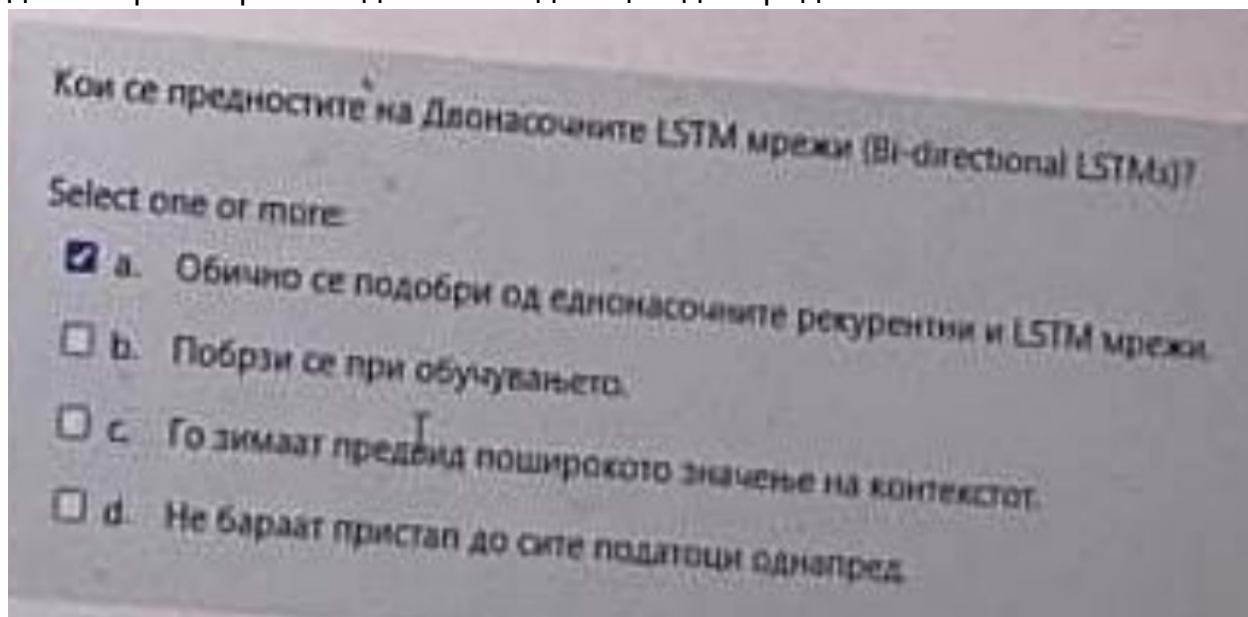
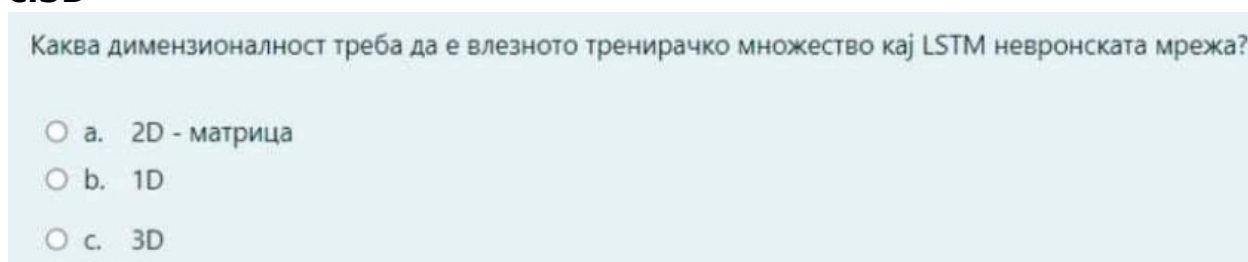


