Projekt z Metód voľnej optimalizácie: Lineárne najmenšie štvorce (Predikcia HDP)

V tomto projekte sa bude odhadovať budúca hodnota HDP na základe množstva peňazí v ekonomike a hodnoty cenového deflátora.

K dispozícii máte súbor `HDP.txt` s kvartálnymi údajmi o americkom HDP za roky 1950 až 1980. Každý z $m = 124$ riadkov prislúcha jednému kvartálu, ktorého označenie sa nachádza v prvom stĺpci. Druhý stĺpec obsahuje premennú Y s hodnotami nominálneho HDP (v biliónoch dolárov), tretí stĺpec $M1$ obsahuje miery ponuky peňazí (takisto v biliónoch dolárov) a stĺpec P hodnoty implicitného cenového deflátora.

Ponuka peňazí v ekonomike značí celkové množstvo peňazí dostupných v ekonomike v danom okamihu. Označením $M1$ rozumieme peniaze vo forme hotovosti a depozitov. Implicitný cenový deflátor, známy aj ako deflátor HDP, je súhrnný cenový index, ktorý odráža vývoj cenovej hladiny tým, že zohľadňuje zmeny cien všetkých tovarov a služieb v ekonomike. Definuje sa aj ako podiel nominálneho a reálneho HDP v danom časovom okamihu.

Na predikovanie hodnoty HDP použijeme lineárny regresný model

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \tag{1}$$

kde $\hat{y}$ je odhad našej závislej premennej y (stĺpec Y s hodnotami HDP) a x_1 a x_2 sú nezávislé premenné modelu, t.j. vektor hodnôt ponuky peňazí $M1$ a vektor cenového deflátora P. Úlohou je odhadnúť parametre β_0, β_1 a β_2 tak, aby sa minimalizovali chyby merania $y_i - \hat{y}_i$, ktoré sa v regresii nazývajú rezíduá. V praxi sa minimalizujú štvorce týchto rezíduí a rieši sa tak úloha voľnej optimalizácie v tvare

$$\text{Min} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2. \tag{2}$$

a) Ukážte, že úloha (2) sa dá prepísať do kompaktného tvaru

$$\text{Min} \|X\beta - y\|_2^2, \tag{3}$$

kde $\beta = (\beta_0, \beta_1, \beta_2)^T$. Ako vyzerajú stĺpce matice X? Explicitne vyjadrite riešenie úlohy (3). V ďalších podúlohách ho využite pri dolaďovaní kódu.

b) Riešte úlohu (3) pomocou gradientnej metódy s konštantným krokom. Experimentujte s voľbou štartovacieho bodu a s dĺžkou kroku. Pre aké dĺžky kroku metóda konverguje?

c) Riešte úlohu (3) pomocou gradientnej metódy s približne optimálnym krokom (na jeho nájdenie použite metódu backtrackingu s parametrami $\alpha = 0.1$ a $\delta = 0.5$) a optimálnym krokom (zvoľte vhodnú metódu na jeho nájdenie). Opäť experimentujte s voľbou štartovacieho kroku a výsledky porovnajte s výsledkami z časti b). Ktorá metóda našla najlepšie riešenie? (V ďalších častiach stačí pracovať s tým.)

Poznámka: V štatistike sa kvalita modelu určuje na základe hodnoty tzv. koeficientu determinácie:

$$R^2 = 1 - \sum_{i=1}^{m} \frac{(\hat{y}_i - y_i)^2}{(\bar{y}_i - y_i)^2},$$

kde $\bar{y}$ je priemer nezávislých premenných y. Koeficient determinácie nadobúda hodnoty z intervalu $[0, 1]$ a hovorí, aký podiel rozptylu závislej premennej je vysvetlený nezávislými premennými. Čím je jeho hodnota bližšie k 1, tým je model lepší.

d) V súbore `HDP_test.txt` nájdete údaje o americkom HDP vo všetkých kvartáloch rokov 1981 až 1983. Aplikujte získané koeficienty modelu (1) na tieto dáta a odhadnite pomocou nich hodnoty HDP v týchto kvartáloch. Posúďte kvalitu tejto predikcie.

e) Podľa makroekonomického IS-LM modelu zvýšenie ponuky peňazí vyvolá zvýšenie HDP (pri zanedbaní ostatných faktorov). Overte, či táto skutočnosť platí pre reálne dáta zo súboru `HDP.txt`. Pracujte teda s modelom (1) pre $\beta_2 = 0$ a na riešenie použite gradientnú metódu podľa vlastného uváženia. Dáta aj nájdenú regresnú priamku vykreslite v dvojrozmernom priestore a opäť overte kvalitu predikcie na dátach zo súboru `HDP_test.txt`.

f) *Nadstavba:* Vymyslite vhodné rozšírenie, či modifikáciu projektu. Porovnajte použité metódy s inými metódami z prednášok, prípadne zdôvodnite, prečo sa niektoré metódy na riešenie úlohy lineárnych najmenších štvorcov nepoužívajú.