

Bioinformatika a strojové učenie

V školskom roku 2019/20 nebudú súčasťou štátnic články z odbornej literatúry, iba okruhy učiva uvedené nižšie.

Okruhy učiva, v zátvorke skratky súvisiacich predmetov:

AOP: Aproximácia optimalizačných problémov

IDZ: Integrácia dátových zdrojov

MBI: Metódy v bioinformatike

NS: Neurónové siete

PaŠ: Pravdepodobnosť a štatistika

SU: Strojové učenie

VPDŠ: Vybrané partie z dátových štruktúr

1. Neurónové siete: viacvrstvový perceptrón, metóda spätného šírenia chyby, hlboké architektúry neurónových sietí (SU,NS)
2. Modelovanie sekvenčných dát: Skryté Markovove modely, podmienená pravdepodobnosť a Bayesove vety, Viterbiho a dopredný algoritmus, príklady využitia v bioinformatike (hľadanie génov a profilové HMM), rekurentné neurónové siete (MBI,PaŠ,NS)
3. Klasifikačné modely: support vector machines, rozhodovacie stromy, náhodné lesy, bagging, boosting (SU)
4. Regresia: lineárna a generalizovaná lineárna regresia, metóda najmenších štvorcov, štatistický model s normálnym rozdelením chýb, regularizácia (PaŠ,SU)
5. Teória strojového učenia: štatistický model strojového učenia, výchyľka vs. rozptyl, preučenie a podučenie, PAC učenie, odhady pomocou VC dimenzie (SU,NS)
6. Strojové učenie bez učiteľa: zhľukovanie, samoorganizujúce sa zobrazenia, analýza hlavných komponentov, využitie na analýzu génovej expresie (SU,NS,MBI)
7. Testovanie štatistických hypotéz: Fisherov exaktný test, Welchov t-test, Mann-Whitneyho U-test, Bonferroniho korekcia viacnásobného testovania, log likelihood ratio test, príklady použitia testov v bioformatike (PaŠ,IDZ,MBI)
8. Stredná hodnota náhodnej premennej: linearita strednej hodnoty, Markovova a Čebyševova nerovnosť (PaŠ)
9. Limitné vety teórie pravdepodobnosti: centrálna limitná veta, Moivrova-Laplaceova veta, slabý zákon veľkých čísel (PaŠ)
10. Sekvenovanie DNA: technológie sekvenovania a ich charakteristiky (Sanger, Illumina, nanopórové sekvenovanie), skladanie genómov, deBruijnove grafy, RNA-seq (MBI)
11. Fylogenetika a komparatívna genomika: metóda spájania susedov, metóda úspornosti, Jukes-Cantorov model a iné substitučné modely, pozitívna a negatívna selekcia a jej vplyv na evolúciu biologických sekvencií (MBI)
12. Zarovnania a algoritmy na reťazcoch: lokálne a globálne zarovnávanie sekvencií, BLAST (jadrá zarovnaní), perfektné hešovanie, Bloomov filter, efektívna reprezentácia

sekvencií (sufixové stromy a polia, Burrowsova–Wheelerova transformácia, FM index) (MBI,VPDŠ)

13. Metóda maximálnej vierohodnosti: odhad parametrov rozdelenia, nevychýlené odhady parametrov, metóda maximálnej vierohodnosti na rekonštrukciu fylogenetických stromov, Felsensteinov algoritmus, EM algoritmus, tréning skrytých Markovových modelov, hľadanie sekvenčných motívov (PaŠ, MBI)
14. Lineárne programovanie: lineárne a kvadratické programovanie, simplexová metóda, dualita, celočíselné lineárne programovanie a jeho využitie na riešenie ťažkých problémov v bioinformatike, využitie lineárneho programovania v aproximačných algoritmoch (deterministické zaokrúhľovanie, iterované zaokrúhľovanie, randomizované zaokrúhľovanie + derandomizácia, primárno-duálne metódy), semidefinitné programovanie a max-cut, využitie duality v support vector machines (kernelové metódy) (AOP, SU, MBI)
15. Aproximovateľnosť: Zložitosť triedy aproximačných algoritmov, PCP veta a jej použitie, AP-redukcia, APX úplné problémy, aproximovateľnosť problému obchodného cestujúceho, polynomiálne aproximačné schémy a príklady PTAS algoritmov (AOP)
16. Aplikácie formálnych jazykov: Knuth-Morris-Pratt algoritmus na hľadanie vzorky v texte, stochastické bezkontextové gramatiky, kovariačný model a rodiny RNA, Nussinovovej algoritmus (MBI, VPDŠ)
17. Modely dátových štruktúr: amortizovaná zložitosť a potenciálová funkcia, I/O model a B-stromy, cache-oblivious model a statický binárny strom s van Emde Boas rozložením, úsporné dátové štruktúry (rank a select), streaming model (VPDŠ)
18. Dátové štruktúry pre intervaly: range minimum query, lowest common ancestor, segmentové stromy, rozsahové stromy (VPDŠ)