

## Osnove strojnog učenja

### Popis pitanja za usmeni ispit, ak. god. 2022./2023.

1. Što je strojno učenje? Koja je razlika između eksplicitnog programiranja računala i strojnog učenja? Navedite kakvi problemi se rješavaju pojedinim pristupom (primjeri).
2. Navedite tipove strojnog učenja te za svaki tip minimalno dva primjera.
3. Nacrtajte tipičan redoslijed radnji kod primjene algoritama strojnog učenja u prediktivnom modeliranju. Objasnite pojedini korak.
4. Objasnite tabličnu organizaciju podatkovnog skupa i način zapisivanja podatkovnog skupa u matričnoj notaciji.
5. Što je eksplorativna analiza podataka (EDA)? Koji grafički prikazi se koriste prilikom EDA? Objasnite ih.
6. Objasnite osnovne postupke u predobradi podataka (podatkovnog skupa).
7. Objasnite kakve su to kategoričke veličine i načine njihovog kodiranja. Navedite primjere.
8. Objasnite načine skaliranja numeričkih veličina.
9. Objasnite podjelu podatkovnog skupa na dva skupa. Koja je uloga svakog skupa podataka?
10. Objasnite nadzirano učenje i princip označavanja dostupnog skupa podataka. Koja dva glavna tipa problema postoje u okviru nadziranog učenja?
11. Objasnite jednostavnu linearnu regresiju. Objasnite kriterijsku funkciju.
12. Objasnite načine procjene parametara linearnog regresijskog modela. Navedite prednosti nedostatke pojedinog pristupa.
13. Što je duljina koraka? Kako utječe na minimizaciju kriterijske funkcije? Objasnite na primjeru modela s jednim parametrom.
14. Objasnite višedimenzionalnu linearnu regresiju. Objasnite kriterijsku funkciju.
15. Objasnite načine procjene parametara višedimenzionalnog regresijskog modela.
16. Objasnite razliku između batch metode gradijentnog spusta i stohastičke metode gradijentnog spusta. Napišite algoritam podešavanja parametara linearnog regresijskog modela stohastičkom gradijentnom metodom. Što je mini-batch stohastička metoda gradijentnog spusta?
17. Navedite metrike za vrednovanje regresijskih modela.
18. Objasnite polinomsku regresiju. Kako se procjenjuju parametri ovakvog modela? Objasnite podusklađivanje i pretjerano usklađivanje na podatke na primjeru polinomske regresije.
19. Na primjeru određivanja parametara pravca objasnite princip RANSAC algoritma. Koje su prednosti i nedostatci RANSAC algoritma?
20. Navedite tri primjera binarne klasifikacije i tri primjera višeklasne klasifikacije. Što su moguće ulazne veličine u model u svakom navedenom primjeru?
21. Objasnite logističku regresiju: model, kriterijska funkcija, optimizacija parametara.
22. Objasnite OvR i OvO pristup na jednostavnom (2D) problemu s tri klase i dvije ulazne veličine. Koje su prednosti i nedostatci pojedinog pristupa?
23. Kako izgleda model multinomijalne logističke regresije? Kako se kodira izlazna veličina za potrebe učenja ovakvog modela (objasnite na primjeru klasifikacije rukom pisanih znamenki). Koliko parametara ima model u ovom slučaju?
24. Skicirajte softmax regresiju u formi potpuno povezanog sloja sa softmax aktivacijskom funkcijom. Napišite izraze za izračunavanje izlaznih vrijednosti.
25. Objasnite matricu zabune. Što predstavlja svaki njen element?