БАЗА ЗА ВТОР КОЛОКВИУМ – ВОБЕД ВО НАКУАТА ЗА ПОДАТОЦИ

1. Кои се предности на Двонасочните LSTM мрежи (Bi-directional LSTM) ?
а.Обично се подобри од еднонасочните рекурентни и LSTM мрежи
б.Побрзи се при обучувањето
ц.Го зимаат предвид поширокото значење на контекстот
д.Не бараат пристап до сите податоци однапред



2. Каква димензионалност треба да е влезно тренирачко множество кај LSTM невронска мрежа ?
а.2D – матрица
б.1D
с.3D



3. Каков вид на учење се реализира кај Автоенкодерите ?
а.Нагледувано (supervised)
б.Полу-нагледувано(semi-supervised)
с.Само-нагледувано(self-supervised)

d. Со поттикување (reinforcement)

Каков вид на учење се реализира кај Автоенкодерите?

Select one or more:

- ☐ a. надгледувано (supervised)
- ☐ b. полу-надгледувано (semi-supervised)
- ☒ c. само-надгледувано (self-supervised)
- ☐ d. со поттикнување (reinforcement)

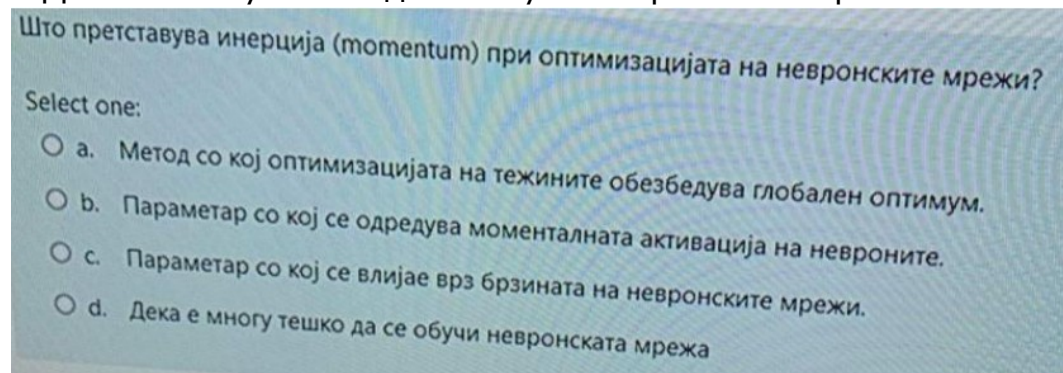
4. Што претставува инерција (momentum) при оптимизација на невронските мрежи ?

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



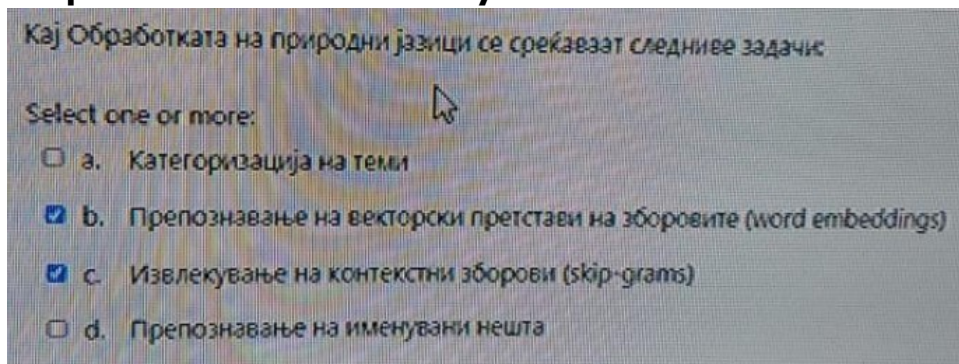
5. Кај Обработката на природни јазици се среќаваат следниве задачи:

