)$ touches constraint.

↱ proportionality constant

$$\boxed{\nabla f(x_m, y_m) = \lambda \nabla g(x_m, y_m)}$$

↳ Lagrange multipliers

$$\begin{cases} \nabla g = \nabla(x^2+y^2) = \begin{pmatrix} 2x \\ 2y \end{pmatrix} \\ \nabla f = \nabla(x^2 y) = \begin{pmatrix} 2xy \\ x^2 \end{pmatrix} \end{cases} \Rightarrow \begin{pmatrix} 2x \\ 2y \end{pmatrix} = \begin{pmatrix} 2xy \\ x^2 \end{pmatrix} \lambda$$

$$\rightarrow \begin{cases} \cancel{2x}y = \lambda \cancel{2x} \rightarrow y = \lambda \\ x^2 = \lambda 2y \\ x^2+y^2=1 \end{cases}$$

3 UNKNOWNS & <u>only</u> 2 equations

↓ CONSTRAINT IS 3rd equation

$$\begin{cases} y = \lambda \\ x^2 = 2y^2 \\ x^2+y^2=1 \end{cases}$$

$$2y^2+y^2 = 1 \rightarrow 3y^2=1 \rightarrow y = \pm\sqrt{1/3}$$

$$x^2 = 2y^2 \rightarrow x^2 = \frac{2}{3} \rightarrow x = \pm\sqrt{\frac{2}{3}}$$

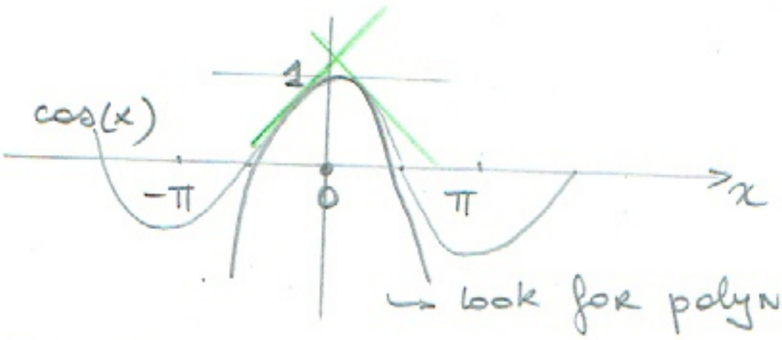
Taylor Series

↳ will be used to derive update rule for Gradient Descent

↳ TOOL to Approximate functions

↳ takes non-polynomial function and finds polynomials that Approximates it near some point

polynomials ↓ easier to handle $\left(\frac{d}{dx}, \int, \dots\right)$



↳ look for polynomial approx near 0

how to approx cos(x) near 0 with polynomial ?

(1) $\boxed{P(x) = c_0 + c_1 x + c_2 x^2}$ → find c_0, c_1, c_2 so that $P(x)$ resembles $\cos(x)$ near $x=0$ → $\cos(x) = 1$

① $(1) \overset{x=0}{\Longrightarrow} P(0) \overset{\text{must be}}{=} 1 \Rightarrow c_0 = 1$ ✓

② now tangent line in $P(0)$ → $\frac{d(\cos)}{dx}(0) = -\sin(0) = 0$

(1) $\frac{dP(x)}{dx} = c_1 + 2c_2 x \xrightarrow{x=0} c_1 \xrightarrow{\text{set to}} 0$ ✓

} First derivatives match!

③ $\frac{d^2(\cos)}{dx^2}(0) = -\cos(0) = -1$ → making sure that 2nd derivatives match makes sure that they curve at same rate !

$\frac{d^2P(x)}{dx^2} = 2c_2 \Rightarrow 2c_2 \overset{\text{must}}{=} -1$

$\Downarrow$

$c_2 = -1/2$

→ polynomial that matches $\cos(x)$ near $x=0$

is $\boxed{P(x) = 1 - \frac{1}{2}x^2}$ (2) → BEST Quadratic APPROX.

