Vilnius universitetas

Matematikos ir informatikos fakultetas

Duomenų mokslo ir skaitmeninių technologijų institutas

Matematiniai dalykai

prof. dr. Olga Kurasova

Olga.Kurasova@mif.vu.lt

Skaliaras, vektorius, matrica

- **Skaliaras** – skaičius, įprastai priklausantis realiųjų skaičių erdvei, $x \in \mathbb{R}$.
- **Vektorius** – tai skaičių rinkinys, $X \in \mathbb{R}^n$, kurio i-tasis elementas žymimas x_i:

$$X = [x_1, x_2, \dots, x_n].$$

- **Matrica** – tai dvimatis masyvas $A \in \mathbb{R}^{m \times n}$, kurio i-tosios eilutės ir j-tojo stulpelio elementas žymimas a_{ij}:

$$A = \begin{bmatrix} a_{11} & a_{22} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

- Įprastai kintamieji **skaičiai** žymimi pasvirusia mažąja raide ($x, y, z, a, \dots$); **vektoriai** ir **matricos** žymimi arba didžiąja raide ($X, Y, Z, A, \dots$), arba pastorinta mažąja ($\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{a}, \dots$)).

Transponuota matrica

- Sukeitus matricos eilutes ir stulpelius, gaunama **transponuota matrica**, kuri žymima A^{T}:

$$A = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

Tenzoriai

- Vektoriai apibendrina skaliarus, matricos apibendrina vektorius, o **tenzoriai** apibendrina matricas.
- **Tenzoriai** – tai daugiamačiai masyvai (matricos).
- Pavyzdyje pateiktas trijų dimensijų tenzorius:

$$T = \begin{bmatrix} [1\ 3] & [2\ 5] & [1\ 6] \\ [3\ 5] & [7\ 8] & [0\ 3] \end{bmatrix}.$$

- Vektoriai yra **pirmos eilės** tenzoriai, matricos – **antros eilės** tenzoriai.

Skaliarinė sandauga

- Dviejų vektorių $X, Y \in \mathbb{R}^n$ **skaliarinė sandauga** (*dot product*) apskaičiuojama taip:

$$X^{\mathrm{T}}\,Y = \sum_{i=1}^{n} x_i y_i\,,$$

čia

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X^{\mathrm{T}} = [x_1, x_2, \dots, x_n].$$

