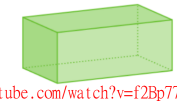
🧀 02 Finding Local Extrema

Let $f : [-1, 2] \to \mathbb{R}, x \mapsto \exp(x^3 - 2x^2)$.

1. Compute $f'(x)$.
2. Plot f and f' (you can use any graphing tool or software).
3. Find all possible candidates x^* for maxima and minima. *Hint:* $\exp$ *is a strictly monotone function.*
4. Compute $f''(x)$.
5. Determine if the candidates are local maxima, minima or neither.
6. Find the global maximum and global minimum of f on $[-1, 2]$.

Խնդիր 5.9 Ունենք 24 մ^2 մակերեսով ստվարաթուղթ (ինչպես նաև մկրատ ու սոսինձ), որով ցանկանում ենք պատրաստել այսպիսի տուփ՝



https://www.youtube.com/watch?v=f2Bp77tiESg

որի ձախ և աջ նիստերը քառակուսիներ են:

Ամենաշատը որքա՞ն կարող է լինել այդ տուփի ծավալը:

Հուշում. *Նախ՝ մի քիչ երկրաչափություն: Վերցրեք ոչ քառակուսի նիստերից մեկը, նշանակեք դրա երկարությունն ու լայնությունը x և y: Կարո՞ղ եք y-ն արտահայտել x-ով: Իսկ ծավալը՝ x-ո՞վ:*

1) $\left(e^{x^3-2x^2}\right)' =$

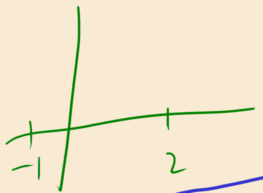
$= e^{x^3-2x^2} \cdot 3x^2 - 4x =$

$= 3e^{x^3-2x^2}\, x\left(x - \frac{4}{3}\right)$

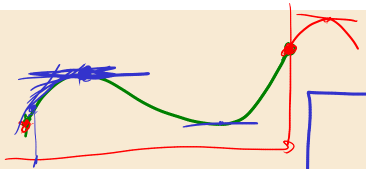
x^3

$f'(x) = 0$

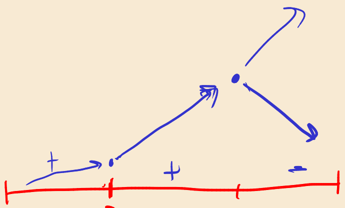
f''



-1 $\quad 2$

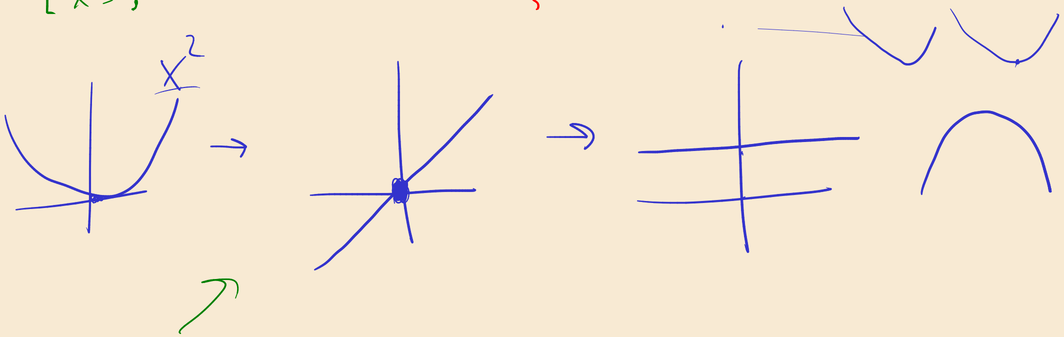


2. $\left[\begin{array}{l} x = -1 \\ x = 2 \\ x = 0 \\ x = \frac{4}{3} \end{array}\right.$



$+ \quad + \quad -$

$0 \quad \frac{4}{3}$



x^2

f'

f_{x_0}

$$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

Hesh. $\begin{bmatrix} f_{xx} & f_{xy} \\ f & f \end{bmatrix}$

03 Convex Function Properties

Consider two convex functions $f, g : \mathbb{R} \to \mathbb{R}$.

