

LTAT.02.006 Andmeteaduse meetodid

9. praktikum

10. november 2020

1. Mündiviske tulemus ära arvamine.

Vaatleme järgmist mängu:

- Su sõbral on 5 münti, mis näevad välja samasugused. Aga tegelikult on neil kõigil erinev tõenäosus saada kull, täpsemalt on need tõenäosused viiel mündil vastavalt: 0.00, 0.25, 0.50, 0.75, 1.00. Sõber valib ühtlaselt juhuslikult ühe nendest müntidest ning viskab seda münti. Sa näed tulemust (kull või kiri) ja pead ära arvama, et mis on sama mündiga tehtud järgmise mündiviske tulemus (ikka veel teadmata milline viiest mündist see on).

Sa mängid seda mängu oma sõbraga 1000 korda, iga kord samade 5 mündiga. Tähistame igas mängus esimese mündiviske tulemus vastavalt $y_1, \dots, y_{1000} \in \{0, 1\}$ ning teise viske tulemus (see mida ennustama peab) $y'_1, \dots, y'_{1000} \in \{0, 1\}$, kus 1 tähistab kulli ja 0 kirja. Ennustamisel saab pakkuda suvalisi reaalarve vahemikust $[0, 1]$, tähistame ennustused $\hat{y}_1, \dots, \hat{y}_{1000}$. Kui 1000 mängu on mängitud, siis hinnatakse ennustusi mõõduga MSE (mean squared error) ehk keskmise ruutveaga:

$$\text{MSE} = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{y}_i - y'_i)^2$$

Kui $\text{MSE} < 0.1$, siis võidab mängu Sina, vastasel juhul võidab sõber.

Selle ülesande eesmärk on leida optimaalne strateegia.

(a) Vaatleme strateegiat, mille järgi ennustus on sama, mis esimese mündiviske tulemus ehk $\hat{y}_i = y_i$. Arvuta oodatav MSE (ehk MSE keskväärtus) selle strateegia puhul.

(b) Vaatleme strateegiat, kus ennustus on alati 0.5 ehk $\hat{y}_i = 0.5$. Arvuta oodatav MSE.

(c) Nüüd leia strateegia, mis võiks ülaltoodutest parem olla. Selleks tuleb otsustada, et mis väärtust \hat{y}_i ennustada $y_i = 1$ korral ning mis väärtust $y_i = 0$ korral. Arvuta oodatav MSE.

(d) Vaatleme modifitseeritud mängu, kus 5 münti asemel on sõbral kõikvõimalike kulli tõenäosustega mündid ning ta valib nende seast ühtlaselt juhuslikult. Teisisõnu, sõber võtab ühtlaselt juhuslikult vahemikust $[0, 1]$ arvu p ning seejärel viskab münti, millel on kulli tõenäosus p . Leia optimaalne strateegia (MSE-d pole vaja arvutada).

(e) Vaatleme punktis (d) toodud mängu modifikatsiooni, kus igas 1000-st alam mängust visatakse münti mitte 1 kord vaid k korda ning on vaja ennustada tulemus neile järgneval korral. Leia optimaalne strateegia (MSE-d pole vaja arvutada).