Išvestinės

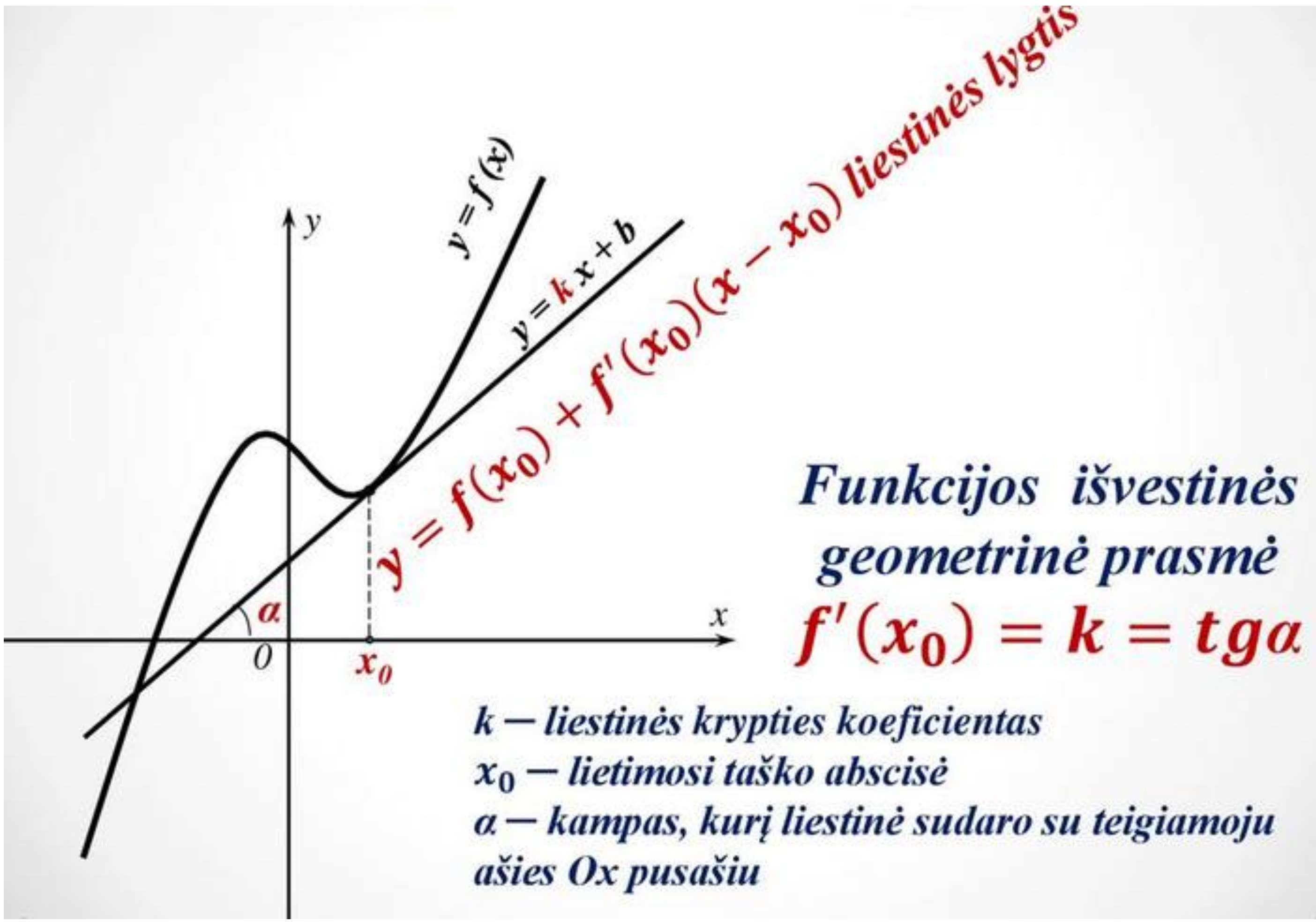
- Tarkime turime **funkciją** $f\colon \mathbb{R} \to \mathbb{R}$.
- Jos **išvestinė** apibrėžiama taip:

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

- Jei tokia riba egzistuoja, t. y. egzistuoja $f'(a)$, sakoma, kad funkcija f yra **diferencijuojama** taške a.
- Jei turime $y = f(x)$, šie reiškiniai yra **ekvivalentiški**:

$$f'(x) = y' = \frac{dy}{dx} = \frac{df}{dx} = \frac{d}{dx} f(x).$$

Išvestinės geometrinė prasmė



https://slideplayer.lt/slide/17056270/

Dalinės išvestinės

- Tarkime turime n kintamųjų **funkciją** $y = f(x_1, x_2, \dots, x_n)$.
- y **dalinė išvestinė** nuo x_i užrašoma:

$$\frac{\partial y}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{h}.$$

- Skaičiuojant $\frac{\partial y}{\partial x_i}$, imama $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ kaip konstantos.

Dalinės išvestinės (pavyzdys)

- Apskaičiuokime funkcijos $y = f(x_1, x_2) = x_1^2 + x_2$ **dalines išvestines**:
    - $\frac{\partial y}{\partial x_1} = 2x_1$ (x_2 laikoma **konstanta**, kurios išvestinė lygi 0);
    - $\frac{\partial y}{\partial x_2} = 1$ (x_1 laikoma **konstanta**, kurios išvestinė lygi 0).

