Domácí úkol – optimalizační úlohy

Úkol 1 (snadné, každý vypracuje úkol samostatně)

Mějme náhodný výběr z normálního rozdělení:

168, 179, 162, 169, 158, 168, 169, 169, 181, 160

Odhadněte střední hodnotu a rozptyl metodou maximální věrohodnosti, a to:

a) pomocí analytického řešení v explicitním tvaru (existují přímo vzorečky pro odhad)
b) sestavením věrohodnostní funkce a nalezením jejího maxima vybranou metodou optimalizace

(Pokud někdo nezná metodu maximální věrohodnosti, doučí se jí sám za domácí úkol, případně se zeptá a nechá si ji vysvětlit).

Praktické tipy:

Napište si věrohodnostní funkci v Rku jako funkci dvou parametrů (střední hodnoty a rozptylu), tak aby výstupem byla hodnota věrohodnostní funkce pro libovolně zvolené kombinace střední hodnoty a rozptylu vzhledem k dostupným datům výše. Tu následně maximalizujte.

Z praktických důvodů (výpočetně mnohem stabilnější) je lepší maximalizovat logaritmus věrohodnostní funkce. Logaritmus věrohodnostní funkce má maximum ve stejném bodě jako samotná věrohodnostní funkce.

Odhad rozptylu musí být kladné číslo. V praktických úlohách toho lze docílit například výběrem vhodného algoritmu, který umožňuje použít restrikce na odhadované parametry (umožňuje hledání řešení pouze v rámci určitého smysluplného intervalu). To nemusí být vždy výhodné, protože takový algoritmus je zpravidla pomalejší a někdy i méně spolehlivý. Druhá možnost je optimalizovat reparametrizované řešení, kdy rozptyl je vyjádřen jako rozptyl = exp(parametr vstupující do optimalizačního algoritmu). I pokud by byla optimální hodnota parametru záporná, funkce exp zajistí, že výsledný odhad rozptylu bude kladné číslo. V této úloze reparametrizace pravděpodobně nebude potřeba, ale proč si ji nevyzkoušet☺.

Optimalizační algoritmy dostupné v různých SW občas umožňují hledat buď pouze minimum, nebo pouze maximum funkce. Úlohu minimalizace funkce lze převést na maximalizační úlohu vynásobením dané funkce -1 (minimum funkce f(x) se nachází ve stejném bodě jako maximum funkce –f(x) a obráceně).

Přečtěte si nápovědu k R funkci: `optim` (a případně i `optimize`)

Úkol 2 (snadné, každý vypracuje úkol samostatně)

Nalezněte minimum funkce:

$f(x,y) = (x - 3{,}14)^2 + (y - 2.72)^2 + \sin(3x + 1{,}41) + \sin(4y - 1{,}73)$

pro $0 < x < 5$, $0 < y < 5$

a) prostřednictvím tradičních optimalizačních algoritmů založených na gradientu (viz funkce `optim`)
b) prostřednictvím evolučních algoritmů (viz balíček `ga` a nápověda k funkcím)
c) prostřednictvím algoritmu PSO (viz balíček `pso` a nápověda k funkcím)

Úkol 3 (středně náročné, vypracujte v týmu)

Formulujte libovolnou regresní úlohu – výběr je zcela na Vašem uvážení (jedna vysvětlovaná proměnná a několik vysvětlujících proměnných). Získejte hodnoty regresních parametrů:

a) metodou nejmenších čtverců (tj. minimalizujte ztrátovou funkci „L2“, která je dána jako součet čtverců reziduí (odchylek pozorovaných hodnot od hodnot vyrovnaných regresní funkcí). Použijte analytické řešení.
b) metodou nejmenších absolutních odchylek (tj. minimalizujte ztrátovou funkci „L1“, která je dána jako součet absolutních hodnot reziduí (odchylek pozorovaných hodnot od hodnot vyrovnaných regresní funkcí). Řešte pomocí optimalizačních metod z Úkolu 2, výsledky porovnejte s výsledky získanými metodou nejmenších čtverců. Jak (a proč) se od sebe výsledky liší?
c) Dobrovolná část úkolu: Řešte zadání z bodu b) pomocí metody IRLS (*iteratively reweighted least squares*), viz třetí přednáška nebo: https://en.wikipedia.org/wiki/Iteratively_reweighted_least_squares