Stručné shrnutí semináře 11

Interpolace funkčních závislostí (fitování). Studujeme chování veličiny y v závislosti na x, máme tedy k dispozici n dvojic (x_i,y_i) ... např. naměřená data. Nejistotu nezávislé veličiny x považujeme obvykle za zanedbatelnou vůči nejistotě závislé veličiny y.
Předpokládáme nějaký konkrétní tvar funkční závislosti $y=f(x)$ a chceme posoudit jeho platnost, resp. získat hodnoty parametrů této závislosti. K tomu lze využít **metodu nejmenších čtverců**, která spočívá v minimalizaci veličiny

$$\chi^2(\alpha,\beta,\gamma,...)=\sum_{i=1}^{N}\frac{\left(f(x_i\,|\,\alpha,\beta,\gamma,...)-y_i\right)^2}{\sigma_{y_i}^2}$$

Hodnoty hledaných parametrů jsou ty, které minimalizují χ^2.

Je-li $f(x)$ **lineární** funkcí parametrů $\alpha, \beta, \chi,...$, lze pro problém minimalizace χ^2 nalézt **analytické** řešení a hovoříme o **lineární regresi**. Některé nelineární funkce lze linearizovat (exponenciálu lze zlogaritmovat).

V ostatních případech analytické řešení nemáme a minimum χ^2 je nutno hledat **numericky**. Pak jde o **nelineární regresi**, iterativní numerický proces, který je nutno uživatelsky kontrolovat. Pro složitější $f(x)$ často velice záleží na počáteční volbě parametrů $\alpha, \beta, \chi,...$. Při nevhodné volbě parametrů může algoritmus uvíznout v lokálním minimu χ^2, a tak dospět k nesprávnému výsledku.

Náhodné poznámky a rady:

- Většinou jsou fyzikální veličiny spojité funkce. $f(x)$ je pak tedy také spojitá funkce (tj. čára v grafu). (Neboli: naměřené body nejsou závislost. Ale využíváme je ke konstrukci té hledané závislosti.)
- Je rozdíl mezi vyhlazováním (tj. interpolování po částech pomocí polynomů – např. metodou *spline*) a výše popsanou regresí. Protože ve fyzice obvykle známe tvar hledané funkce $f(x)$, proto se ho vždy snažíme použít při fitování. Interpolace pomocí *spline* sice může být v grafu hezké vodítko pro oko, ale nemá žádný fyzikální význam.
- Některé algoritmy, veličiny nebo charakteristiky mohou mít v různých programech odlišnou implementaci či definici. Vždy je dobré se přesvědčit, jak je to v daném programu zadefinováno. Obecně vždy byste měli mít jistotu, že program dělá právě a jenom to, co chcete, aby dělal.
- Obvykle program umí znázornit chybové úsečky (nejistoty fitovaných dat), a často je i umí zohlednit (jako váhy jednotlivých bodů) při výpočtu hodnot hledaných parametrů funkce $f(x)$. To ale ještě neznamená, že je umí promítnout také do nejistot těch získaných parametrů. Zpravidla tomu tak není. Ke správnému započtení by totiž byly potřeba dodatečné předpoklady o charakteru nejistot fitovaných dat (např. že jsou normálně rozděleny, jak jsou korelovány, apod.).
- Fit přímkou je sice častý případ lineární regrese, nicméně lineární regrese je obecnější pojem – jde o fitování jakoukoliv funkcí $f(x; \alpha, \beta, \chi, ...)$, která je lineární kombinací svých parametrů $\alpha, \beta, \chi, ...$ (např. polynom n-tého stupně proměnné x), tj. vede na lineární problém řešitelný analyticky. Někdy lidé nesprávně hovoří o lineární regresi, když mají na mysli fit lineární závislostí (přímkou), a o nelineární regresi, když mají na mysli fit nelineární funkcí (např. parabolou).
- Reziduální analýza a další nástroje (statistické charakteristiky) mohou pomoci posoudit, jak je zvolený model vhodný.