Gradientas

- Tarkime turime n kintamųjų **funkciją** $f\colon \mathbb{R}^n \to \mathbb{R}$. Jos įvestis yra n-matis vektorius $X = (x_1, x_2, \dots, x_n)^{\mathrm{T}}$, o išvestis – skaliaras.
- **Gradientą** sudaro dalinės išvestinės:

$$\nabla f(X) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n}\right]^{\mathrm{T}}.$$

- **Antigradientas** yra priešingos krypties vektorius, $-\nabla f(X)$.

Gradientinis nusileidimas

1. Paimamas **pradinis taškas** x_0, $(t = 0)$.
2. Skaičiuojamas **naujas taškas**: $x_{t+1} = x_t - \eta \nabla f(x_t)$.
3. Jei nėra tenkinama algoritmo **sustojimo sąlyga**, atliekamas antras žingsnis.

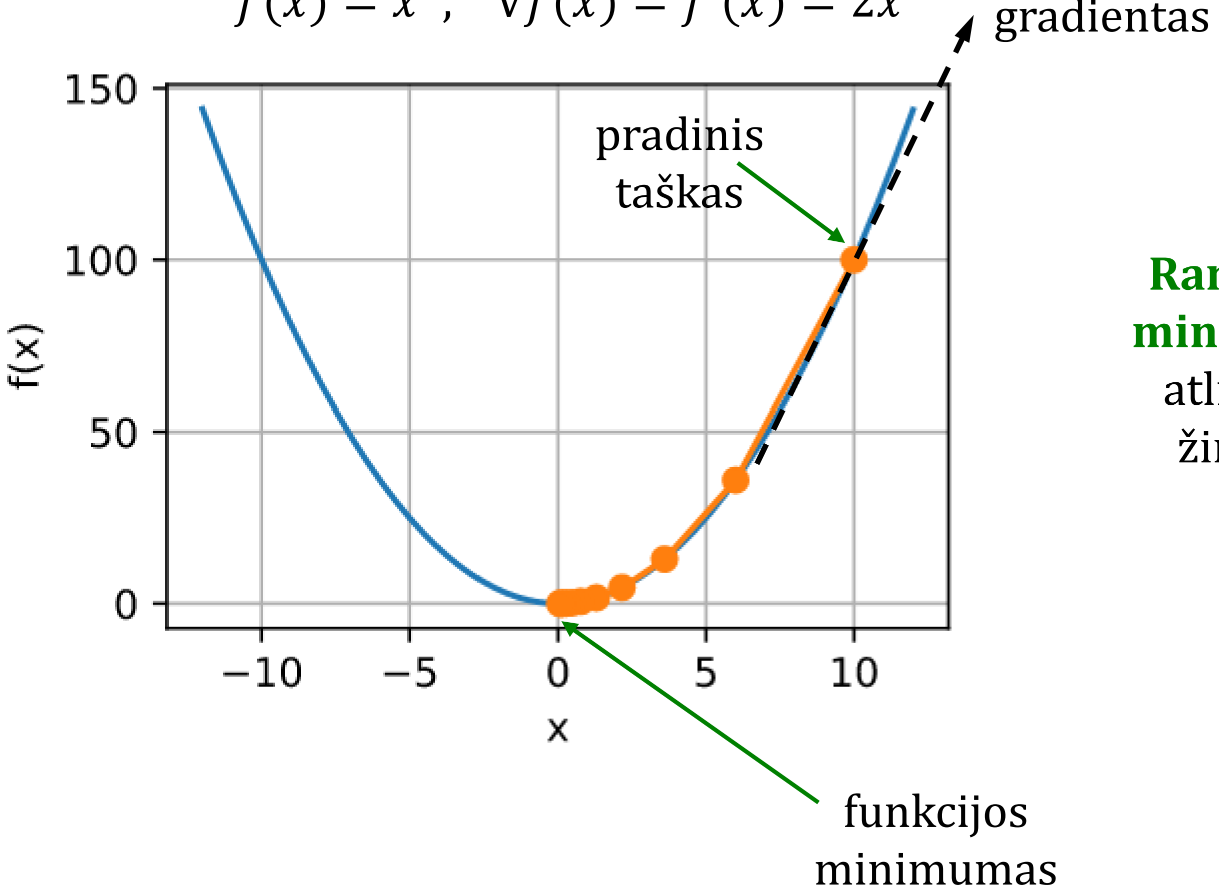
- Čia η yra parametras, įtakojantis optimizavimo žingsnį, dar vadinamas **mokymo greičiu** (*learning rate*).
- $-\nabla f(x_t)$ yra antigradientas. Optimizavimo metu „judama“ antigradiento kryptimis, t. y. funkcijos mažėjimo kryptimi.

