

Вовед во науката за податоци

1. Кои се предности на Двонасочните LSTM мрежи (Bi-directional LSTM)?

- a. Обично се подобри од еднонасочните рекурентни и LSTM мрежи ***
- b. Побрзи се при обучувањето
- c. Го зимаат предвид поширокото значење на контекстот ***
- d. Не бараат пристап до сите податоци однапред

2. Каква димензионалност треба да е влезно тренирачко множество кај LSTM невронска мрежа?

- a. 2D – матрица
- b. 1D
- c. 3D ***

3. Каков вид на учење се реализира кај Автоенкодерите ?

- a. Нагледувано (supervised)
- b. Полу-нагледувано(semi-supervised) ***
- c. Само-нагледувано(self-supervised) ***
- d. Со поттикување(reinforcement)

4. Што претставува инерција (momentum) при оптимизација на невронските мрежи?

- a. Метод со кој оптимизацијата на тежините обезбедува глобален оптимум. ***
- b. Параметар со кој се одредува моменталната активација на невроните
- c. Параметар со кој се влијае врз брзината на невронските мрежи
- d. Дека е многу тешко да се обучи невронската мрежа

5. Кај Обработката на природни јазици се среќаваат следниве задачи:

- a. Категоризација на теми ***
- b. Препознавање на векторски претстави на зборовите (word embeddings) ***
- c. Извлекување на контекстни зборови (skip-grams) ***
- d. Препознавање на именувани нешта ***

6. На кој начин се добиваат embedding на зборовите при тренирање на BERT модел?

- a. Се зима излезот на моделот
- b. Се искористуваат синусни и косинусни растојанија
- c. Преку тежините земени од скриените слоеви ***

7. Во кој случај би било најдобро да се употреби Sigmoid како излезно ниво кај невронските мрежи?

- a. Кога влезовите во мрежата се дискретни вредности
- b. Кога како мрежа за пресметка на загуба во мрежата се користи MSE (Mean Squared Error)
- c. Кога бројот на влезови е поголем од бројот на излези во невронската мрежа
- d. Кога сакаме да добиеме побрзо процесирање на резултатите на GPU
- e. Кога имаме бинарна класификација ***

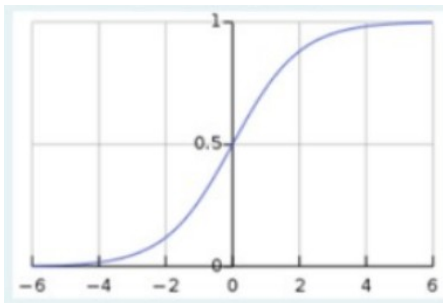
8. Во кој случај би било најдобро да се употреби Softmax како излезно ниво кај невронските мрежи?

- a. Кога имаме класификација во повеќе од две класи ***
- b. Кога сакаме да добиеме по брзо процесирање на резултатите на GPU
- c. Кога како мерка за пресметка на загуба во мрежата се користи MSE
- d. Кога бројот на влезови е поголем од бројот на излези во невронската мрежа
- e. Кога имаме длабока невронска мрежа

9. Нека е дадена реченицата: "It was a bright cold day in April, and the clocks were striking" Skip-gram со големина на прозорец три за зборот day е:

- a. a bright cold
- b. in April, and
- c. was bright cold April clocks were
- d. a bright cold in April and ***

10. Која активациска функција е претставена на графикот?



- a. relu
- b. sigmoid
- c. linear

11. Што е точно за моделот seq2seq?

- a. Крајниот скриен слој на енкодерскиот дел е влезен слој за декодерскиот дел. ***
- b. Обуката се одвива како и кај другите Рекурентни невронски мрежи. ***
- c. Предноста на seq2seq е што целото значење на реченицата е претставено во крајниот скриен слој на енкодерскиот дел. ??? (во лекциите пишува: "the entire "meaning" of the 1st sequence is expected to be packed into this one embedding", одлучи си дали тоа е исто со понуденото)
- d. При тестирањето се генерираат збор по збор, се додека не се добие на излез знак за крај на реченицата. ***

12. Кои карактеристики треба да ги има активациската функција кај невронските мрежи?

- a. Да има некаква нелинеарност ***
- b. Да овозможи градиентите да останат доволно големи и преку неколку скриени слоја ***
- c. Да дава активација само за позитивни влезови
- d. Да е заоблена

13. Кои од наведените карактеристики се новитети кај Трансформер моделите? (не сум сигурен за ова)

- a. Positional embeddings ***
- b. Self Attention layer ***
- c. Feedforward Network
- d. Tokenization

14. Еден од најдобрите јазични модели BERT се потпира на трансформер архитектура. Кој дел од трансформер архитектура се користи во BERT?

