

Семинарски рад из курса Увод у теорију узорака

Стефан Малбашић

10. април 2023. године

1 Списак тема

1. Истраживачки рад о начинима узорковања и оцењивања.

Потребно је одабрати базу података која ће представљати популацију. Сваки студент мора користити другачију базу података. У уводу је потребно дати кратак опис истраживања односно проблематике којом ћете се бавити у истраживању, а затим дефинисати циљ истраживања и питања на која ћете у овом раду покушати да одговорите. Такође, треба издвојити променљиве које представљају обележја од интереса и анализирати их (преко сумарних статистика, табличним приказом, визуелизацијом података и сл.). У централном делу рада треба илустровати различите планове узорковања и техника оцењивања и међусобно их упоредити (у смислу квалитета оцена). На крају, у закључку, треба интерпретирати резултате и изнети своје закључке о различитим начинима узорковања и оцењивања.

2. Бутстреп метода узорковања.

Описати Bootstrap методу и илустровати је на примерима. Bootstrap метода подразумева коришћење рачунара и омогућава добијање резултата везаних за тачкасте и интервалне оцене параметара расподеле посматраног обележја. Метода може да се користи и за тестирање статистичких хипотеза, што је такође потребно показати у раду.

3. Подузорковање и Џекнајф метод.

Постоји неколико метода за оцењивање параметара које су засноване на подузорковању, а једна од најпознатијих је џекнајф метода. Описати на чему се заснива ова метода и која је основна идеја. Такође, потребно је приказати рачунарску имплементацију ове методе и исту илустровати на одређеној бази података.

4. Нацрти невероватносног узорковања.

На курсу су приказани нацрти узорковања који се заснивају на теорији вероватноћа, због чега су оцене параметара основног скупа на основу статистика непристрасне и постоји могућност одређивања грешке узорка. Описати нацрте невероватносног узорковања (квотни узорак, пригодни узорак, намерни узорак...). Навести ситуације у којима се користе ови планови узорковања и упоредити их са вероватносним. Примену илустровати на реалним подацима.

5. Методе реузорковања: унакрсна валидација и к-фолд валидација.

Методе реузорковања су један од основних алата у модерној статистици. Описати принцип унакрсне валидације и к-фолд валидације, као и када се која метода користи. Упоредити пристрасност и дисперзију оцена добијених овим методама. Кросвалидацију и класификационе проблеме демонстрирати на реалним подацима.

2 Упутство

Пријава тема за семинарски рад се може извршити путем мејла: stefan.malbasic[at]matf.bg.ac.rs.

Одрађен семинарски рад заједно са одбраном носи максималан **30 поена** и изарада се ради самостално. Одбрана се састоји од презентације у последњем термину часова вежби **22. мај 2023. године**. Уколико буде више пријављених кандидата за исту тему, предност ће имати кандидат који је први послао пријаву. Тему 1. може радити више кандидата, али све база на којој се врши истраживање мора разликовати за сваког кандидата.

На крају рада је неопходно да постоји библиографија коришћене литературе. Код никако не сме бити у виду слике ни на једном месту, јер онда не можемо покренути исти. Радови се предају у .pdf формату, направљеном у L^AT_EX или Rmarkdown. Неопходно је послати и презентацију у облику слајдова (PDF, PowerPoint, Prezi и сл). Приликом слања рада, у Subject обавезно ставити наслов теме.

Код треба предати заједно са радом као R фајл (средите га претходно, треба да буде лепо форматиран и искоментарисан). Код се може убацити као део документа, али само ако су то кратки исечци који имају сврху и са смисленом величином излаза. Никако не остављати у раду код који обухвата више страна документа или чији излаз обухвата неколико страна. Рад превашодно мора бити читљив.

3 Рокови за предају

- Рок за пријаву теме: 20. април 2023. у 23:59 часова.
- Рок за предају семинарског рада: 19. мај 2023. у 23:59 часова.