Gradientinis nusileidimas

x_0 =10, $\quad$ $x_{t+1} = x_t - \eta \nabla f(x_t)$

$\boldsymbol{\eta = 0,2}$

$f(x) = x^2, \quad \nabla f(x) = f'(x) = 2x$



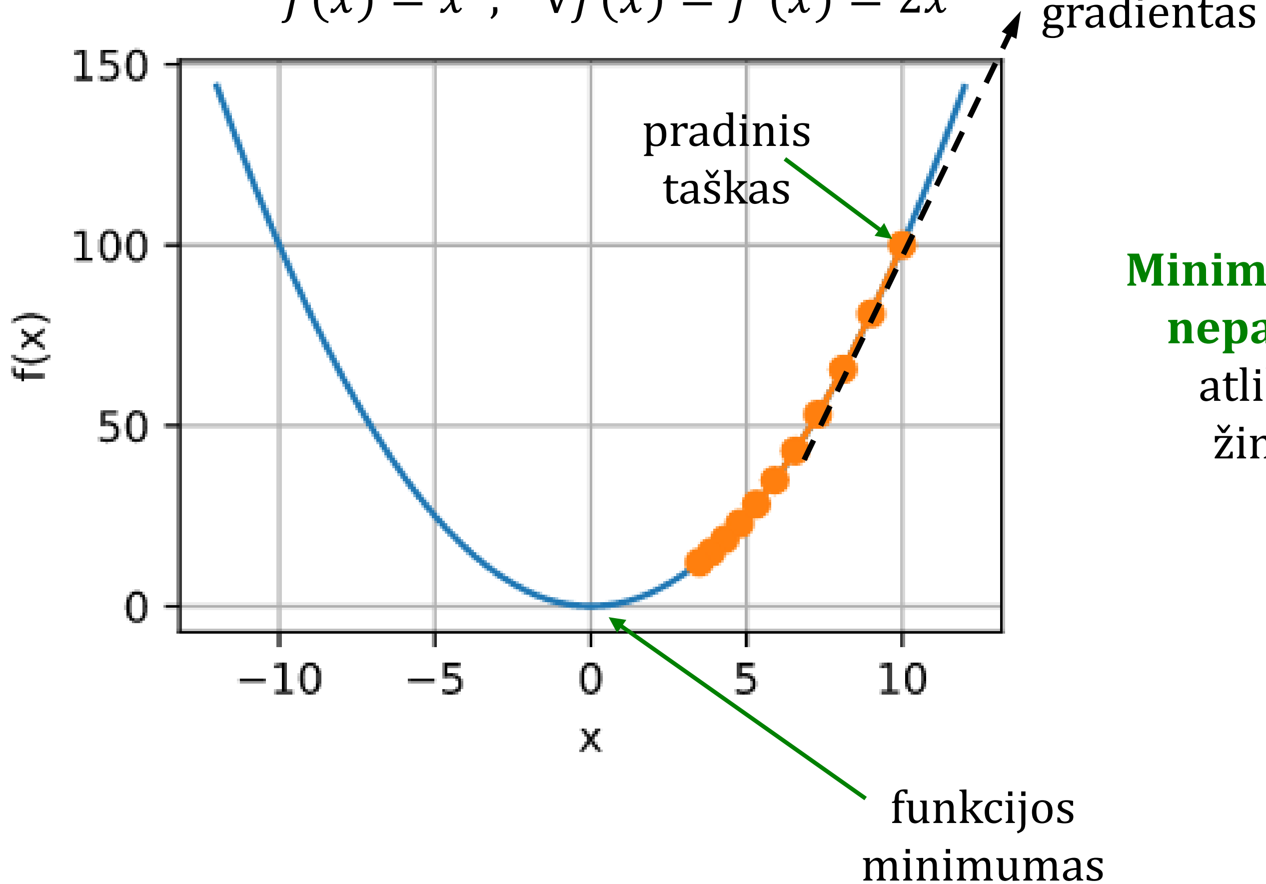
Randamas minimumas atlikus 10 žingsnių

Gradientinis nusileidimas

$x_0 = 10,$ $\qquad x_{t+1} = x_t - \eta \nabla f(x_t)$

$\boldsymbol{\eta = 0,05}$