1. Show that $f + g$ is convex.

2. Now, assume that g is additionally non-decreasing, i.e., $g(y) \geq g(x)$ for all $x \in \mathbb{R}$, for all $y \in \mathbb{R}$ with $y > x$. Show that $g \circ f$ is convex. ✓

f, g

$h(x) = g(f(x)) = g(f(x))$

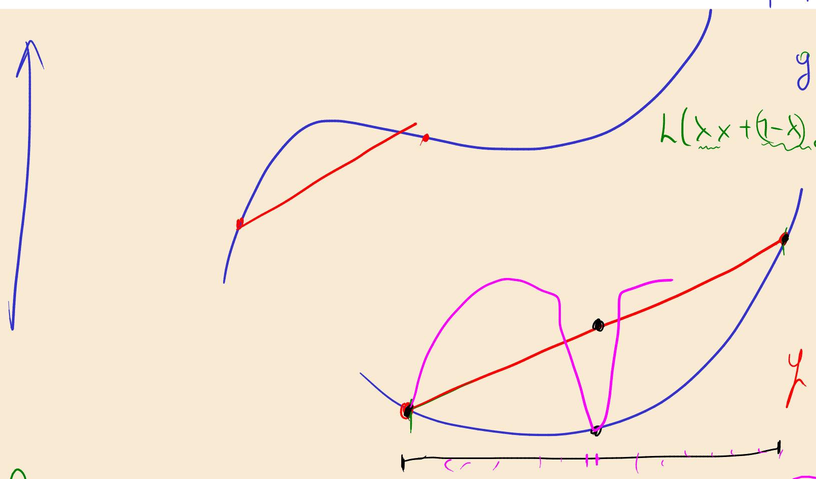
$\lambda h(x) + (1-\lambda) h(y) \geq h(\lambda x + (1-\lambda) y)$

$g(f(\lambda x + (1-\lambda) y)) \leq g(\lambda f(x) + (1-\lambda) f(y))$

$g(\lambda f(x) + (1-\lambda) f(y)) \leq \lambda g(f(x)) + (1-\lambda) g(f(y))$

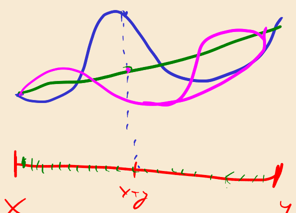
$h(\lambda x + (1-\lambda) y)$ $\quad g(f(\lambda x + (1-\lambda) y))$ $\quad \lambda h(x) + (1-\lambda) h(y)$

$h(\lambda x + (1-\lambda) y) \leq$



(~~convex~~)

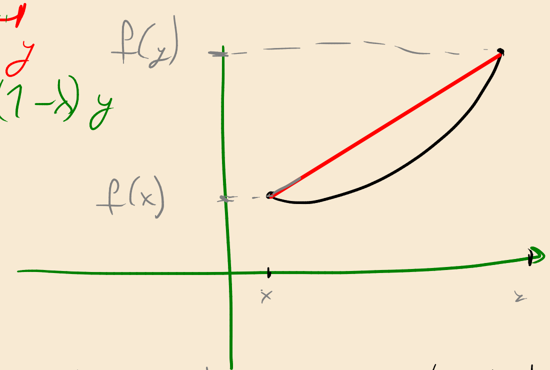
(concave)



f, x, y

$$\frac{f(x) + f(y)}{2} \geq f\left(\frac{x+y}{2}\right)$$

x, y, $\lambda x + (1-\lambda) y$



$\lambda x + (1-\lambda) y$ $\qquad \lambda \in [0,1]$

$\lambda f(x) + (1-\lambda) f(y) \geq f(\lambda x + (1-\lambda) y)$

$\lambda = \frac{1}{2}$

$\boxed{x + (y-x) c}$

$c \in [0,1]$

$x + cy - xc = x(1-c) + cy$

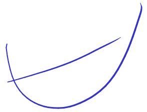
$c = 1 - \lambda$

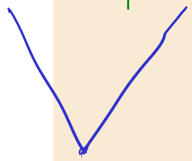
$f(x) + f(y) ...$

🧀 🧀 04 Testing Convexity in ML Functions

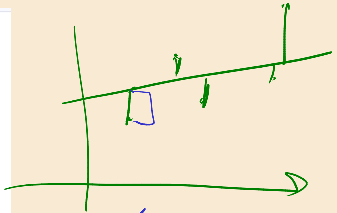
Determine whether the following ML-related functions are convex, concave, or neither on the given intervals:

1. **Mean Squared Error**: $L(w) = \frac{1}{2}(w-3)^2$ on $\mathbb{R}$
2. **ReLU Activation**: $\mathrm{ReLU}(x) = \max(0, x)$ on $\mathbb{R}$
3. **Sigmoid Function**: $\sigma(x) = \frac{1}{1+e^{-x}}$ on $\mathbb{R}$

