Odhady fraktální dimenze bayesovským způsobem

Ing. David Blatský

Katedra softwarového inženýrství
Fakulta jaderná a fyzikálně inženýrská
České vysoké učení technické v Praze

7. leden 2015 v Praze

školitel: prof. Ing. Pavel Sovka, CSc.
školitel specialista: doc. Ing. Jaromír Kukal, Ph.D.

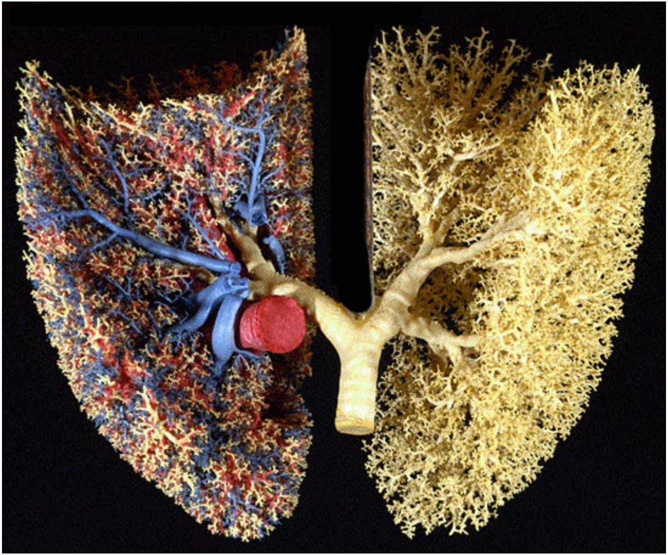
Fraktál

- neexistuje přesná definice
- množiny, jejichž geometrický motiv se opakuje
- nekonečně členitý útvar $\times$ geometricky hladký
- Mandelbrot: "Tvar tvořený částmi, které jsou podobné celku."
- Zelinka: "Fraktál je objekt, jehož geometrická struktura se opakuje v něm samém. Fraktály se dělí na soběpodobné a soběpříbuzné."

Fraktály v přírodě

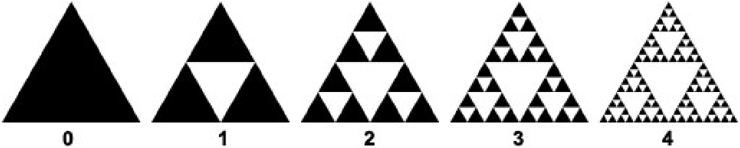
Obrázek : Listy kapradí [3]

Fraktály v přírodě

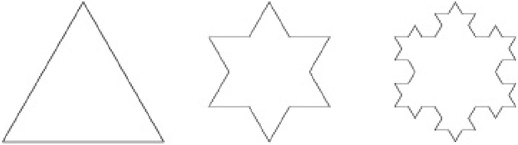


Obrázek : Lidské plíce [3]

Geometrické fraktály



Obrázek : Sierpinského trojúhelník [3]

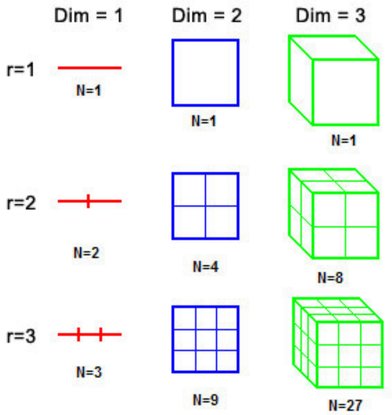


Obrázek : Kochova vločka

Dimenze

- zvětšování $r = 1, 2, 3$

$$N = r^D$$
$$\ln N = lnr^D$$
$$D = \frac{\ln N}{\ln r}$$

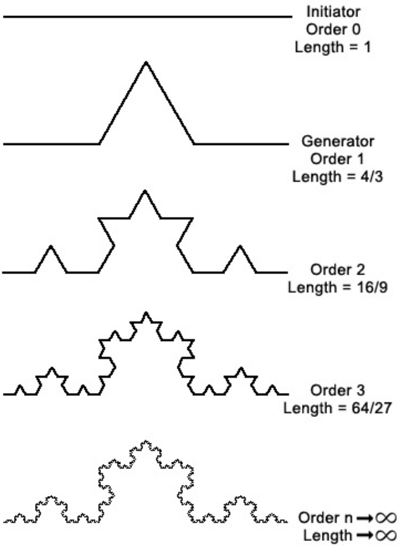


Soběpodobnostní dimenze

- stejný přístup

$$N = 4$$
$$r = 3$$
$$D_S = \frac{\ln N}{\ln r}$$
$$D_S = \frac{\ln 4}{\ln 3}$$
$$D_S = 1,26$$

