

Kontekstinis duomenų, aprašomų erdviniais apibendrintais tiesiniais modeliais, klasifikavimas

Atsiskaitymas už ketvirtą doktorantūros semestrą

Doktorantė: Eglė Zikarienė

Vadovas: prof. dr. Kęstutis Dučinskas

Konsultantas: prof. dr. Julius Žilinskas

2019-10-24

Tyrimo objektas:

- Erdvinių duomenų modeliai,
- Klasifikavimo procedūra, paremta Bajeso diskriminantine funkcija (BDF)

Tyrimo tikslas:

Sudaryti ir ištirti BDF išraiškas eksponentinės šeimos skirstiniams ir eliptinių šeimos t skirstiniui, siekiant praplėsti klasifikavimo procedūros pritaikymo galimybes.

Planuojami rezultatai:

- BDF išraiškos, Bernulio, Binominis, Puasono, Gamma, Beta, t skirstiniams;
- Tikslios tikrųjų klasifikavimo klaidų formulės;
- Vidutinių klasifikavimo klaidų tikimybių asimptotinės aproksimacijos;
- Pasiūlytos procedūros teorinis pagrindas;
- Pasiūlytos procedūros algoritminė realizacija;

4-to doktorantūros semestro darbo planas

Egzaminai:

- Lygiagretieji ir paskirstytieji skaičiavimai (9 kreditai, 2019 m. spalio 28 d., įvertinimas 10)

Disertacijos rengimas:

- Teorinis tyrimas. Erdvinių duomenų, aprašomų apibendrintais tiesiniais modeliais, kontekstinio klasifikavimo klaidos tikimybinių analitinių išraiškų ir aproksimacijų išvedimas naudojant klasikinius ir Bajeso parametrų įvertinimus.

Nagrinėjamas duomenų modelis

Atsitiktinis laukas $\{Z(s) : s \in D \subset R^p\}$.

Stebėjimo $Z(s)$ modelis populiacijoje Ω_l yra

$$Z(s) = x'(s)\beta_l + \varepsilon(s).$$

Nagrinėjamas $Z_0 = Z(s_0)$ klasifikavimo uždavinys su mokymo imtimi T į vieną iš populiacijų, kai naudojama fiksuota, ne stochastinė mokymo imties struktūra.

Klasifikavimo procedūra

Bajeso diskriminantinė funkcija

$$W(Z_0|T) = \ln \left(\frac{p(Z_0|T, \Omega_1)P(\Omega_1)}{p(Z_0|T, \Omega_2)P(\Omega_2)} \right) = \ln \left(\frac{p(Z_0|T, \Omega_1)}{p(Z_0|T, \Omega_2)} \right) + \ln \left(\frac{P(\Omega_1)}{P(\Omega_2)} \right)$$

Sudarant BDF naudojamos sąlyginio tankio funkcijos. Į BDF įvedama klasifikuojamo taško erdvinė priklausomybė su mokymo imtimi.

Įterptoji BDF

Įterptoji Bajeso diskriminantinė funkcija:

$$\begin{aligned} W\left(Z_0|T, \hat{\theta}\right) &= \ln \left(\frac{p\left(Z_0|T, \Omega_1, \hat{\theta}_1\right) P\left(\Omega_1\right)}{p\left(Z_0|T, \Omega_2, \hat{\theta}_2\right) P\left(\Omega_2\right)} \right) = \\ &= \ln \left(\frac{p\left(Z_0|T, \Omega_1, \hat{\theta}_1\right)}{p\left(Z_0|T, \Omega_2, \hat{\theta}_2\right)} \right) + \ln \left(\frac{P\left(\Omega_1\right)}{P\left(\Omega_2\right)} \right) \end{aligned}$$

$\hat{\theta}_l$ - nežinomų parametrų vektorius θ_l įverčiai.