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 –

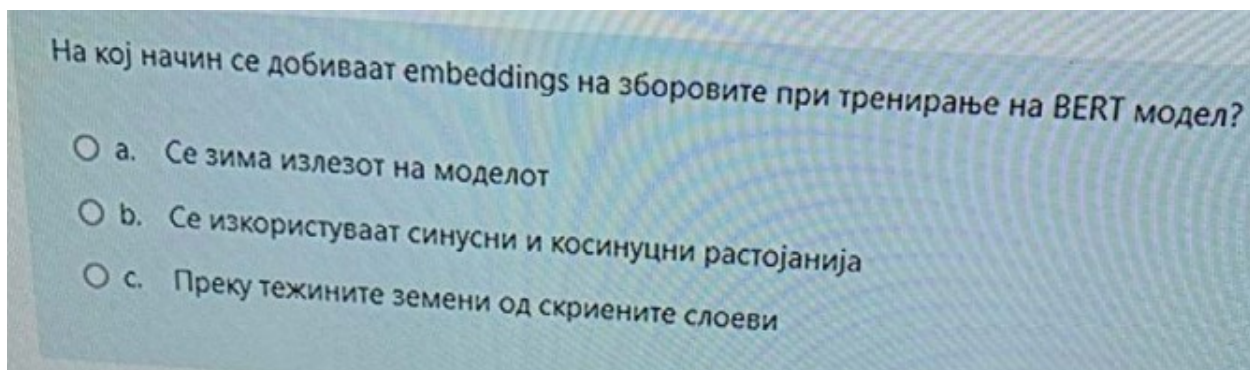


6. На кој начин се добиваат embedding на зборовите при тренирање на BERT модел?

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?

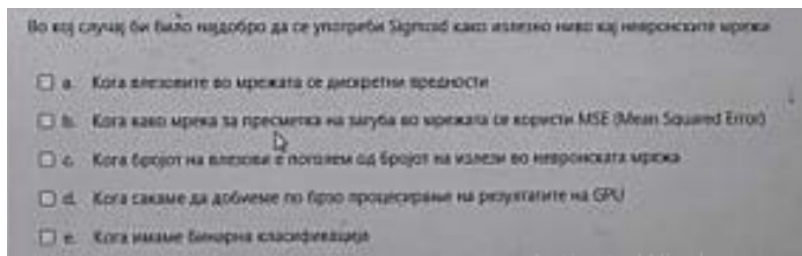
a. Кога влезовите во мрежата се дискретни вредности

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

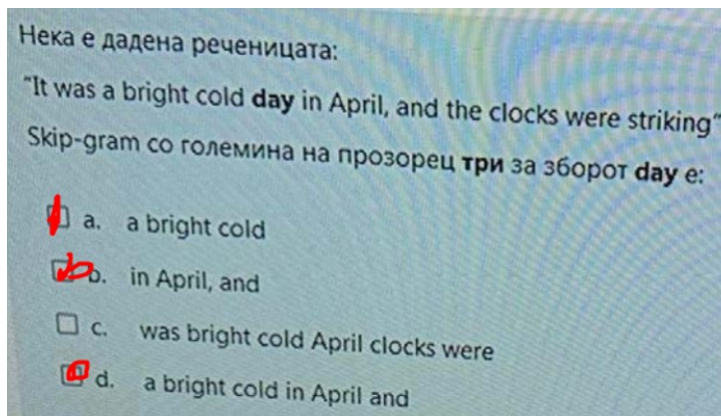
e. Кога имаме длабока невронска мрежа

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 (Mean Squared Error)
- ☒ d. Кога бројот на влезови е поголем од бројот на излези во невронската мрежа
- ☐ e. Кога имаме длабока невронска мрежа