- geometrické fraktály



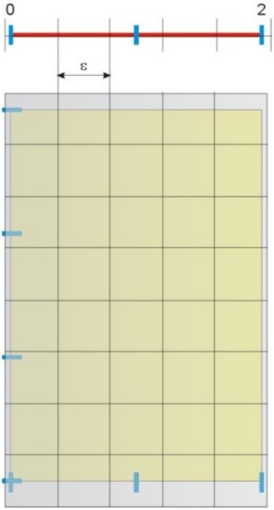
Kapacitní dimenze

- pokrytí intervalu boxy s rozměrem $a > 0$
- $C(a)$ - počet boxů
- A - délka přímky, obsah obdélníku, ...

$$C(a) \approx A\left(\frac{1}{a}\right)^{D}$$

$$D \approx \frac{\ln C(a)}{\ln \frac{1}{a}} - \frac{\ln A}{\ln \frac{1}{a}} \approx \frac{C(a)}{\ln \frac{1}{a}}$$

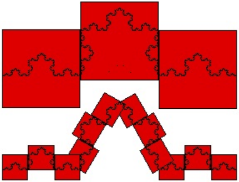
$$D_K = \lim_{a \to 0} \frac{C(a)}{\ln \frac{1}{a}}$$



Obrázek : Dělení intervalu a obdélníku boxy velikosti a

Boxcounting - odhadování dimenze

- vychází z předchozího vztahu
  - není možné jít $a \to 0$
- ve formě $\ln C(a) = A - D_0 \ln a$
  - $D_0 = D_K$ – kapacitní dimenze
  - A – neznámá konstanta
- vypočítat $C(a)$ pro několik a, následně proložit přímkou



Multinomické rozdělení

- $n \in \mathbb{N}$ – počet různých jevů
- $p_j > 0$ – pravděpodobnost j-tého jevu
  - $j = 1, \dots, n$
  - $\sum_{j=1}^{n} p_j = 1$
- potom j má $\mathrm{Mul}(p_1, \dots, p_n)$
- $N \in \mathbb{N}$ realizací, $N_j \in \mathbb{N}_0$, $\sum_{j=1}^{n} N_j = N$
- počet různých jevů $K = \sum_{N_j > 0} 1 \leq \min(n, N)$

Naivní odhad Hartleyovy entropie

- Hartleyova entropie – $H_0 = \ln n$
- odhad – $\hat{H}_{0,\mathrm{naive}} = \ln K$
- vychýlení
  - $K \in \{1, \dots, n\}$ potom $\mathrm{E}\hat{H}_{0,\mathrm{naive}} = \mathrm{E} \ln K < \mathrm{E} \ln n = \ln n = H_0$
- $\ln C(a) = A_0 - \hat{D}_{0,\mathrm{naive}} \ln a$
  - ale $C(a) = K \rightarrow \ln C(a) = \ln K = \hat{H}_{0,\mathrm{naive}}$
- $\hat{H}_{0,\mathrm{naive}} = A_0 - \hat{D}_{0,\mathrm{naive}} \ln a$

Revidovaný odhad Hartleyovy entropie I

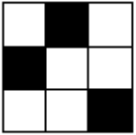
- máme $\mathrm{Dir}(p_1, \dots, p_n)$ s $\alpha_j = \alpha^* > 0$
- $\hat{\mathrm{p}}(K \mid n, N) =$
  $$\mathrm{P}\left(\sum_{N_j>0} 1 = K \,\middle|\, n, \sum_{j=1}^{n} N_j = N\right) =$$
  $$= \binom{n}{K} \frac{\Gamma(N+1)\Gamma(n\alpha^*)}{\Gamma(N+n\alpha^*)} \sum_{\vec{N}\in\mathbb{D}_{K,N}} \prod_{j=1}^{K} \frac{\Gamma(N_j+\alpha^*)}{\Gamma(N_j+1)\Gamma(\alpha^*)},$$
  kde $\mathbb{D}_{K,N} = \{\vec{x} \in \mathbb{N}^K | \sum_{j=1}^{K} x_j = N\}$
- $S_{K,N} = \sum_{n=K}^{\infty} \hat{\mathrm{p}}\left(K \mid n, N\right)$

