

Neparametrická analýza funkcionálních dát

Daniela Kuruczová a Jan Koláček

13.9.2016

Přírodovědecká fakulta, Masarykova univerzita

Úvod

Cieľ práce: Predstaviť zovšeobecnenie jadrových metód na funkcionálne dáta, otestovať metódy výberu vyhladzovacieho parametru a predstaviť možný smer využitia.

Obsah:

- Jadrové metódy pre funkcionálne dáta
- Voľba vyhladzovacieho parametru
- Aplikácia
- Záver

Teoretická část

- Funkcionálna náhodná veličina (FNV) – rozšírenie pojmu náhodná veličina na nekonečne dimenzionálne priestory.
- FNV značíme \mathcal{X} , pozorovanie FNV značíme χ .
- Vzdialenosť, preklatie dimenzionality, semimetrika

$$d_q^{\text{deriv}}(\mathcal{X}_i, \chi) = \sqrt{\int \left(\mathcal{X}_i^{(q)}(t) - \chi^{(q)}(t) \right)^2 dt}$$

$$d_q^{\text{PCA}}(\mathcal{X}_i, \chi) = \sqrt{\sum_{k=1}^q \left(\int [\mathcal{X}_i(t) - \chi(t)] v_k(t) dt \right)^2}$$

Jadrová regresia pre funkcionálne dáta

- Dvojice (\mathcal{X}_i, Y_i) nezávislé, rovnako rozdelené, $\varepsilon_i \sim N(0, \sigma^2)$ pre $i = 0, 1, \dots, n$

$$Y_i = r(\mathcal{X}_i) + \varepsilon_i = \mathbb{E}(Y_i | \mathcal{X}_i) + \varepsilon_i$$

- Nadarajov–Watsonov odhad regresnej funkcie

$$\hat{r}(\chi) = \frac{\sum_{i=1}^n Y_i K\left(\frac{d(\chi, \mathcal{X}_i)}{h}\right)}{\sum_{i=1}^n K\left(\frac{d(\chi, \mathcal{X}_i)}{h}\right)}$$

- Metódy globálnej voľby

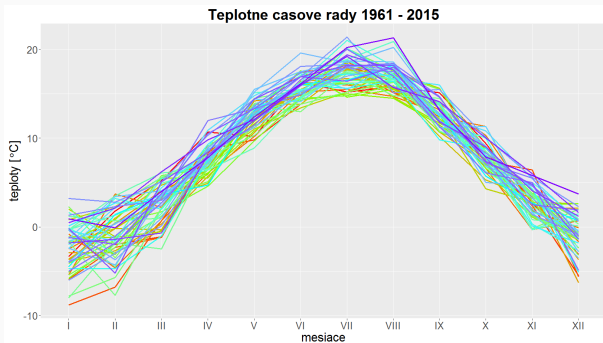
- Metóda krížového overovania $CV(h) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{r}_{-i}(\mathcal{X}_i, h))^2$
- Metóda penalizačných funkcií

$$PF(h) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{r}(\mathcal{X}_i, h))^2 \equiv \left(\frac{\mathcal{K}\left(\frac{d(x, \mathcal{X}_i)}{h}\right)}{\sum_{i=1}^n \mathcal{K}\left(\frac{d(x, \mathcal{X}_i)}{h}\right)} \right)$$

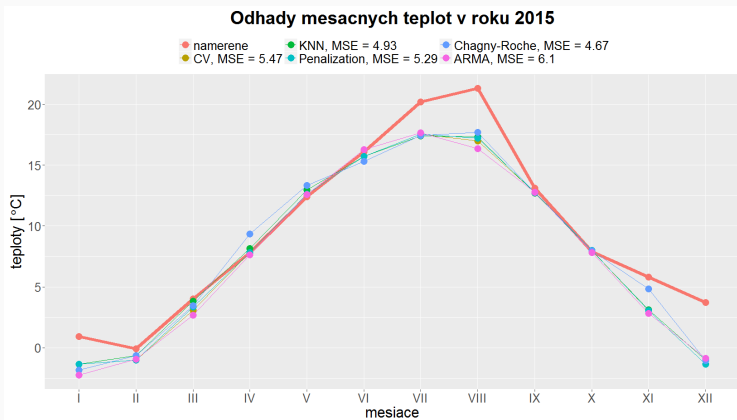
- Metódy lokálnej voľby

- Metóda k najbližších susedov
- Metóda Chagny–Roche

Praktická část



Obr. 1: Krivky mesačných teplôt



Obr. 2: Odhad pre rok 2015

Záver

- Oblasť analýzy funkcionálnych dát – relatívne nová, stále sa rozvíja
- Potrebné algoritmy pre ladenie parametrov (q , κ v Chagny-Roche)
- Dôraz na využívanie reálnych dát

Ďakujem za pozornosť.