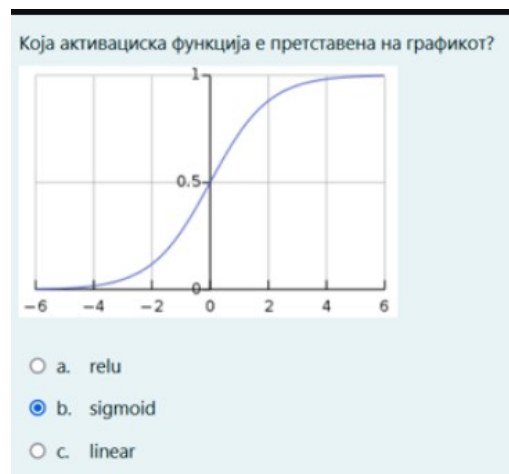
9. Нека е дадена реченицата: "It was a bright cold day in April, and the clocks were striking" Skip-gram со големина на прозорец три за зборот day е:

- a bright cold
- in April, and
- was bright cold April clocks were
- a bright cold in April and



10. Која активациска функција е претставена на графикот?

sigmoid



11. Што е точно за моделот seq2seq?

- Крајниот скриен слој на енкодерскиот дел е влезен слој за декодерскиот дел. - ТОЧНО
- Обуката се одвива како и кај другите Рекурентни невронски мрежи.
- Предноста на seq2seq е што целото значење на реченицата е претставено во крајниот скриен слој на енкодерскиот дел. - ТОЧНО
- При тестирањето се генерираат збор по збор, сè додека не се добие на излез знак за крај на реченицата. (Testing generates decoder outputs one word at a time, until we generate a token.)

Што е точно за моделот seq2seq?

Select one or more:

☐ 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.

☐ d. При тестирањето се генерираат збор по збор, сè додека не се добие на излез знак за крај на реченицата.

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

- Да има некаква нелинеарност - ТОЧНО
- Да овозможи градиентите да останат доволно големи и преку неколку скриени слоја

- Да дава активација само за позитивни влезови
- Да е заоблена

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

Select one or more:

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

13. Кои од наведените карактеристики се новитети кај Трансформер моделите?

- **Positional embeddings**
- **Self Attention layer**
- **Feedforward Network**
- Tokenization

Кои од наведените карактеристики се новитети кај Трансформер моделите?

- ☒ 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 ?

ENCODER

Еден од најдобрите јазични модели BERT се потпира на трансформер архитектурата. Кој дел од трансформер архитектурата 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 ?

DECODER

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

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

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

- Бришење од меморијата при тестирање
- **Случајно поставување на активацијата и тежините на врските на некои неврони на нули.** (Randomly set some neurons and their connections to zero)
- Трајно бришење од меморијата.
- Откривање на недостатоци и нивно отфрлање

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

Select one:

- ☐ a. Бришење од меморијата при тестирање.
- ☐ b. Случајно поставување на активацијата и тежините на врските на некои неврони на нула.
- ☐ 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

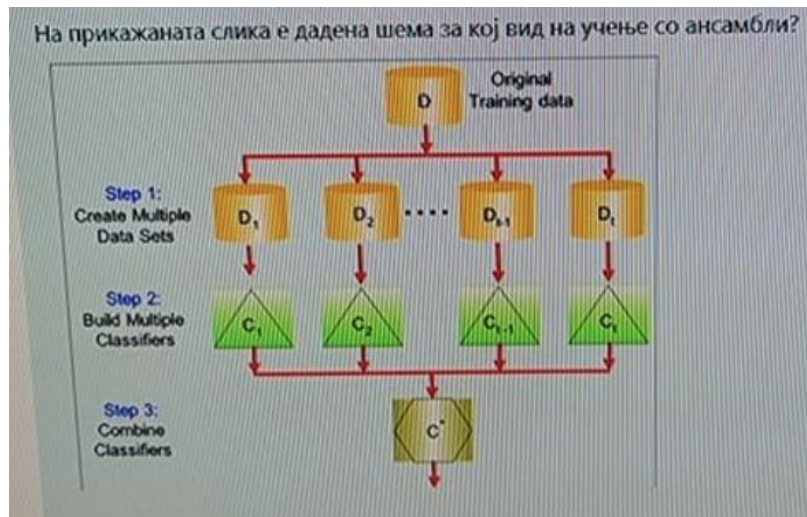
Кои од следниве репозиториуми/библиотеки се користат за едноставно споделување на претренирани NLP модели.