$\cos(0.1) = 0.995$
$P(0.1) = 0.995$ → (2) is really good Approx for $\cos(x)$

you can add c_3x^3 → $P(x) = \underbrace{1 - \frac{1}{2}x^2} + \underbrace{c_3x^3}_{1.2.3\,c_3}$

$$\frac{d^3P(x)}{dx^3} = 1.2.3\, c_3\, x^0$$

now $\frac{d^3\cos}{dx^3}(0) = \sin(0) = 0$

must match $\Rightarrow c_3 \overset{\text{must}}{=} 0$

→ $\boxed{P(x) = 1 - \frac{1}{2}x^2}$ → BEST Qubic APPROX

you can add c_4x^4

now $\frac{d^4\cos(0)}{dx^4} = \cos(0) = 1$.

$$\frac{d^4p(x)}{dx^4} = 1.2.3.4\, c_4 = 24\, c_4$$

must mach $24\, c_4 = 1$

→ $c_4 = \frac{1}{24}$

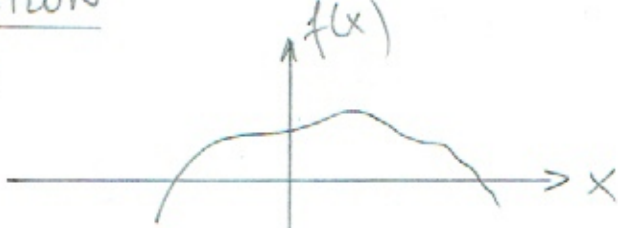
→ $P(x) = 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4$

$$\boxed{P(x) = 1 + 0\frac{x^1}{1!} - 1\frac{x^2}{2!} + 0\frac{x^3}{3!} + 1\cdot\frac{x^4}{4!} + \cdots}$$

Polynomial approximation of cos(x) in the neighborhood of $x=0$

Generalization

for $x=0$

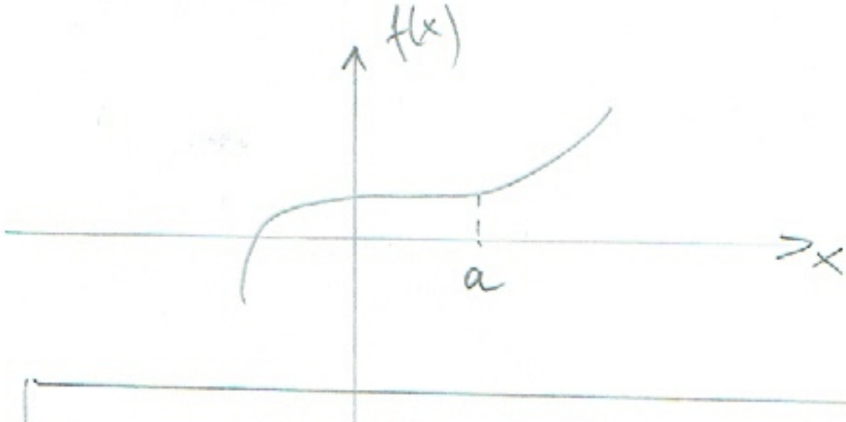


calculate $f'(x), f''(x), f'''(x), \ldots$

$$\rightarrow P(x) = f(0) + f'(0)\,x + f''(0)\frac{x^2}{2!} + f'''(0)\frac{x^3}{3!} + \cdots$$

for $a=0$

Generalization for $a=x$

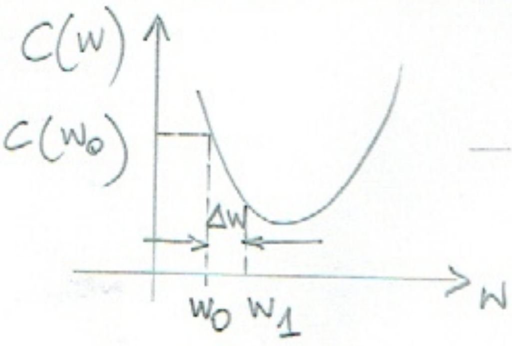


$$\rightarrow \boxed{P(x) = f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2!} + f'''(a)\frac{(x-a)^3}{3!} + \cdots}$$

Applied to machine learning $\rightarrow x = w_0 + \Delta w = w_1$

$a = w_0$

$$\boxed{C(w_0+\Delta w) \approx C(w_0) + \Delta w \cdot \nabla C(w_0)}$$



will be used to determine Gradient Descent for multivariate regression