

Teme za seminarski iz LSM

Bojana Milošević, Luka Perović, Blagoje Ivanović

Decembar 2023

1. 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.

2. Primena butstrepa u regresionim modelima

[Link 1](#)

[Link 2](#)

[Link 3](#)

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

3. Pravičnost u modeliranju*

Objasniti problem, navesti nekoliko definicija i ilustrovati ih na samostalnoj empirijskoj studiji (ili replicirati neke od rezultata iz literature). Posebno je važno objasniti značaj u nekim oblastima (npr u osiguranju).

[Literatura](#)

4. 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.

5. 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.

6. Lokalna polinomijalna regresija

[Literatura](#)

[Literatura](#)

Predstaviti Nadaraja–Votson regresiju i još neki metod lokalne polinomijalne regresije (npr. loess), primeniti ih na nekim podacima i uporediti medjusobno, kao i sa običnom polinomijalnom regresijom.

7. Primena uopštenih linearnih modela u neživotnom osiguranju

Literatura: Wüthrich, Mario V., and Michael Merz. Statistical foundations of actuarial learning and its applications. Springer Nature, 2023. (dostupna na zahtev)

Opisati nekoliko modela (npr Puasonovu regresiju) i kako se mogu primeniti u neživotnom osiguranju i primeniti ih na realnim podacima. Poželjno je nekoristiti isti skup podataka kao u knjizi.

8. 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).

9. 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

10. 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.

11. 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.

12. 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).

13. 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 learnim modelima.

14. 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

- *Obavezno pročitati sve stavke i napomene na ovoj stranici i poštovati data uputstva i pravila*
- Teme su okvirno za 2-3 osobe (ne više od 4). Moguće je raditi i u manjim i većim grupama. Ako se radi o grupi o 4 osobe projekat treba da bude sadržajniji nego što bi bio da ga je radila grupa od 3 osobe.
- Prijavljivanje se vrši na prijava.lsm.matf.online
- Rokovi za predaju:
 - Rok za predaju predloga projekta:
21.12.2023. u 23:59.
 - Prvi rok za predaju radova:
08.01.2024. u 10:00
Odbrana radova će biti 09.01.2023.

– Drugi rok za predaju radova:

- Rasporedi odbrana će biti naknadno objavljeni

Administracija

- Sva (pisana) komunikacija u vezi sa seminarskim se vrši putem mejla: `kurslsm@gmail.com`. To uključuje i pitanja za seminarski (koja su uvek dobrodošla) i prijave samostalnih tema i ostale informacije, da bismo ubrzali komunikaciju i mogli zajedno da vidimo mejl. Prilikom slanja pitanja, u “Subject” obavezno uključiti i naslov teme.
- Kolege koje su prošle godine radile seminarski, nisu u obavezi da ga ponovo rade, već im se može preneti broj bodova sa prošlogodišnjeg seminarskog, samo neka se jave na gorepomenuti mejl.
- Kolege koje polažu ceo ispit iz Elemenata statističkog učenja (dakle, nisu prethodno položili LSM), moraju odabrati temu sa zvezdicom.
- Za kolege koje su ranije položili LSM i polažu samo razliku za ESU, važi dogovor od ranije.

Izrada radova

- Literatura navedena uz temu nije sveobuhvatna i ne treba se ograničavati na nju. Ona je data kao jedna polazna tačka za istraživanje i upoznavanje sa temom. Ohrabrujemo vas da nadete svoje izvore i proširite (ili zao-
bidete) datu literaturu, u cilju maksimizovanja kvaliteta konačnog rada. Za sve nedoumice nas možete pitati.
- Na kraju rada je obavezno da postoji bibliografija korišćene literature (pomoć: \LaTeX). Kod nikako ne sme biti u vidu slike ni na jednom mestu, jer onda ne možemo pokrenuti isti.

Predaja radova

- Radovi se predaju *isključivo u pdf formatu, napravljenom u LaTeX-u*. Neophodno je poslati i prezentaciju u obliku slajdova (PDF, PowerPoint, Prezi i sl.). Prezentacija se može malo korigovati do odbrane. Prilikom slanja rada, u Subject obavezno staviti naslov teme. Kod treba predati zajedno sa radom kao R fajl (sredite ga prethodno, treba da bude lepo formatiran i iskomentarisan). Kod se može ubaciti kao deo dokumenta, ali samo ako su to kratki isečci koji imaju svrhu, i sa smislenom veličinom izlaza. Nikako ne ostavljati u radu kod koji obuhvata više strana dokumenta ili čiji izlaz obuhvata nekoliko strana. Rad prevažno mora biti čitljiv.

Odbrana

- Pripremiti prezentaciju rada zajednički, u trajanju do 15min. Uvežbati je zajedno, kako bi vreme zaista bilo ispoštovano.
- Nakon prezentacije svakom članu će biti postavljena neka pitanja u vezi sa radom. Svi članovi treba da budu upoznati sa svim delovima rada.
- Odbrane će biti održane online i neophodno je da prisustvuju svi članovi timova, kao i svi timovi koji brane seminarski u odgovarajućoj grupi.

Poeni

- 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
- Broj poena će biti saopšten nakon što budu održane sve odbrane.
- Broj poena za rad je isti za sve članove tima dok se za odbranu može dobiti različit broj poena.