Select one or more:

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

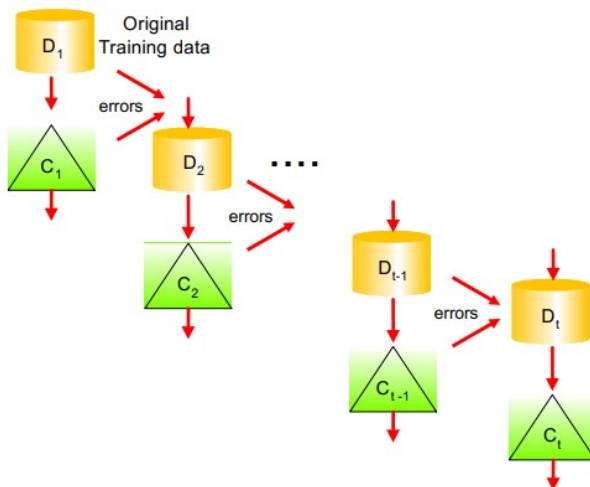
18. На прикажана слика е дадена шема за кој вид на учење со ансамбли ?

Bagging



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

Boosting



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

⌚ Ако бројот на примероци што припаѓаат на дадена класа го надмине дозволениот

број

⌚ Ако сите примероци во јазелот припаѓаат на истата класа ***

⌚ Ако бројот на циклуси надмине даден праг

⌚ Ако бројот на примероци во под-јазлите спадне под даден праг (`min_samples_leaf`) ***

⌚ Ако бројот на јазлите во дрвото надмине даден праг ***

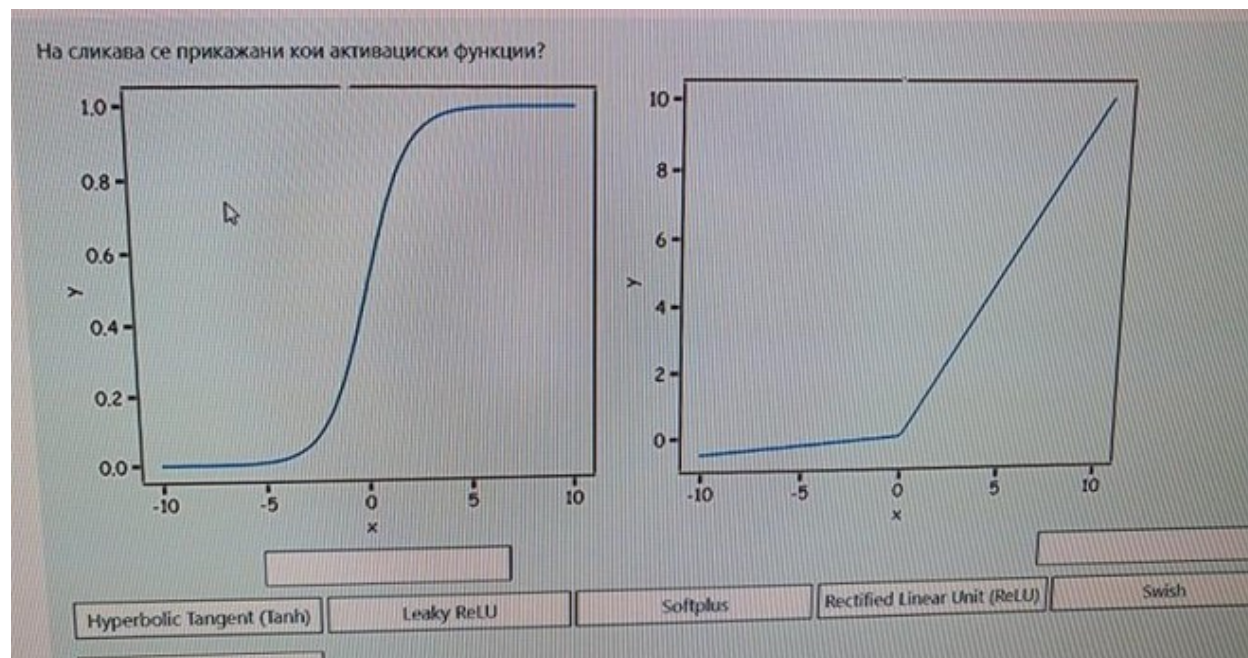
Кои од наведените можат да се користат како критериуми за прекин на понатамошното делење на јазлите кај дрвата за одлучување (Stopping Conditions)?

