

Spikker

- kommentaarid

Sinine - asjad, mille õigsust peaks kontrollima

Miks ma käesolevat asja teen?

Panen kirja mõned põhjused, miks ma tegelen selle asjaga:

- I Sellest on kellegile kasu - loodan realselt, et saan hakkama mingi toreda asjaga, millest keegi (näiteks Marguse psühholoogidest sõbrad) kasu saab.
- II Saab kätte selle paberi, mis teeb minust parema inimese.
- III Midagi uut, olen juba päris pikalt tarkust ühes formaadis kuula õppejõudu, tööta läbi tema poolt valitud materjalid, esita see õppejõule, äkki selline formaat, kus tuleb ise otsida ja ise mõelda meeldib.
- IV Väljakutse, kuigi viimaste aastatega on enesemotivatsioon kõvasti paranenud, ei ole see veel seal, kus ta olla võiks. Loodetavasti "treenin" seda aspekti.

Mida me teen?

Meil on valikvastutusega küsimustik milles on n küsimust, kõik küsimused 5 palli skaalal. Vaja oleks mudelit, mille põhjal hinnata vastust ühele küsimusele, kui on teada vastused $n - 1$ küsimusele. Mudel oleks kasutatav näiteks vastuste kvaliteedi hindamisel. Näiteks, juhul kui küsimustiku viimasele küsimusele pakutud vastus on mudeli põhjal äärmiselt ebatõenäoline, võime eeldada, et tegemist on erindiga (vastaja "väsis" ja viimane küsimusele vastati kõige esimese variandiga).

Olemas on mitmeid erinevaid lahendusi. Margus pakkus välja järgmise mõtte :

Kolmas mõte oleks aga lineariseerida muutuja valides kordajad "optimaalselt" nii, et kokku oleks (lineaarsete s.o. pearsoni) korellatsioonide summa maksimaalne.

Minu ülesanne: analüüsida, kas see lahendus idee on mõistlik ja kui on mõistlik, siis välja mõelda, kuidas see lahendus täpselt võiks välja näha.

Alteervatiivid

1. Kõige triviaalsem viis: kõikide küsimuste vastused kujutama hulgale $\{1,2,3,4,5\}$, leiame mudeli, mille võime kirjeldada valitud küsimust on suurim. Tegemist on ilmselt vaikimisi variandiga ehk loodud mudelit peab võrdlema
2. Selle asemel, et kujutada hulgale $1,2,3,4,5$, leiame sobivad vastused nii, et mudeli kirjeldav jõud oleks suurim. Oht selles, et meie mudel kirjeldab väga hästi olemasolevat valimit, ütle mata suurt midagi üldkogumi kohta. Huvitav oleks, kas lihsalt nii midagi teha ei annaks ? Overfittimise vastu saaks, aga vaja oleks valimit, mis oleks piisavalt suur, et seda kaheks jagada (midagi, mille pealt mudelit ehitada ja midagi, mille pealt seda validifitseerida.
3. Midagi veel? Peab uurima.

Marguse idee lahtikirjutatuna

Meil on muutujad $K_1, K_2, \dots, K_n, K_i \in \{1, 2, 3, 4, 5\}$. Teisendame need muutujateks $L_1, L_2, L_3, \dots, L_n$, tuues sisse väärtused $\{a_i, b_i, c_i, d_i, e_i\}$ $i \in 1, 2, \dots, n$, nii et kehtib järgnev : $K_i = 1 \implies L_i = a_i, K_i = 2 \implies L_i = b_i, \dots, K_i = 5 \implies L_i = e_i$. Lisame kitsendused:

1. $E(L_i) = 0$, st $p_{ia} * a_i + p_{ib} * b_i + p_{ic} * c_i + p_{id} * d_i + p_{ie} * e$
2. $D(L_i) = 1$, st $p_{ia} * (a_i)^2 + p_{ib} * (b_i)^2 + p_{ic} * (c_i)^2 + p_{id} * (d_i)^2 + p_{ie} * (e_i) = 1$

Olgu meil P ($n \times n$ maatriks, aga missugune ?) ja x :

$$P = Pr[K_{i/5} = (i \bmod 5) + 1 \ \& \ K_{j/5} = (j \bmod 5) + 1]_{i,j}$$

(Ei oska seda kuidagi mõistlikult lugeda, otse kirjast kopeeritud)

$$x = (a_1, b_1, c_1, d_1, e_1, a_2, b_2, c_2, d_2, e_2, \dots, a_n, b_n, c_n, d_n, e_n)$$

Siis $x^T P x = \sum_{i=1}^n \sum_{j=1}^n COV(L_i, L_j)$ (Kirja põhjal $x^T P x$, aga kui P ikkagi on maatriks, siis peaks vast ikka x^T olema ?)

Selle põhjal saame moodustada *Quadratically constrained quadratic programm*-i, kus *objective function* (Eesti keels kulufunktsioon?) on $\max x^T P x$ ja piiravateks funktsioonideks on funktsioonid $R_i^T x = 0, i \in 1, 2, \dots, n$, kus $R_i = (\underbrace{0, 0, \dots, 0}_{(i-1)*5}, p_{ia}, p_{ib}, p_{ic}, p_{id}, p_{ie}, \underbrace{0, 0, \dots, 0}_{(n-1)*5})$ ja funktsioonid $x^T P_i x = 1, i \in 1, 2, \dots, n$, kus

$$P_i = \begin{pmatrix} p_{ja} & 0 & \dots & 0 \\ 0 & p_{jb} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_{je} \end{pmatrix}, \quad p_j \alpha = \begin{cases} 0 & L_j \neq i \\ p_i \alpha & i = j \geq 0 \end{cases}, \alpha \in \{a, b, c, d, e\}$$