



## Data Normalization

- scop: să transformă / ajusteze datele pt a fi la o scară similară
- utilizat în performanță pt ML (machine learning)

## Tehnici de normalizare

- clipping (tăiere)
  - log scaling (scalare jurnal)
  - scalare min-max
  - standardezare
- mean = medie
- std dev = deviația standard de la media unor date
- outlier = o val. fără diferență de celelalte din setul de date (poate afecta rezultatul)  
↳ efect semnificativ asupra mediei și a abaterii standard

## CLIPPING

= limitarea datelor care împiedică toate valenile corect. de a deacopla / de desvăluții unei anumite valori la o val. fixă

$$x_i^{\text{new}} = \begin{cases} x_i, & x_i < \text{val fixă} \\ \text{val fixă}, & x_i \geq \text{val fixă} \end{cases}$$

- se utilizează când există outliers extreme

# Objects - clasificare și detectie

## Regresia

- var. trănsă este continuă →
  - predicția este o val. numerică
- ex: estimare temperatură,  
precipitare venituri
- calitate predicție:
  - însumare dif (ridicate la patrat)
  - dintre val reală și cea preizată de afg.

## Clasificarea

- var. trănsă este dicționată →
  - predicția este o clasă / categorie
- ex: clasificare emulziuni,  
detectare tip obiect din img.
- calitate predicție:
  - nr. dif. dintre val reală  
și val preizată

## Algoritmul gradient descendente

- problemele de regresie

- identifică relația de dependență dintre date întrege

poate fi:  
 - liniară  
 - ne-liniară

Pentru pt a identifica relația de dependență:

- colectarea datelor
- explorarea datelor (plot)
- selecționează model regresie (ex: liniară, polinomială, ...)
- antrenare model
- evaluare model (performanță → metrii de evaluare)
- interpretare rezultate

- pe baza unor date se donează precizia unor valori  
associate cu acele date

- valoare preizăre pe un interval continuu

↪ var. dependență poate lua o gamă  
largă de valori dintr-un interval  
continuu.

↪ nu conține găuri

- datele se normaliză și primul atribută și output

$$x = (x_1, x_2, \dots, x_n)$$

- regresorul  
(model liniar  
de predicție)

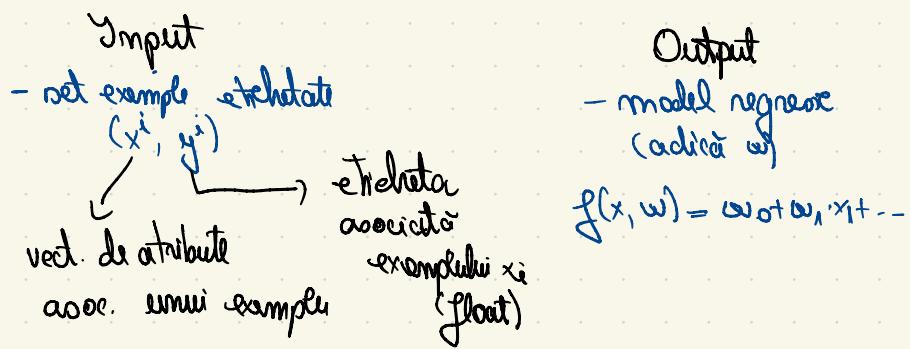
$$y = f(\mathbf{x}, w) = w_0 + w_1 \cdot x_1 + w_2 \cdot x_2 + \dots$$

# Metodologia rezolvării unui problema de regresie liniică

## • Antrenare

### Algorithm

- Least Square
- Gradient Descent



## • Testare

### Algorithm

- regresorul învățat pt val. output

$$y_{\text{new}} = f(x_{\text{new}}, w)$$

### Input

- exemplu ne-etichetat  $\mathbf{x}_{\text{new}}$   $\rightarrow$  vector de atribută

### Output

- val. predictă pt  $\mathbf{x}_{\text{new}}$

# metoda de gradient decrescător