Select one or more:

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

21. На сликава се прикажани кои активациски функции ?

1.SIGMOID 2.RELU



22. Кај Наивните Баесови класификатори, за атрибутите A_i за дадена класа C може да претпоставиме

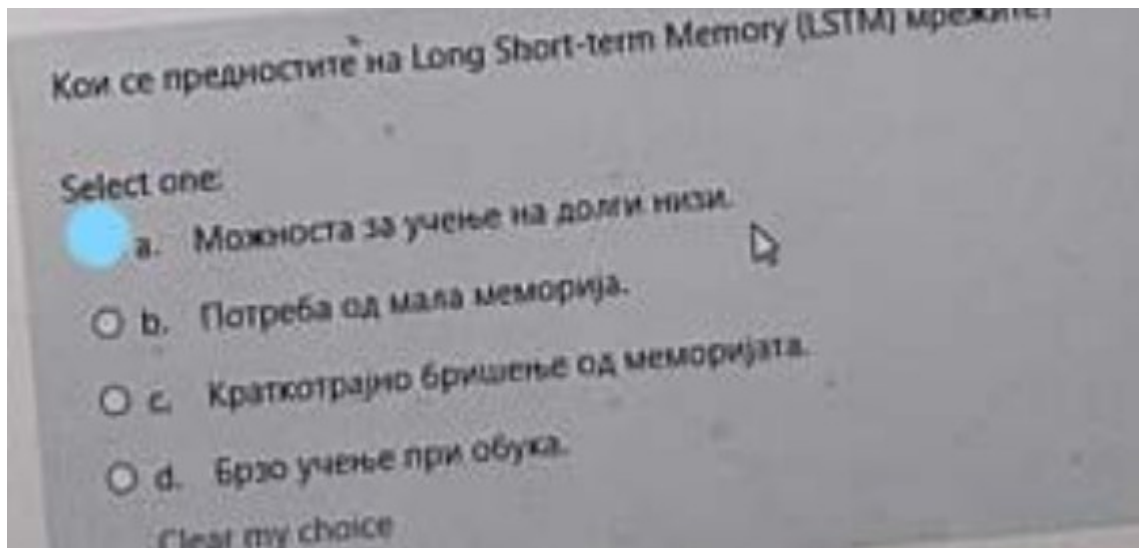
В,с

Кај Наивните Баесови класификатори, за атрибутите A_i за дадена класа C може да претпоставиме

