

Teme za seminarski iz LSM

Blagoje Ivanović, Bojana Milošević

Decembar 2022

1. Regresioni modeli na longitudinalnim podacima

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). Applied longitudinal analysis (Vol. 998). John Wiley & Sons. Knjiga je dostupna na zahtev.

Opisati tip podataka, ilustrovati primerom, definisati linearni model na ovim podacima, i prikazati tok statističkog zaključivanja u njemu. Sve ilustrovati primerima (bar jednim van knjige).

2. Linearna i kvadratna diskriminanta analiza *

[Literatura](#)

Opisati metode i primeniti na nekim realnim podacima.

3. Hijerarhijski linearni modeli

[Podaci](#)

[Uvodna literatura](#)

[Literatura](#)

Napraviti hijerarhijski linearni model u dva nivoa u kome je zavisna promenljiva konačna ocena iz matematike (podaci student-mat.csv), a faktore i promenljivu po kojoj je vrši hijerarhija odabrati po želji.

4. Klasterovana linearna regresija *

[Link](#)

Opisati metod i primeniti na nekim realnim podacima.

5. Bajesova logistička regresija

Obavezno je uključiti samostalni primer.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). Bayesian data analysis. Chapman and Hall/CRC.

Knjiga je dostupna na zahtev.

6. Primena butstrepa u regresionim modelima

[Link 1](#)

[Link 2](#)

[Link 3](#)

Opisati metode i objasniti kada ih je potrebno primenjivati. Ilustrovati samostalnim primerima.

7. Primena permutacijskih testova u zaključivanju u regresionim modelima

[Uvodna literatura](#)

[Literatura](#)

[Literatura](#)

Predstaviti neke permutacijske testove, objasniti motiv za njihovo korišćenje i uporediti sa nekim drugim testovima koji se koriste za istu nultu hipotezu.

8. Funkcionalni linearni regresioni modeli *

Literatura: Ramsay, James O., and Bernard W. Silverman. "Functional data analysis." (2001): 5822-5828.

Opisati ove modele i primeniti ih na nekim realnim podacima.

9. Višestruko testiranje hipoteza

[Osnovna literatura](#)

Osmisliti empirijsko istraživanje kojim se porede metode korekcije. Imajte u vidu da kvalitet testa određuju njegova mera (u odnosu na nivo značajnosti) i moć.

10. Evolucija korona virusa

[Podaci](#)

Osmisliti istraživanje o faktorima koji utiču na širenje korona virusa. Dozvoljeno je da, ukoliko smatrate korisnim, izađete iz okvira kursa.

Napomena. Predlog istraživanja treba pre početka rada poslati na e-mail kurslsm@gmail.com

11. Test saglasnosti sa logističkim regresionim modelom

[Literatura](#)

Opisati testove iz rada i uporediti ih na način na koji su to uradili autori u radu (nije neophodno odabrati modele kao i autori, ali treba pratiti korake opisane u radu).

12. Uvod u kauzalno (uzročno) zaključivanje (*Causal inference*) *

Literatura:

- Poglavlja 6 i 7 iz knjige:
Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction, Guido W. Imbens, Donald B. Rubin

Objasniti značaj randomiziranih eksperimenata, zatim neki od osnovnih metoda za ocenjivanje prosečnog uticaja tretmana.

Za više detalja pisati na kurslsm@gmail.com

13. Cirkularna regresija (*Circular regression*)

Literatura:

[Link 1](#)

Jammalamadaka, S. Rao and SenGupta, A. (2001). Topics in Circular Statistics, Section 8.3, World Scientific Press, Singapore.

Knjiga je dostupna na zahtev.

Objasniti model (uz uvođenje cirkularnih promenljivih), prikazati ocenjivanje i statističko zaključivanje i to sve ilustrovati na nekom primeru sa realnim podacima.

14. Faktorska analiza

Literatura:

[Link](#)

(uz literaturu navedenu u tom master radu)

Objasniti metod detaljno i primeniti ga na realnim podacima. Poželjno je objasniti sličnosti i razlike sa PCA.

15. Analiza sentimenta iz tekstualnih podataka *

[Skup podataka](#)

Napraviti model koji na osnovu teksta klasifikuje recenzije filmova u pozitivne ili negativne. Odraditi potrebno pretprocesiranje i uporediti više metoda za vektorizaciju, npr. bag-of-words i TF-IDF. Model može biti logistička regresija (ukoliko je deo ispita ESU, potrebno je primeniti i još neki model statističkog učenja po izboru).

16. Ansambli - metode pojačavanja (boosting) *

Literatura:

[Link 1](#) (Poglavlje 10)

[Link 2](#) (poglavlje 6.2)

Opisati metode pojačavanja, Adaboost i gradijentno pojačavanje, i primeniti metode na nekom skupu podataka. Uporediti dobijene modele sa lilearnim modelima.

17. Cross-sell model *

Opisati *cross-sell* i *upsell* pojmove i na koji način se može pristupiti pravljenju modela za ciljanje kupaca.

Napraviti neki *cross-sell* model na nekim realnim podacima. Može se koristiti i ovaj [skup podataka](#) iz osiguranja.

18. Metode za odabir hiperparametara *

Literatura:

[Link 1](#)

[Link 2](#)

[Link 3](#)

Opisati metode za odabir hiperparametara i primeniti ih na realnim podacima, za model LASSO ili grebene regresije.

Napomene

Prijava tema i članova grupa za seminarski se može izvršiti putem aplikacije na adresi <http://prijava.lsm.matf.online>. Ukoliko neko želi da obradi temu koja nije na spisku, predlog teme može da pošalje na e-mail kurslsm@gmail.com.

Seminarske radove slati na e-mail kurslsm@gmail.com.

Obavezno pratiti uputstva data na <http://lsm.matf.online>

Maksimalan broj poena za pismeni deo je 15 poena. Isto važi i za odbranu. Za predlog projekta se dobija 5 poena i neophodno je predati ga pre predaje seminarskog. Odbrana se sastoji od prezentacije u trajanju do 15 min i pitanja svakom od članova grupe. Poeni za pismeni deo su jedinstveni za sve članove grupe, dok se poeni na odbrani mogu razlikovati.