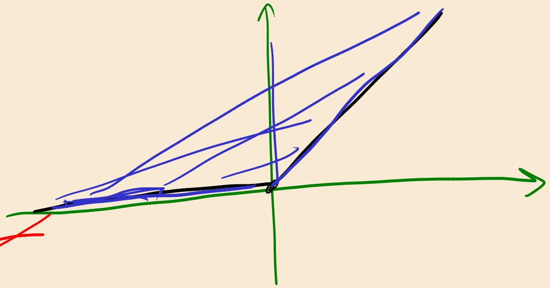


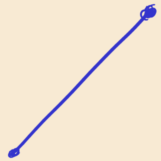


$|x|$

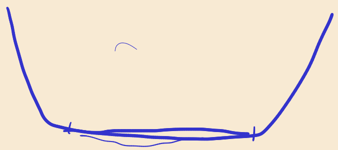


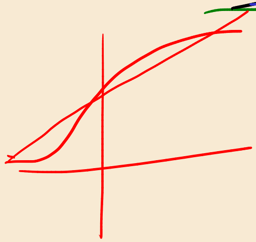
$f(\lambda x)+7\dots)$





$3x+509$





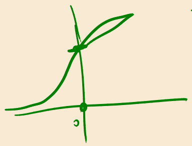
$$\sigma'(x) = \sigma(x) - \sigma^2(x)$$

$$\left(\sigma(x) - \sigma^2(x)\right)'_x = \sigma'(x) - 2\sigma(x)\sigma'(x) =$$

$$= \sigma'(x)\left(1 - 2\sigma(x)\right) = \qquad \sigma(x) \in [0, 1]$$

$$= \left(\sigma(x) - \sigma^2(x)\right)\left(1 - 2\sigma(x)\right) =$$

$$= \underset{+}{\sigma(x)}\,\underset{+}{\left(1 - \sigma(x)\right)}\left(1 - 2\sigma(x)\right)$$



$\sigma(x) > \frac{1}{2} \xrightarrow{(x>0)}$ concave

$\sigma(x) < \frac{1}{2} \xrightarrow{(x<0)}$ convex

🧀 🧀 05 L2-Regularized Linear Regression

Consider the L2-regularized mean squared error loss function:

$$R_\lambda(w) = \frac{1}{n}\sum_{i=1}^{n}(wx_i - y_i)^2 + \lambda w^2$$

where $\{(x_i, y_i)\}_{i=1}^n$ are training data points, w is the model parameter, and $\lambda > 0$ is the regularization parameter.

1. Find the optimum w^* and determine if it's a minimum or maximum.
2. Is the function $R_\lambda(w)$ convex? Justify your answer.
3. Is the minimizer unique? Explain why this is important for machine learning.

🧀 🧀 🧀 06 Logistic Loss and Its Properties

ⓘ **Context**

Consider the logistic loss function:

$$\ell(z;y) = -y\log\sigma(z) - (1-y)\log(1-\sigma(z))$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function, z is the logit (linear combination of features), and $y \in \{0,1\}$ is the true binary label.

1. **Task**: Show that $\frac{d}{dz}\ell(z;y) = \sigma(z) - y$.

2. **Bonus**: Check if the function $g(z) = (y - \sigma(z))^2$ is convex with respect to z.

$y = \{0, 1\}$ $\quad x \quad z, y \quad z = \hat{w}x \quad$ cross-entropy

$\left(f(x^{(i)}) - y^{(i)}\right)^2 \quad \sigma(x^{(i)}),\ y^{(i)}$

$\sigma(z) \cdot 1-\sigma(z) \quad y=0 \quad \sigma(z) \quad y=1 \quad \sigma(z)$

$y=1$: $\sigma(z)$; $y=0$: $1-\sigma(z)$

if $y=1 \to \sigma(z)$, $y=0 \to 1-\sigma(z)$ $\to \sigma(z)^y \cdot (1-\sigma(z))^{1-y}$

$\log a^x \Rightarrow x\log a$

$\log ab = \log a + \log b$

$-\left(y\log(\sigma(z)) + (1-y)\log(1-\sigma(z))\right)$

log-likeh. MLE

$(1-y)\ \sigma(z) \quad y=1 \quad -y(\sigma(z))\ \ 1-y(1-\sigma z)$

MLE $\quad P(x_1) \quad x_1 \quad x_2$

if $y=1$, $f = \sigma(z)$; else: if $y=0$, $f = 1-\sigma(z)$

$\sigma(z)^y \cdot (1-\sigma(z))^{1-y}$

$y=1\ \sigma(z) \cdot 1 \quad y=0\ 1 \quad 1-\sigma(z)$

$\log a \cdot b = \log a + \log b$

$\log a^x = x \cdot \log a$

$\log\left(\sigma(z)^y \cdot (1-\sigma(z))^{(1-y)}\right) =$

$= \log(\sigma(z)^y) + \log(1-\sigma(z))^{1-y} =$

$= y\log(\sigma(z)) + (1-y)\log(1-\sigma(z))$

Part 1: Derivative of logistic loss

First, recall that $\sigma(z) = \frac{1}{1+e^{-z}}$ and $\frac{d\sigma}{dz} = \sigma(z)(1-\sigma(z))$.