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

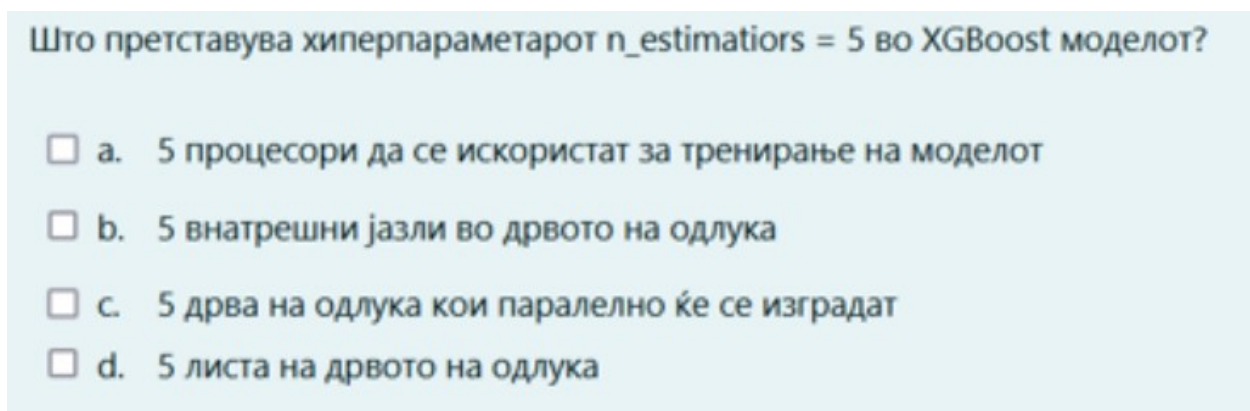
23. . Кои се предности на Long Short-term Memory (LSTM) мрежите ?

- **Можноста за учење на долги низи**
- Потреба од мала меморија
- Краткотрајно бришење од меморијата
- Брзо учење при обук



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

С 5 дрва на одлука кои паралелно ќе се изградат



25. 75. Кои од следниве се називи на алгоритми за оптимизација кај невронските мрежи? - Adam , Adagrad

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

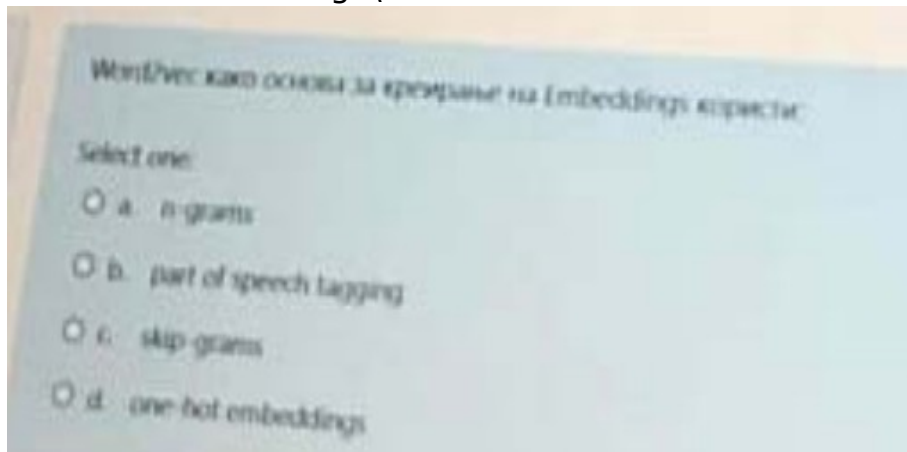
- Се зима излезот на моделот -
- Се искористуваат синусни и косинусни растојанија
- Преку тежините земени од скриените слоеви

На кој начин се добиваат embeddings на зборовите при тренирање на BERT модел?

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

27. Word2vec како основа за креирање на Embeddings користи:

- n-grams
- part of speech tagging
- **skip-grams**
- one-hot embeddings\



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

Topic Modeling

За што се користи Latent Dirichlet Allocation (LDA) алгоритмот

Select one:

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

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

??

Колку често можат да се ажурираат тежините кај невронските мрежи?

Select one or more:

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

30. Кои особености ги има Преносното учење (Transfer Learning)?

Кои особености ги има Преносното учење (Transfer Learning)?

Select one or more:

- ☐ a. Врши пренос на моментите во друга невронска мрежа.
- ☐ b. Овозможува подобрување на перформансите.
- ☐ c. Може да го научи преносното значење на зборовите.
- ☒ d. Врши пренос на испуштените јазли (drop-out) во друга невронска мрежа.
- ☐ e. Користи означени податоци од други или сродни области.