$f(x) = x^2, \quad \nabla f(x) = f'(x) = 2x$



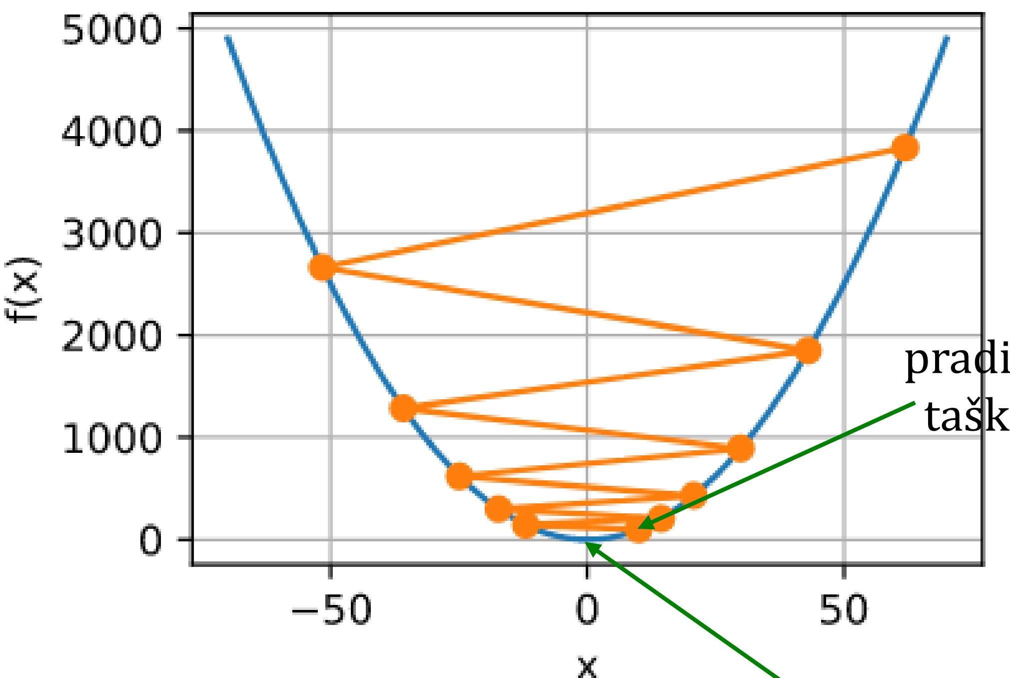
Minimumas dar nepasiektas atlikus 10 žingsnių

Gradientinis nusileidimas

$x_0 = 10,$ $\quad\quad x_{t+1} = x_t - \eta \nabla f(x_t)$

$\boldsymbol{\eta = 1,1}$

$$f(x) = x^2, \quad \nabla f(x) = f'(x) = 2x$$



pradinis taškas

funkcijos minimumas

Procesas diverguoja