***t* skirstinio BDF**

t skirstinio atveju gauta BDF išraiška (Dučinskas ir Zikarienė 2015):

$$W(Z_0|T, \theta) = \left(Z_0 - \frac{1}{2}(\mu_{1t}^0 + \mu_{2t}^0) - \alpha'(t - X\beta) \right) (\mu_{1t}^0 - \mu_{2t}^0)$$

Eksponentinių skirstinių šeima

Sąlyginio tankio išraiška eksponentinių skirstinių šeimai

$$p(Z_0|T;\theta) = \exp\{A_0(T;\theta)T_0(Z_0) - B_0(T;\theta) + C_0(Z_0)\}$$

Prielaida: tarp erdvinių duomenų galioja porinis (ang. pairwise) ryšys:

$$p(z) = f\left(\alpha + \sum_{1 \leq i \leq n} z_i \beta_i + \sum_{1 \leq i < j \leq n} z_i z_j \beta_{ij}\right)$$

Auto-Gamma modelis

Sąlyginis skirstinys: $Z_0^l | T \sim \Gamma(\alpha^l, \gamma_0(T))$,

čia α^l - mastelio parametras kiekvienai klasei l ,
 $\gamma_0(T)$ - formos parametras, kurio išraiška

$$\gamma_0(T) = \theta_0 + \sum_{t=1}^n \theta_{0t} z_t.$$

Auto-Gamma modelis II

Sąlyginis tankis:

$$p(Z_0^l = z_0 | T) = \exp \left\{ \begin{aligned} & -\frac{1}{\alpha^l} z_0 + \left(\theta_0 + \sum_{t=1}^n \theta_{0t} z_t \right) \ln z_0 + \\ & + \left(\theta_0 + \sum_{t=1}^n \theta_{0t} z_t \right) \ln \left(\frac{1}{\alpha^l} \right) - \\ & - \ln \left(\Gamma \left(\theta_0 + \sum_{t=1}^n \theta_{0t} z_t \right) \right) - \ln z_0 \end{aligned} \right\}$$

Auto-Gamma modelis III

BDF:

$$\begin{aligned} W(Z_0|T) &= \ln\left(\frac{p_1(Z_0|T, \Omega_1)}{p_2(Z_0|T, \Omega_2)}\right) + \ln\left(\frac{P(\Omega_1)}{P(\Omega_2)}\right) = \\ &= \left(-\frac{1}{\alpha^1} + \frac{1}{\alpha^2}\right)z_0 + \gamma_0\left(\ln\left(\frac{1}{\alpha^1}\right) - \ln\left(\frac{1}{\alpha^2}\right)\right) + \gamma = \\ &= \left(\frac{\alpha^1 - \alpha^2}{\alpha^1\alpha^2}\right)z_0 + \left(\theta_0 + \sum_{t=1}^n \theta_{0t}z_t\right)\ln\left(\frac{\alpha^2}{\alpha^1}\right) + \gamma \end{aligned}$$

$$\text{kai } \gamma = \ln\left(\frac{P(\Omega_1)}{P(\Omega_2)}\right).$$

Auto-Beta modelis

Sąlyginis skirstinys: $Z_0^l | T \sim \text{Beta}(p^l, q)$,

čia p, q - mastelio parametrai, p - skirtingi kiekvienai klasei l ,

$$p^l(T) = \theta_0^l + \sum_{t=1}^n \theta_{0t} z_t.$$

Auto-Beta modelis II

Sąlyginis tankis:

$$p(Z_0^l = z_0 | T) = \exp \left\{ \begin{aligned} & \left(\left(\theta_0^l + \sum_{t=1}^n \theta_{0t} z_t \right) - 1 \right) \ln z_0 + \\ & + (q - 1) \ln (1 - z_0) + \\ & + \ln \left(\frac{\Gamma \left(\theta_0^l + \sum_{t=1}^n \theta_{0t} z_t + q \right)}{\Gamma \left(\theta_0^l + \sum_{t=1}^n \theta_{0t} z_t \right) \Gamma(q)} \right) \end{aligned} \right\}$$

Auto-Beta modelis III

BDF:

$$\begin{aligned}
 W(Z_0|T) &= \ln\left(\frac{p_1(Z_0|T, \Omega_1)}{p_2(Z_0|T, \Omega_2)}\right) + \ln\left(\frac{P(\Omega_1)}{P(\Omega_2)}\right) = \\
 &= (p^1 - p^2) \ln z_0 + \left(\frac{\Gamma(p^1 + q)\Gamma(p^2)}{\Gamma(p^1)\Gamma(p^2 + q)}\right) = \\
 &= (\theta_0^1 - \theta_0^2) \ln z_0 + \ln\left(\frac{\Gamma\left(\theta_0^1 + \sum_{t=1}^n \theta_{0t} z_t + q\right)\Gamma\left(\theta_0^2 + \sum_{t=1}^n \theta_{0t} z_t\right)}{\Gamma\left(\theta_0^1 + \sum_{t=1}^n \theta_{0t} z_t\right)\Gamma\left(\theta_0^2 + \sum_{t=1}^n \theta_{0t} z_t + q\right)}\right) + \gamma \\
 \gamma &= \ln\left(\frac{P(\Omega_1)}{P(\Omega_2)}\right)
 \end{aligned}$$

Tikroji klaidos tikimybė

- Prielaida: laikysime, jog, kai $W(Z_0|T, \theta) = 0 \Rightarrow z_0 \in \Omega_1$.