Revidovaný odhad Hartleyovy entropie II

- $\hat{\mathrm{p}}\left(n \mid K, N\right) = \frac{\hat{\mathrm{p}}(K \mid n,N)}{S_{K,N}}, n \geq K$
- $\hat{H}_{0,\text{Bayes}} = \mathsf{E}H_0 = \sum_{n=K}^{\infty} \hat{\mathrm{p}}\left(n \mid K, N\right) \ln n =$
  $= \sum_{n=K}^{\infty} \frac{\hat{\mathrm{p}}(K \mid n,N) \ln n}{S_{K,N}} = \frac{\sum_{n=K}^{\infty} \hat{\mathrm{p}}(K \mid n,N) \ln n}{\sum_{n=K}^{\infty} \hat{\mathrm{p}}(K \mid n,N)} > \ln K$
- $\hat{H}_{0,\text{Bayes}} = \frac{\sum_{j=0}^{\infty} b_j \ln\left(K+j\right)}{\sum_{j=0}^{\infty} b_j},$
  kde $b_j = \binom{K+j}{j} \frac{\mathrm{B}((K+j)\alpha^*,N)}{\mathrm{B}(K\alpha^*,N)}$

Experimenty I

- $\hat{H}_{0,\text{Bayes}} = A - \hat{D}_0 \ln a$
- 21 2D frakrálů, rekurzivní expanze matice
  - $\mathbb{G}_{u,v} \in \{0,1\}^{v \times v}$
  - u počet nenulových prvků
  - $v > 1$ dimenze matice
  - $D_0 = \frac{\log u}{\log v}$



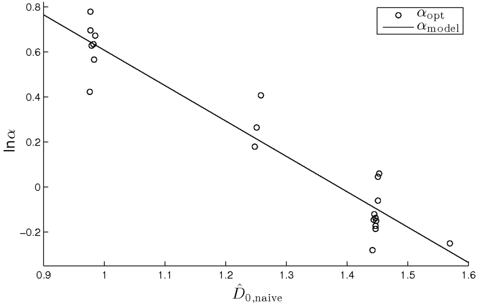
Obrázek : Bázová matice pro generování fraktálu $F_{3,3,7}$

Experimenty II

- náhodná rotace a náhodné posunutí
- položena síť o hraně $a = 12, 16, 20, \ldots, 480, 500$
- vypočítán $\hat{H}_{0,\text{naive}}$ $20\times$
- vypočítán $\hat{D}_{0,\text{naive}}$ z průměru $\hat{H}_{0,\text{naive}}$
- navržen α_{opt} pro $\hat{D}_{0,\text{Bayes}}$

Návrh nejlepšího α

- vztah mezi α_{opt} a $\hat{D}_{0,\mathrm{naive}}$
  - vyzkoušeny 3 modely
  - nejlepší je exponenciální $\ln \alpha = A + B\hat{D}_{0,\mathrm{naive}}$
  - $A = 2.178,\ B = -1.571$



Obrázek : α_{opt} hodnoty exponenciálního modelu a jeho lineární regrese

Závěr

- vyvinut odhad $\hat{H}_{0,\mathrm{Bayes}}$ pro Dirichletovo apriorní rozdělení
- procedura pro zlepšení odhadu:
  - odhadnout $\hat{D}_{0,\mathrm{naive}}$
  - určit α podle navrženého modelu
  - znovu odhadnout kapacitní dimenzi $\hat{D}_{0,\mathrm{Bayes}}$ pro α
- plánované rozšíření:
  - odhady informační dimenze D_1
  - testování na 1D, 3D, ... fraktálech

Aplikace

- odhadování Alzeimerovy choroby z PET snímků mozku
- CT a MRI snímky plic odhalování rakoviny

Zdroje

1 MANDELBROT, The fractal geometry of nature. 1982

2 ZELINKA, Fraktální geometrie - principy a aplikace.

3 fractalfoundation.org

4 ksr.tul.cz/fraktaly