

Úvod do Umelej Inteligencie

Cvičenie 8 - Regresia

November 9, 2022

8. LINEÁRNA A KVADRATICKÁ REGRESIA

Budeme programovať funkcie ktoré k zadaným dátam vypočítajú lineárnu a kvadratickú regresiu.

Regresia - vzorce:

Keďže odvodenie vzorcov na výpočet regresie pomocou minimalizácie regresnej chyby je zdĺhavé, uvádzame vzorce už odvodené - nemusíte ich nijak meniť ani upravovať, stačí implementovať do kódu v takej verzii ako sú tu:

Lineárna regresia: Pre n fitovaných bodov bude výsledná priamka $y = ax + b$ - výpočet parametrov a, b :

$$a = \frac{n \sum (x_i y_i) - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$
$$b = \frac{\sum y_i - a \sum x_i}{n}$$

Kvadratická regresia: Pre n fitovaných bodov bude výsledná parabola $y = ax^2 + bx + c$ - výpočet parametrov a, b, c :

$$a = \frac{S(x^2 y) S(x x) - S(x y) S(x x^2)}{S(x x) S(x^2 x^2) - (S(x x^2))^2}$$
$$b = \frac{S(x y) S(x^2 x^2) - S(x^2 y) S(x x^2)}{S(x x) S(x^2 x^2) - (S(x x^2))^2}$$
$$c = \frac{\sum y_i - b \sum x_i - a \sum x_i^2}{n}$$

Pričom:

$$S(x x) = \sum x_i^2 - \frac{1}{n} (\sum x_i)^2$$
$$S(x y) = \sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i$$
$$S(x x^2) = \sum x_i^3 - \frac{1}{n} \sum x_i \sum x_i^2$$

$$S(x^2 y) = \sum x_i^2 y_i - \frac{1}{n} \sum x_i^2 \sum y_i$$

$$S(x^2 x^2) = \sum x_i^4 - \frac{1}{n} (\sum x_i^2)^2$$

Program:

Po spustení *regression.py* sa vygenerujú dva typy dát - body náhodnej priamky + šum a body náhodnej paraboly + šum. Tieto sa posielajú ako argumenty do funkcií *linear_regression* a *quadratic_regression*, ktoré máte za úlohu naprogramovať. Dáta sú reprezentované ako list dvojíc (x, y). Následne sa dáta (aj s vašimi vypočítanými regresiami) vykreslia v grafe.

Súčasťou grafov je aj vypočítaná regresná chyba - Mean Squared Error (MSE), teda suma štvorcov rozdielov medzi skutočnými bodmi a predikovanými hodnotami:

$$MSE = \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2, \text{ kde } \tilde{y}_i = ax_i + b \text{ v lineárnej regresii a } \tilde{y}_i = ax_i^2 + bx_i + c \text{ v kvadratickej.}$$

Okrem toho sa vykreslia aj testovacie dáta - tieto ležia na tej istej priamke resp. parabole, no pre x z iného intervalu, než boli fitované dáta. Rovnako sa pre ne vypočíta aj regresná chyba.

Ak váš program funguje správne, chyba kvadratickej regresie na fitovaných dátach by nikdy nemala byť vyššia než chyba lineárnej regresie.

Úloha 1 (0.5b): Do funkcie *linear_regression* v súbore *regression.py* naprogramujte vzorce na výpočet parametrov lineárnej regresie.

Úloha 2 (0.5b): Do funkcie *quadratic_regression* v súbore *regression.py* naprogramujte vzorce na výpočet parametrov kvadratickej regresie.

Výpočty si môžete výrazne zjednodušiť knižnicou *numpy* (napr. na sumy, maticové operácie atď.), no samotné **vzorce naprogramujte sami**, t. j. použitie existujúcich funkcií na výpočet regresie, napríklad *np.polyfit()* nie je povolené. Odporúčame ich ale použiť pri testovaní/debugovaní pre kontrolu správnosti vášho riešenia.