- a. Decoder
- b. Првите 9 нивоа од Encoder делот
- c. Encoder ***
- d. Decoder + Encoder

15. Еден од најдобрите јазични модели GPT-2 се потпира на трансформер архитектура. Кој дел од трансформер архитектура се користи во GPT-2?

- a. Decoder + Encoder
- b. Првите 9 нивоа од Encoder делот
- c. Decoder ***
- d. Encoder

16. Што претставува поимот отфрлање (dropout) во контекст на невронски мрежи?

- a. Бришење од меморијата при тестирање
- b. Случајно поставување на активацијата и тежините на врските на некои неврони на нули ***
- c. Трајно бришење од меморијата.
- d. Откривање на недостатоци и нивно отфрлање

17. Кои од следниве репозиториуми/библиотеки се користат за едноставно споделување на претренирани NLP модели?

- a. HuggingFace Transformers library ***
- b. PyTorch Hub ***
- c. GitHub
- d. TensorFlow-Hub ***

18. Кое од наведените можат да се користат како критериуми за прекин на понатамошното делење на јазли кај дрвата за одлучување (Stopping Conditions)?

- a. Ако бројот на примероци што припаѓаат на дадена класа го надмине дозволеният број
- b. Ако сите примероци во јазелот припаѓаат на истата класа ***
- c. Ако бројот на циклуси надмине даден праг
- d. Ако бројот на примероци во под-јазлите спадне под даден праг (min_samples_leaf) ***
- e. Ако бројот на јазлите во дрвото надмине даден праг ***

19. Кај Наивните Баесови класификатори, за атрибут A_i за дадена класа C може да претпоставиме:

- a. Условна зависност меѓу атрибутите, за таа класа
- b. $P(A_1, A_2, \dots, A_n | C) = P(A_1 | C) \times P(A_2 | C) \times \dots \times P(A_n | C)$ ***
- c. Условна независност меѓу атрибутите, за таа класа ***
- d. $P(A_1, A_2, \dots, A_n | C) = P(A_1 | C) + P(A_2 | C) \leftrightarrow P(A_n | C)$

20. Кои се предности на Long Short-term Memory (LSTM) мрежите?

- a. Можноста за учење на долги низи ***
- b. Потреба од мала меморија
- c. Краткотрајно бришење од меморијата
- d. Брзо учење при обука

21. Што претставува хиперпараметарот $n_estimators = 5$ во XGBoost модел?

- a. 5 процесори да се искористат за тренирање на моделот
- b. 5 внатрешни јазли во дрвото на одлука
- c. 5 дрва на одлука кои паралелно ќе се изградат ***
- d. 5 листа на дрвото на одлука

22. Word2vec како основа за креирање на Embeddings користи:

- a. n-grams
- b. part of speech tagging
- c. skip-grams ***
- d. one-hot embeddings

23. За што се користи Latent Dirichlet Allocation (LDA) алгоритмот?

- a. Topic Modeling ***
- b. Part-of-Speech (POS) tagging
- c. Named Entity Recognition
- d. Open Information Extraction

24. Колку често можат да се ажурираат тежините кај невронските мрежи?

- a. Ажурирање во серии (batch) ***
- b. Ажурирање во случајно расфрлани мини-серии (mini-batches) ***
- c. Ажурирање после секој примерок во множество за обука ***
- d. Ажурирање во моменти
- e. Ажурирање во конволуции

25. Кои особености ги има Преносното учење (Transfer Learning)? (не сум сигурен за а, с и d)

- a. Врши пренос на моментите во друга невронска мрежа
- b. Овозможува подобрување на перформансите *** (точно, го има во лекции)
- c. Може да научи преносно значење на зборовите
- d. Врши пренос на испуштениите јазли (drop-out) во друга невронска мрежа
- e. Користи означени податоци од други или сродни области *** (точно, го има во лекции)

26. Кои мерки ги користиме за сличност помеѓу два кластера?

- a. Бројот на елементи кои се наоѓаат во кластерите
- b. Сличноста помеѓу два случајно избрани елементи од двата кластера
- c. Најмалата различност помеѓу два елементи од кластерите ***
- d. Сличноста помеѓу центроидите на двата кластера ***

27. Кои од наведените параметри се дел од хиперпараметрите за тренирање на XGBoost моделот?

- a. `n_estimators` ***
- b. `min_depth`
- c. `learning_rate` ***
- d. `max_depth` ***

28. На кои од наведените модели за кластирање потребно е да се наведе бројот на кластери?

