

ETTEKANDE TEEMAD

Alternatiivina toodud ülesannete lahendamisele võib kursuse viimasel (või eelviimasel) nädalal loengu ajal rääkida lühidalt ka mõnel allpool toodud teemal (et äkki kellelgi haakub oma huvidega, töödega, uurimistöö temaatikaga, on lihtsalt soov õppejõu elu lihtsamaks teha, ...)

Kuigi ilmselt jätkuks mitmete teemade puhul juttu terveks loengukursuseks (või mitmekski), võiks antud aines arvestuse sooritatud saamiseks rääkida siiski vaid ~20-30 minutit.

Järgnevalt nimekiri mõningatest teemadest, mille kõigi puhul tuleks avada meetodi-algoritmi matemaatiline taust (midagi tõestama ja tuletama ei ole vaja hakata, seda nii lühikese ettekande jooksul ilmselt ei jõuakski) ning tuua üks-kaks näidet meetodi-algoritmi reaalsest (või hüpoteetilisest) kasutamisest (soovitavalt geneetikaga haakuvates uuringutes-artiklites).

- χ^2 -testi võimsus ja valimi maht seoses šansside suhtega (või suhtelise riskiga assotsiatsioonanalüüsi näitel).
- Permutatsioonitestid ja Monte Carlo meetodid.
- Mitmene testimine (Bonferroni meetod, Bonferroni-Holmi meetod, FDR, p- ja q-väärtus).
- Juhumets (*random forest*) ja võimendatud regressioonipuud (*boosted regression trees*).
- Hierarhiline ja k-keskmiste meetodil klasterdamine (algoritmide olemus ja rakendamine näiteks liikide sarnasuse-põlvnemise-rände uurimisel (näiteks mõni Mait Metspalu töödest) või liikide erinevuse hindamisel (näiteks eesti maalamba uuringud vmt) ning tulemuste esitamine/illustreerimine).
- Veel mõnedest geneetikalastest (sh erinevate -oomikate alastes) uuringutes (potentsiaalselt rakendatavatest mitmemõõtmelise statistika (ja samaaegselt ka masinõppe) meetoditest-algoritmidest (nimekiri on siin pikk).
- Indiviididevaheline sugulus (erinevad suguluskoefitsiendid).
- Populatsioonide sisene ja vaheline geneetiline varieeruvus, selle mõõtmine (vt näiteks <http://www.loujost.com/Statistics%20and%20Physics/Diversity%20and%20Similarity/PopGen.htm>)
- Geeni ekspressioon (mida enesest kujutab ja mõni näide sellest, kuidas ja mida ja miks analüüsida).
- Geenianadmehete imiteerimine ja/või haplotüüpide ja haplotüübiblokkide hindamine (näiteks varjatud Markovi mudelitega).
- ...

Erinevad andmeanalüüsi tulemuste visualiseerimise viisid – mis tüüpi analüüsides ja andmete puhul kasutatakse, kuidas konstrueeritakse (mis arvutustel põhinevad) – lisaks tuleks tuua üks-kaks konkreetset näidet teadusartikleist (koos artiklis käsitletava probleemi lühikirjeldusega ja selgitusega, kuidas kasutatud joonis võimaldab probleemi lahendada / vastuseid leida).

Et nende jooniste puhul ilmselt väga pikalt rääkida pole, grupeerisin osad teemad paarikaupa.

- Vulkaanidiagramm (*volcano plot*) (ja *S-plot*), kolmnurkdiagramm (*triangular plot*, *ternary plot*). *Forest plot*.
- *Heatmap*. *Bi-plot*.
- Andmete ja analüüsitulemuste tsirkulaarne esitamine (näiteks paketi *Circos* abil).

Kuna kaks viimast teemat haakuvad klasterdamise jm mitmemõõtmelise statistika ja masinõppe meetoditega, siis võib tekitada ka teemasid nagu näiteks 'hierarhiline klasterdamine ja *heatmap*', 'peakomponentanalüüs ja *bi-plot*', 'koinertsusanalüüs ja *bi-plot*', 'gruppidevaheline peakomponentanalüüs ja *bi-plot*', 'klasterdamise tulemuste esitamine programmi *Circos* abil' jne.

Muidugi on ettekannete teemadena mõeldavad ka mistahes muud geneetikas ja -oomikates rakendatavad matemaatika-statistika-arvutiteaduse meetodid, mida mina eraldi välja tuua ei osanud. Samas võib ettekande aluseks olla ka mõni konkreetne meditsiinilise, ökoloogilise, ajaloo-alase jne geneetika (vm -oomika) uuring/artikkel, mille puhul tuleks anda ülevaade nii probleemi olemusest kui ka selle lahendamiseks kogutud andmetest, nende analüüsimise meetoditest ja tulemuste esitamisest.