Základy matematiky a statistiky pro humanitní obory II

Vojtěch Kovář

Fakulta informatiky, Masarykova univerzita
Botanická 68a, 602 00 Brno, Czech Republic
xkovar3@fi.muni.cz

část 9

Obsah přednášky

Derivace

Integrál

Parciální derivace

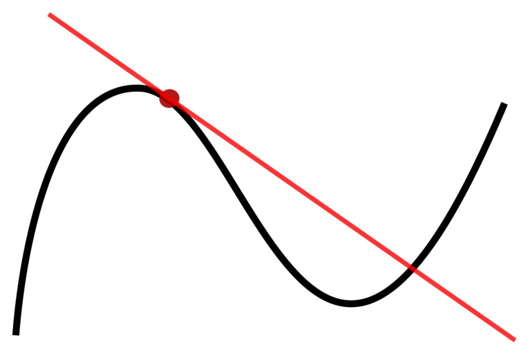
Strojové učení

Neuronová síť

Derivace

- Základní pojem diferenciálního počtu
- Směrnice tečny ke grafu funkce v určitém bodě
    - směrnice je číslo vyjadřující sklon přímky: y = ax + b
    - definována přes limitu: $f'(a) = \lim_{x \to a} \frac{f(x)-f(a)}{x-a}$
- Pro všechny body: nová (odvozená = derivovaná) funkce
    - např. $f : y = x^2 \quad \to \quad f' : y = 2x$
    - např. $f : y = 4x + 1 \quad \to \quad f' : y = 4$
- Využití: vyšetřování průběhu funkce
    - kladná derivace = funkce roste
    - záporná derivace = funkce klesá
    - nulová derivace = lokální minimum nebo maximum

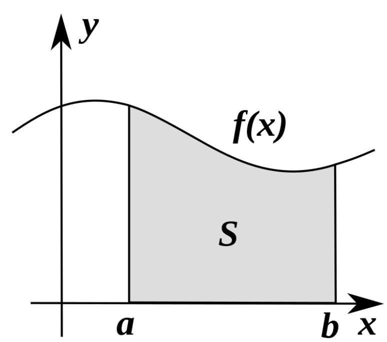
Příklad: derivace v tomto bodě je záporná



Integrál

- Inverzní operace k derivaci
    - tzv. základní věta integrálního počtu
- Opět z dané funkce „vyrábíme" jinou funkci
- Neurčitý integrál = primitivní funkce
    - množina funkcí, jejichž derivací dostaneme danou funkci
- Určitý integrál
    - změna primitivní funkce na daném intervalu
    - odpovídá ploše pod křivkou funkce
    - viz též Riemannův integrál
    - využití: výpočet obsahů, objemů, ...

Určitý integrál: příklad



zdroj: https://commons.wikimedia.org/wiki/File:Integral_as_region_under_curve.svg

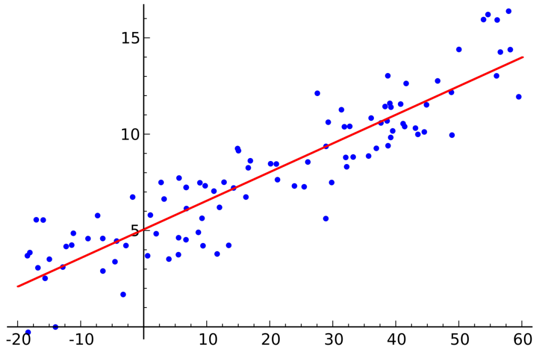
Parciální derivace

- U funkcí více proměnných
    - např. $f : z = 4y - 2x + 5$
    - např. $f : y = 4x_1 + 8x_2 + x_3 + 5x_4 - 8$
- Všechny proměnné až na jednu bereme jako konstanty
    - např. $f(x, y) = 4y^2 - 2x + 5$
    - $\frac{df}{dy} = 8y$
    - $\frac{df}{dx} = 2$
    - směrnice tečny v bodě „ve směru dané osy"

Strojové učení (supervised = s učitelem)

- Zobecnění z (mnoha) konkrétních příkladů
- 2 základní typy problémů:
- Klasifikace
    - cílem je přiřadit objekt do třídy
    - na základě jeho atributů (tzv. features)
    - features jsou číselné nebo řetězcové hodnoty
    - často mluvíme o vektoru hodnot
    - např. „umět rozlišit jabka a hrušky na základě barvy a tvaru"
- Regrese
    - cílem je na základě atributů přiřadit číselnou hodnotu
    - např. „na základě parametrů automobilu přiřadit cenu"