26. Objasnite točnost, preciznost, odziv i F1 mjeru. Što je preciznost-odziv krivulja i ROC krivulja. Kako ih dobivamo?
27. Kako se izračunavaju metrike za evaluaciju u slučaju višeklasne klasifikacije? Pokažite na vlastitom primjeru s tri klase.
28. Objasnite regularizaciju. Na koji način se modificira kriterijska funkcija? Skicirajte efekt regularizacije na jednostavnim primjerima (regresija i klasifikacija).
29. Objasnite algoritam  $K$  najbližih susjeda (KNN). Kako se najčešće definira mjera udaljenosti? Kako broj susjeda utječe na rezultate KNN algoritma? Koje su prednosti i nedostaci algoritma KNN?
30. Skicirajte princip određivanja optimalne vrijednosti nekog hiperparametra pomoću jednostavne provjere (validacije). Koji su nedostaci ovakvog pristupa?
31. Objasnite  $k$ -struku unakrsnu validaciju. Kako se određuje optimalna vrijednosti jednog ili više hiperparametara pomoću unakrsne validacije.
32. Od čega se sastoji stablo odlučivanja, koje su prednosti i nedostaci stabla odlučivanja? Što predstavlja mjera nečistoće (npr. entropija) i kako se ona koristi u procesu učenja? Kako je moguće spriječiti pretjerano usklađivanje na podatke stabla odlučivanja?
33. Objasnite temeljnu pretpostavku strojeva s potpornim vektorima. Kako se dobiva procjena izlazne veličine pomoću ovog algoritma? Kako se dobiva nelinearna (složena) granica odluke? Objasnite što predstavljaju hiperparametri  $C$  i  $\gamma$ .
34. Što je cilj nenadziranog učenja? Kakve podatke koristimo i zašto su baš takvi?
35. Objasnite algoritam  $K$  srednjih vrijednosti na jednostavnom problemu s dvije ulazne veličine.. Koji su načini inicijalizacije centara? Objasnite princip određivanja optimalnog broja grupa  $K$ .
36. Objasnite kvantizaciju boje digitalne slike pomoću algoritma  $K$  srednjih vrijednosti i kako se može postići kompresija slike.
37. Objasnite hijerarhijsko aglomerativno grupiranje na jednostavnom primjeru s dvije ulazne veličine (značajke). Na koje se sve načine može definirati udaljenost (sličnost) dva primjera ili grupe?
38. Objasnite princip rada DBSCAN algoritma.
39. Što je smanjivanje dimenzionalnosti? Zašto se ono radi? Koja su dva pristupa smanjivanju dimenzionalnosti?
40. Objasnite analizu glavnih komponenti. Kako se određuje potrebni broj komponenti? Na konkretnom primjeru (npr. skup podataka koji sadrže lica) objasnite kako se postiže kompresija podataka.
41. Skicirajte umjetni neuron. Koje aktivacijske funkcije se najčešće koriste i kako izgledaju? Kako se izračunava izlaz iz neuron na temelju ulaznih vrijednosti?
42. Izvedite izraze za podešavanje parametara umjetnog neurona s dvije ulazne veličine.
43. Što je umjetna neuronska mreža? Kakva je to unaprijedna potpuno povezana neuronska mreža? Skicirajte potpuno povezanu mrežu s 3 ulazne veličine, 6 neurona u skrivenom sloju i 4 neurona u izlaznom sloju. Koliko ovakva mreža ima parametara?
44. Kako se strukturira izlazni sloj neuronske mreže prilikom rješavanja regresijskih problema, a kako prilikom rješavanja klasifikacijskih problema?
45. Navedite osnovne korake prilikom izgradnje neuronske mreže. Objasnite načelno princip rada algoritma unazadne propagacije (engl. *backpropagation algorithm*). Što nam omogućava ovaj algoritam?
46. Napišite u općem obliku izraze koji opisuju unaprijednu propagaciju kroz mrežu s jednim skrivenim slojem.
47. Što je epoha i veličina *batcha*? Što je stopa učenja? Što je inicijalizacija parametara mreže?
48. Objasnite princip sprječavanja pretjeranog usklađivanja na podatke pomoću ranog zaustavljanja (engl. *early stopping*).

49. Objasnite tehniku nasumičnog izbacivanja neurona tijekom učenja (engl. *dropout*).
50. Što je 2D konvolucija. Objasnite/skicirajte na jednostavnom primjeru filtriranja digitalne slike.
51. Što je konvolucijski sloj i zašto se koristi? Skicirajte primjer primjene konvolucijskog sloja na neki ulazni volumen, naznačite ulazne i izlazne dimenzije. Što je *stride*? Što je *padding*? O čemu ovisi broj parametara konvolucijskog sloja?
52. Što je sloj sažimanja po maksimalnoj vrijednosti (eng. *max pooling*) i zašto se koristi? Skicirajte primjer primjene ovog sloja na neki volumen, naznačite ulazne i izlazne dimenzije.
53. Za danu konvolucijsku mrežu napišite dimenzije pojedinih volumena/izlaza iz slojeva te ukupan broj parametara mreže.
54. Što je augmentacija skupa podataka za učenje? Zašto se koristi?
55. Navedite najpopularnije „gotove“ strukture konvolucijskih neuronskih mreža koje se mogu koristiti kao ekstraktor značajki (što je karakteristično za svaku pojedinu mrežu?). Na kojem skupu su najčešće već istrenirane ove mreže i koje su karakteristike ovog skupa?
56. Što je učenjem prijenosom (engl. *transfer learnign*) i kako se odvija?