

Teme za seminarski iz LSM

Blagoje Ivanović, Bojana Milošević

Novembar 2020

1. Primena regresionih modela u saobraćaju

[Podaci](#)

[Opis podataka](#)

Napraviti najbolji model kojim bi se prikazao uticaj različitih faktora na broj povređenih ili nastradalih u nesrecama u Kaliforniji. Razmotriti dvodimenzioni Puasonov regresioni model za broj povređenih i broj umrlih, kao i neki model koji objedinjeno posmatra umrle i povređene. **(3 osobe)**

2. Primena linearnih modela u medicini

[Podaci](#)

U R paketu NHANES su dostupni podaci o 75 različitih zdravstvenih i demografskih promenljivih. Odabрати neku zavisnu promenljivu i prediktore i sprovesti istraživanje uticaja prediktora na zavisnu promenljivu. **(3 osobe)**

Napomena: Za ovu temu se može prijaviti više grupa (najviše tri tročlane), ako su predložena istraživanja dovoljno različita. Predloge slati na e-mail kurslsm@gmail.com.

3. Detekcija autlajera

Predstaviti metode iz rada (i primeniti na nekim realnim podacima)

- [Penalized Weighted Least Squares for Outlier Detection and Robust Regression \(link\)](#) **(3 osobe)**

4. 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.

5. Neuronske mreže (višeslojni perceptroni)

[Dodatna literatura](#)

Opisati osnovnu strukturu i funkcionisanje neuronskih mreža, postupak propagacije unazad i način na koji se mreža trenira.

Napraviti ručnu implementaciju neuronske mreže ([pomoćna literatura](#)). Modifikovati algoritam da može da prihvati 2 skrivena sloja, kao i softmax aktivacionu funkciju na izlaznom sloju i napraviti mrežu za klasifikaciju ručno pisanih cifara iz skupa podataka MNIST ([link](#)), sa što većom tačnošću. **(3-4 osobe)**

6. Evolucija korona virusa

Podaci

Osmisliti istraživanje o faktorima koji utiču na širenje korona virusa. Dozvoljeno je da, ukoliko smatrate korisnim, izađete iz okvira kursa. **(3 osobe)**

Napomena: Za ovu temu se može prijaviti više grupa (najviše tri tročlane), ako su predložena istraživanja dovoljno različita. Predloge slati na e-mail kurslsm@gmail.com.

7. Rešavanje problema heteroskedastičnosti

Literatura

Predstaviti metode iz rada i primeniti na realnim podacima. **(2-3 osobe)**

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). **(3 osobe)**

9. Imputacija nedostajućih podataka primenom regresionih modela

Osnovna literatura

Predstaviti metode iz rada i uporediti ih sa klasičnim metodama kao što su imputacija srednjom vrednošću ili uzoračkom medijanom. **(3 osobe)**

10. 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. **(3 osobe)**

Za više detalja pisati na kurslsm@gmail.com

11. Gama regresija

Osnovna literatura

Opisati model, prikazati kako se ocenjuju njegovi parametri i predstaviti rezidualne i njihove osobine. Sve ilustrovati na realnim podacima. **(3 osobe)**

12. Regresioni modeli preživljavanja

Osnovna literatura

Predstaviti neki regresioni model preživljavanja, na primer AFT i primeniti ga na realnim podacima. **(3 osobe)**

13. Retki (*sparse*) regresioni modeli

(Slobodna tema) Predložiti temu u vezi sa retkim (*sparse*) regresionim modelima

Osnovne informacije mogu se naći u [ovom radu](#) **(2-3 osobe)**

14. Rad sa nebalansiranim podacima

Pomoćna literatura

Pomoćna literatura

Opisati značaj problema neizbalansiranosti klasa u podacima, predstaviti često korišćene metode za rešavanje istog (npr. *under-sampling*, *over-sampling*, *SMOTE*, *ansambli*) i primeniti ih na nekim realnim podacima. **(2-4 osobe)**

Napomena. Kao osnovni klasifikator koristiti logističku regresiju.

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. 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.