- a. K-Means Clustering ***
- b. AffinityPropagation Clustering
- c. DBSCAN Clustering
- d. Agglomerative Clustering

29. Што се Skip-grams?

- a. N-grams кои се појавуваат во дадена реченица но не се појавуваат во дадениот контекст
- b. Множество од не-последователни зборови (со одредено поместување), кои се појавуваат во некоја реченица ***
- c. Стоп зборовите кои се појавуваат најчесто
- d. Множество од сите зборови во реченицата

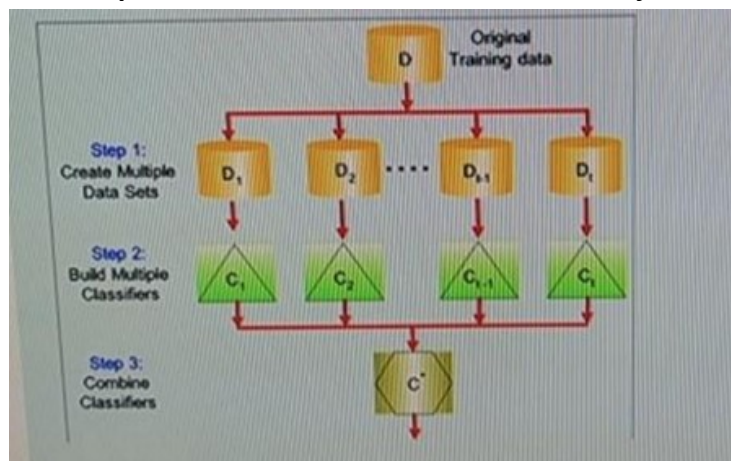
30. Кои од следниве се називи на алгоритми за оптимизација кај невронските мрежи?

- Adam , Adagrad (some others are: Momentum, Adadelata, RMSProp)

31. Што претставува Parts of Speech Tagging?

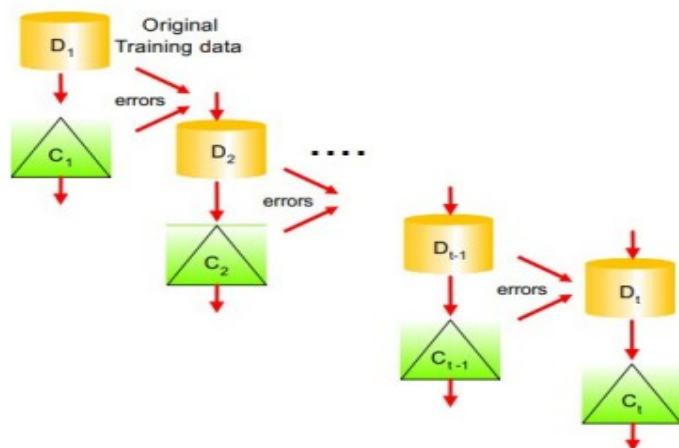
- Natural language processing (NLP) task that involves assigning a specific grammatical category (such as noun, verb, adjective, etc.) to each word in a sentence.

32. На прикажана слика е дадена шема за кој вид на учење со ансамбли?



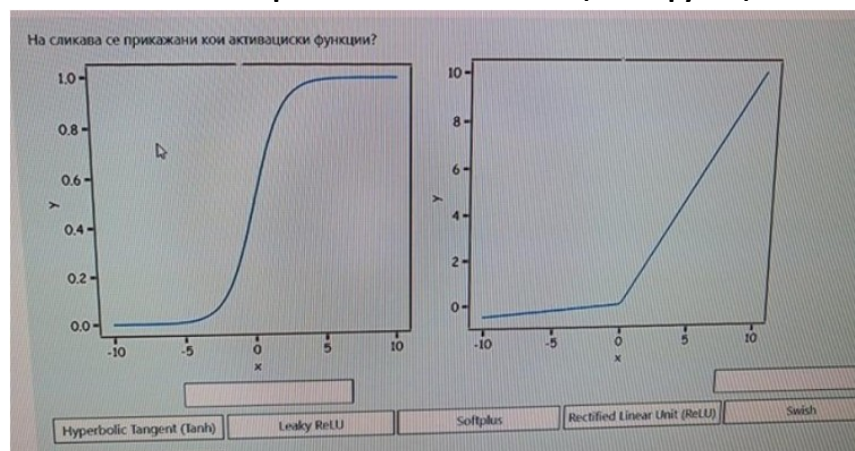
- bagging

33. На прикажана слика е дадена шема за кој вид на учење со ансамбли?



- boosting

34. На сликава се прикажани кои активациски функции?



1) Sigmoid 2) Leaky ReLU