$\log(x) = \frac{1}{x}$

Let's compute the derivative term by term:

$$\frac{d}{dz}\ell(z;y) = \frac{d}{dz}\left[-y\log\sigma(z) - (1-y)\log(1-\sigma(z))\right]$$

e

$\log(\sigma(z))$

$\ln$ $\log$

For the first term:

$$\frac{d}{dz}[-y\log\sigma(z)] = -y\cdot\frac{1}{\sigma(z)}\cdot\frac{d\sigma}{dz} = -y\cdot\frac{1}{\cancel{\sigma(z)}}\cdot\cancel{\sigma(z)}(1-\sigma(z)) = -y(1-\sigma(z))$$

For the second term:

$\log$ $\boxed{10}$ 0.2 info

$$\frac{d}{dz}[-(1-y)\log(1-\sigma(z))] = -(1-y)\cdot\frac{1}{1-\sigma(z)}\cdot\frac{d}{dz}[1-\sigma(z)]$$

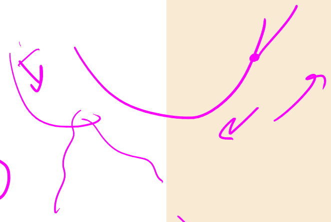
$$= -(1-y)\cdot\frac{1}{\cancel{1-\sigma(z)}}\cdot(-\sigma(z)\cancel{(1-\sigma(z))}) = (1-y)\sigma(z)$$

12 — 12.5 — 1 . . . 13

Combining both terms:

residual

$$\frac{d}{dz}\ell(z;y) = -y(1-\sigma(z)) + (1-y)\sigma(z)$$
$$= -y + y\sigma(z) + \sigma(z) - y\sigma(z) = \sigma(z) - y$$



$y=1$ 0.5 1 $-(-0.5)$

$\sigma(z)(1-\sigma(z))$

$\sigma(z) \in [0,1]$

$(\sigma(z) - y)^2$

y ...

🧀🧀🧀 07 Lipschitz Continuity & Gradient Clipping

Context: Lipschitz Continuity

A function $f : \mathbb{R} \to \mathbb{R}$ is called **L-Lipschitz continuous** if there exists a constant $L \geq 0$ such that:

$$|f(x) - f(y)| \leq L|x - y|$$

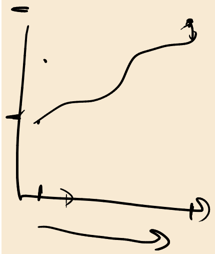
$$\frac{|f(x)-f(y)|}{|x-y|} \leq L$$

for all x, y in the domain. The smallest such constant L is called the **Lipschitz constant**.

This property is crucial in deep learning for gradient clipping, ensuring gradients don't explode during training.

Consider the sigmoid function $\sigma(z) = \frac{1}{1+e^{-z}}$.

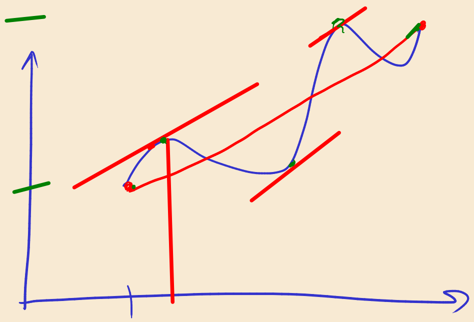
Task: Prove that $\sigma(z)$ is L-Lipschitz continuous and find the **optimal** (smallest possible) Lipschitz constant L.



10 →

10·L

t.



$f(y)$

$f(x)$

$$\frac{|f(y)-f(x)|}{|y-x|} = f'(c).$$

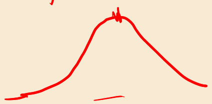
$f(y) - f(x) < f'(c)_{y-x}$

$\frac{1}{4}$

$\exists\, c \quad f'(c)$

Don't forget mean value theorem (Լա Գրանժ)

$f''(z) = \sigma(z)(1-\sigma(z))(1-2\sigma(z))$



$\sigma(z) = \frac{1}{2}$

$x(1-x)$

$x = \frac{1}{2}$

$f'(z)$

$\underset{\frac{1}{2}}{f(z)}\,(1-\underset{\frac{1}{2}}{f(z)})$

$\frac{1}{4}$

🧀 🧀 08 Taylor Series Expansions

Find the Taylor series expansion around the given point for each function:

1. $f(x) = e^x$ around $x = 0$ (Maclaurin series)
2. $g(x) = \ln(x)$ around $x = 1$
3. $h(x) = \cos(x)$ around $x = 0$ (first 4 non-zero terms)

$e^{x})$