- Kai (Dučinskas 2009) $l=1,2$

$$P(\theta) = \sum_l \pi_l P_{0l} = \pi_1 P_{01} \left((-1)W(Z_0|T, \theta) \geq 0 | \Omega_1 \right) + \\ + \pi_2 P_{02} \left(W(Z_0|T, \theta) > 0 | \Omega_2 \right)$$

- Nagrinėtiems modeliams BDF funkcijoje priklausomybė nuo Z_0 yra tiesinė.

Tikroji klaidos tikimybė II

- Bendras BDF pavidalas: $W(Z_0|T) = az_0 + b$
- Tikroji klaidos tikimybė:

$$\begin{aligned} P(\theta) &= \sum_l \pi_l P_{0l} = \pi_1 P_{01} \left(z_0 \leq -\frac{b}{a} \mid \Omega_1 \right) + \pi_2 P_{02} \left(z_0 > -\frac{b}{a} \mid \Omega_2 \right) = \\ &= \pi_1 F_{Z_0|T} \left(-\frac{b}{a} \right) + \pi_2 \left(1 - F_{Z_0|T} \left(-\frac{b}{a} \right) \right) \end{aligned}$$

Tikroji klaidos tikimybė III

| Modelis\Parametrai | a | b |
|--------------------|--|---|
| Auto-Beta | $(\theta_0^1 - \theta_0^2)$ | $\ln \left(\frac{\Gamma\left(\theta_0^1 + \sum_{t=1}^n \theta_{0t} z_t + q\right) \Gamma\left(\theta_0^2 + \sum_{t=1}^n \theta_{0t} z_t\right)}{\Gamma\left(\theta_0^1 + \sum_{t=1}^n \theta_{0t} z_t\right) \Gamma\left(\theta_0^2 + \sum_{t=1}^n \theta_{0t} z_t + q\right)} \right) + \gamma$ |
| Auto-Gamma | $\left(\frac{\alpha^1 - \alpha^2}{\alpha^1 \alpha^2} \right)$ | $\left(\theta_0 + \sum_{t=1}^n \theta_{0t} z_t \right) \ln \left(\frac{\alpha^2}{\alpha^1} \right) + \gamma$ |

Tikrosios klasifikavimo klaidos vertinimai

- Plug-in estimator
- Resubstitution estimator (R-method)
- Leave-One-Out estimator
- Jackknife error rate estimator
- DS method estimator
- O and OS estimator
- Posterior probability estimator

Pranešimai tarptautinėse konferencijose

- Zikarienė E., Dučinskas K. Application of spatial beta regression for modelling of the algae concentration index. Spatial statistics 2019, Sitges, Ispanija 2019 m. liepos 9 - 13 d.
- Zikarienė E., Dučinskas K. Implementation of generalized additive models for spatial beta regression. Computer data analysis and modeling: stochastics and data science. Minskas, Baltarusija 2019 m. rugsėjo 18 - 22 d.

Mokslinė publikacija

- Zikarienė E., Dučinskas K. Implementation of generalized additive models for spatial beta regression. Proceedings of the XII International Conference. p. 341-343.

2019-10/2020-10 metų darbo planas

Disertacijos rengimas:

Erdvinių duomenų, aprašomų apibendrintais tiesiniais modeliais, kontekstinio klasifikavimo klaidos tikimybių analitinių išraiškų ir aproksimacijų išvedimas naudojant klasikinius ir Bajeso parametrų įvertinius diskrečių požymių atveju.

Planuojama veikla:

Dalyvavimas ir rezultatų pristatymas tarptautinėje konferencijoje

Planuojamas mokslinių tyrimų publikavimas:

Pateikiami teoriniai rezultatai, susiję su siūlomų originalių klasifikavimo procedūrų klaidų tikimybių analitinių išraiškų ir aproksimacijų išvedimu.