Lineární regrese



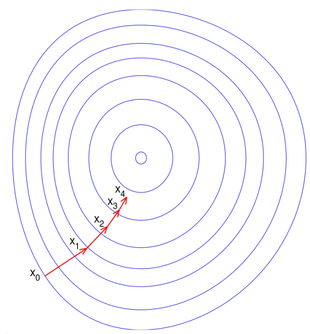
Lineární regrese

- Snažíme se co nejlépe proložit přímku množinou bodů
  - např. ve dvourozměrném prostoru
  - přímka: $y = ax + b$
  - snažíme se najít parametry a a b
  - tak, abychom minimalizovali chybovou funkci (loss function)
- Chybová funkce e
  - součet rozdílů (čtverců) odchylek od přímky
  - pro každou dvojici a a b dostaneme číselnou hodnotu chyby
  - e je tedy funkcí dvou proměnných, a a b
  - potřebujeme najít a a b tak, že chyba $e(a, b)$ je minimální

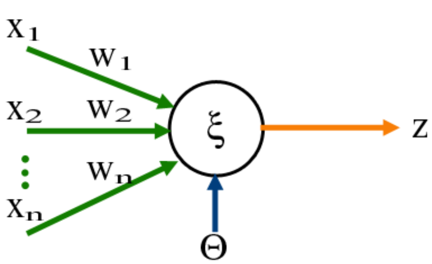
Lineární regrese

- Steepest gradient descent
  - začneme v náhodném bodě (a, b)
  - spočítáme parciální derivace chybové funkce $\frac{de}{da}$ a $\frac{de}{db}$
  - vektor $(\frac{de}{da}(a), \frac{de}{db}(b))$ se nazývá gradient
  - posuneme náš odhad (a, b) opačným směrem a opakujeme
  - (tedy $\frac{de}{da}(a)$ říká, kam máme posunout parametr a)
  - ($\frac{de}{db}(b)$ říká, kam máme posunout parametr b)
  - obě derivace nulové = jsme v lokálním minimu
  - a tedy chyba je nejmenší možná
  - (v některých příznivých situacích)
- Totéž lze dělat pro libovolnou funkci a libovolný počet dimenzí

Steepest gradient descent

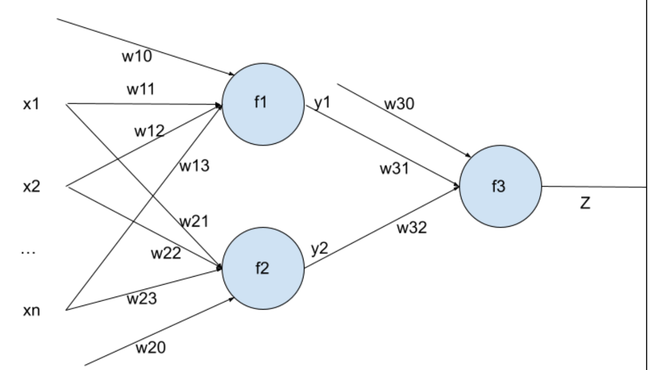


Neuron



výpočet: $Z = x_1w_1 + x_2w_2 + \ldots + x_nw_n + w_0$
(neuron je tedy jednoduchá lineární funkce)
trénování: optimalizace $(w_0..w_n)$ pomocí steepest gradient descent

Neuronová síť



Trénování neuronové sítě

- Backpropagation: algoritmus zpětné propagace chyby
  - začneme s náhodnými vahami w_{ij}
  - provedeme výpočet na trénovacích příkladech
  - provedeme gradient descent a úpravy parametrů w_{ij}
  - opakujeme, dokud nedosáhneme minima (nulových derivací podle všech parametrů w_{ij})
- Proč „zpětná propagace"
  - chyba e je složená funkce
  - derivace složené funkce je součinem derivací jejích částí
  - např. pro $\frac{de}{dw_{11}}$ musíme tedy nejdřív spočítat $\frac{de}{dZ}$
  - potom $\frac{df_3}{dy_1}$ a $\frac{df_3}{dy_2}$
  - a nakonec $\frac{df_1}{dw_{11}}, \frac{df_1}{dw_{12}}, \ldots, \frac{df_2}{dw_{21}}$ atd.
  - postupujeme tedy zprava doleva, opačně než výpočet