

Bakalaureusetöö teemade pakkumised

Evely Kirsiaed, arvutusmeetodite lektor
evely.kirsiaed@ut.ee

September 2026

1. FISTA meetodi praktilised täiustused hõredas regressioonis

Töös uuritakse klassikalist ISTA (*iterative shrinkage-thresholding*) algoritmi ja selle kiirendatud varianti FISTA (*fast iterative shrinkage-thresholding algorithm*) Lasso-tüüpi ülesande

$$\min_x \|Ax - b\|_2^2 + \lambda \|x\|_1$$

lahendamisel (sihifunktsiooniks on sileda kaofunktsiooni ja mittesileda trahvifunktsiooni summa). Lasso-tüüpi ülesanne (*LASSO – Least Absolute Shrinkage and Selection Operator*) on masinõppes ja statistikas kasutatav regressioonanalüüsi meetod, mis tegeleb üheaegselt nii mudeli parameetrite hindamise kui ka tunnuste valikuga. Lassol on unikaalne omadus suruda vähem oluliste tunnuste kordajad nulliks, mis hoiab ära mudeli ülesobitamise, nn. *overfitting*’u.

Töö võimalik struktuur: esitada meetodi kirjeldus ja edasiarendused koos koonduvustulemustega, realiseerida ISTA, klassikaline FISTA ja adaptiivse taaskäivitamisega FISTA ning võrrelda neid Lasso-tüüpi ülesannetel (nii sünteetiliste kui pärisandmete korral).

Võimalikud lähteartiklid:

- [1] Beck, A., Teboulle, M. (2009). A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, **2**(1), 183–202. (FISTA alusartikkel)
- [2] Chambolle, A., Dossal, C. (2015). On the Convergence of the Iterates of the ‘Fast Iterative Shrinkage/Thresholding Algorithm’. *Journal of Optimization Theory and Applications*, **166**, 968–982. (FISTA koondumine + modifikatsioon)
- [3] Kim, D., Fessler, J. A. (2018). Another look at the fast iterative shrinkage/thresholding algorithm (FISTA). *SIAM Journal on Optimization*, **28**(1), 223–250. (FISTA modifikatsioon)
- [4] O’Donoghue, B., Candès, E. (2015). Adaptive Restart for Accelerated Gradient Schemes. *Foundations of Computational Mathematics*, **15**, 715–732. (adaptiivne taaskäivõtamine)

2. Koordinaatlaskumise koordinaadi valikureeglite võrdlus

Koordinaatlaskumise meetodis muudetakse ühe iteratsiooni jooksul ainult üht või väikest hulka muutujaid. Töö keskne küsimus võiks olla, kuidas mõjutab koordinaadi valik algoritmi kiirust.

Võrrelda võiks näiteks tsüklilist valikut, juhuslikku valikut, suurima gradiendikomponendi ehk Gauss–Southwelli reeglit, gradiendi suurusel põhinevat tõenäosuslikku valikut. Gauss–Southwelli valik võib vähendada iteratsioonide arvu, kuid koordinaadi leidmine ise on kallim, juhuslik valik on odavam ja selle kohta on olemas põhjalik koonduvusteooria. Iseseisva täiendusena saab lisada mõne hübriidreegli.

Meetodeid võiks katsetada ruutfunktsionaali minimiseerimisel, LASSO ehk hõreda lineaarse regressiooni, logistilise regressiooni vmt korral.

Võimalikud lähteartiklid:

- [1] Nesterov, Y. (2012). Efficiency of Coordinate Descent Methods on Huge-Scale Optimization Problems. *SIAM Journal on Optimization*, **22(2)**, 341–362. (meetodi alusartikkel)
- [2] Nutini, J., Schmidt, M., Laradji, I., Friedlander, M., Koepke, H. (2015). Coordinate Descent Converges Faster with the Gauss–Southwell Rule Than Random Selection. *Proceedings of the 32nd International Conference on Machine Learning*, PMLR 37, 1632–1641. (juhusliku valiku võrdlus GS reegluga)
- [3] Richtárik, P., Takáč, M. (2014). Iteration Complexity of Randomized Block-Coordinate Descent Methods for Minimizing a Composite Function. *Mathematical Programming*, **144**, 1–38. (meetodi laiendus)

3. Adaptiivsete gradientmeetodite stabiilsus ja sammuvahetus

Teema keskenduks masinõppes kasutatavatele SGD, AdaGrad, Adam ja AMSGrad meetoditele. Eesmärk ei ole suure närvivõrgu treenimine: algoritme saab analüüsida hästi valitud kumeratel ja mittekumeratel testfunktsioonidel ning väikesel regressiooni- või klassifitseerimisülesandel.

Töös võiks uurida:

- millal Adam võngub või koondub halvasti;
- kuidas mõjutavad tulemust sammu pikkus ja momendiparameetrid;
- kas AMSGrad või jooneotsinguga AdaGrad on stabiilsem;
- kuivõrd vastavad katsed artiklites esitatud koonduvustulemustele.

Kirjanduses on toodud nii Adami probleemseid näiteid kui ka tingimusi, mille korral Adam-tüüpi meetodid koonduvad. Reprodutseerida võiks Reddi jt kontranäite. Meetodeid SGD, AdaGrad, Adam ja AMSGrad võiks võrrelda ning iseseisva täiendusena realiseerida automaatse sammuvahetuse või algoritmi, mis vahetab ebastabiilsuse tuvastamisel Adami ja SGD vahel.

Võimalikud lähteartiklid:

- [1] Kingma, D. P., Ba, J. (2015). Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations (ICLR)*. (Adam alusartikkel)
- [2] Reddi, S. J., Kale, S., Kumar, S. (2018). On the Convergence of Adam and Beyond. *6th International Conference on Learning Representations (ICLR)*. (Näited, kus Adam ei koonu; parandatud AMSGrad algoritm)
- [3] Zhou, D., Chen, J., Cao, Y., Tang, Y., Yang, Z., Gu, Q. (2020). On the Convergence of Adaptive Gradient Methods for Nonconvex Optimization. *Proceedings of the 12th Annual Workshop on Optimization for Machine Learning*. (AdaGrad ja AMSGrad koonduvusanalüüs)