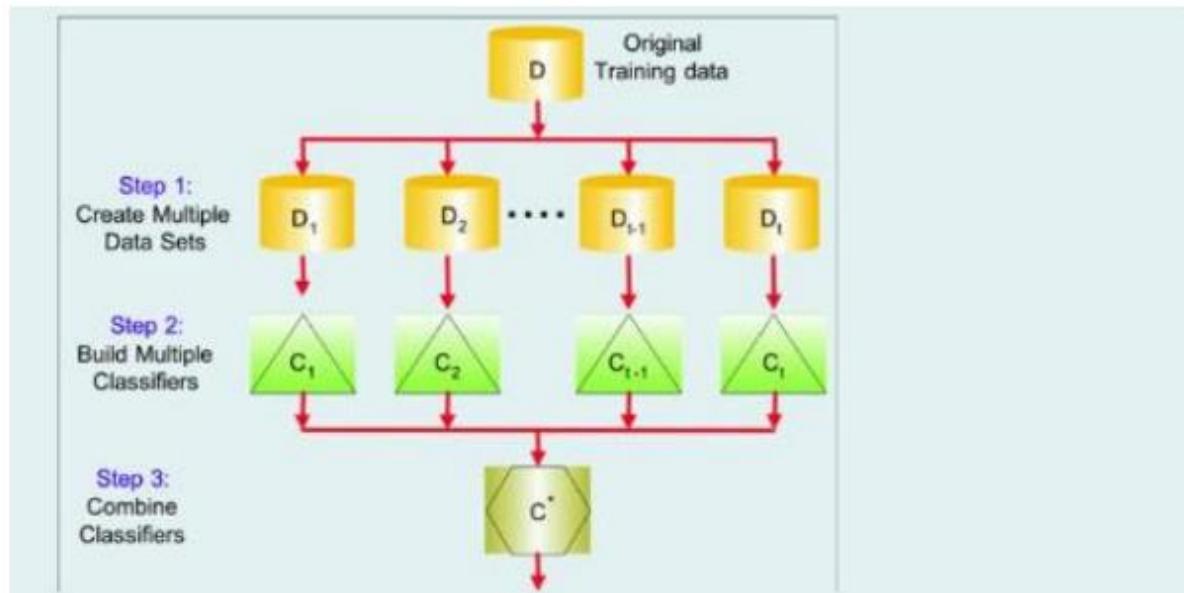


Шема за кој вид на учење со ансамбли?



- **Bagging**

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.
  - Noun (boat, plane, Obama)
  - Verb (goes, spun, hunted)
  - Pronoun (She, Her)
  - Preposition (in, on)
  - Adverb (quietly, then)
  - Conjunction (and, but)
  - Participle (eaten, running)
  - Article (the, a)

Кои од следниве репозиториуми/библиотеки се користат за едноставно споделување на претренираните NLP модели?

- **HuggingFace transformers library, PyTorch Hub, TensorFlow-Hub**

Кои од следните се називи на алгоритми за оптимизација кај невронските мрежи?

- Adam , Adagrad

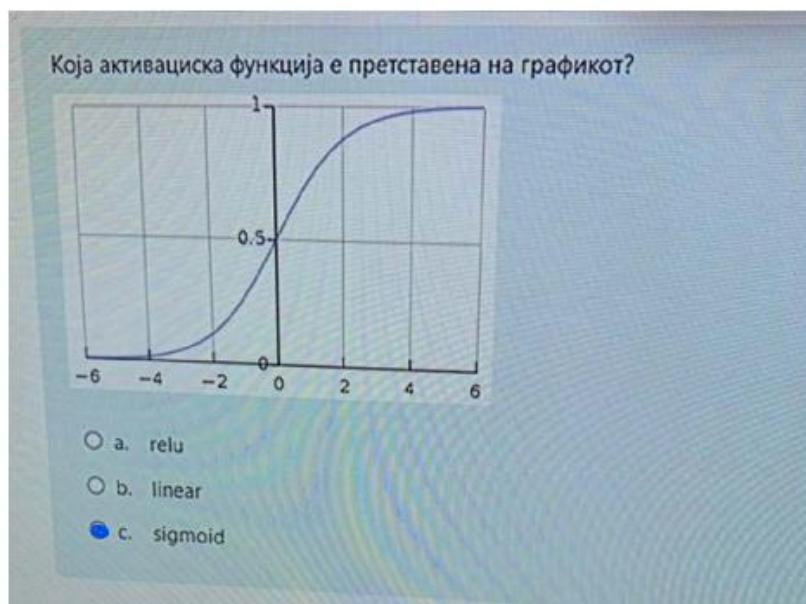
На кои од наведените модели за кластирање потребно е да се наведе бројот на кластери?

-K-Means Clustering



13. Која активациска функција е претставена на графикот?

- relu
- linear
- sigmoid – ВОА Е ТОЧНО



14. Каков вид на учење се реализира кај Автоенкодерите?

- Надгледувано (supervised)
- Полу-надгледувано (semi-supervised)
- Само-надгледувано (self-supervised)
- Со поттикнување (reinforcement)

Кои од наведените карактеристики се новитети кај Трансформер моделите?

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

Каква димензионалност треба да е влезното тренирачко множество кај LSTM невронската мрежа?

- a. 2D - матрица
- b. 1D
- c. 3D

2D – for non temporal data.

3D – for temporal data or sequences.

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.

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.

Кои мерки може да ги користиме за сличност помеѓу два кластера?

Select one or more:

- 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.

Со кој од дадените алгоритми за кластерирање може да се добијат шест кластери како што се прикажани на сликата



- a. K-means
- b. Hierarchical Clustering
- c. DBSCAN
- d. K-means++

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

Кои се предностите на Long Short-term Memory (LSTM) мрежите?

Select one:

a. Можноста за учење на долги низи.

b. Потреба од мала меморија.

c. Краткотрајно бришење од меморијата.

d. Брзо учење при обука.

[Clear my choice](#)

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

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

\*\*When do we use Softmax? -> discrete numbers/multi-class/multi-label, генерирање на текст или во предвидување на следната реченица.

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

Select one or more:

a. Да има некаква нелинеарност.

b. Да овозможи градиентите да останат доволно големи и преку неколку скринети слоја.

c. Да дава активација само за позитивни влезови.

d. Да е заоблена.

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

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

Кои се предностите на двонасочните LSTM мрежи (Bi-directional LSTM)?

Select one or more:

- 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.

Што претставува инерција (momentum) при оптимизацијата на невронските мрежи?

Select one:

- 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

Кај Обработката на природни јазици се среќаваат следниве задачи:

Select one or more:

- 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

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

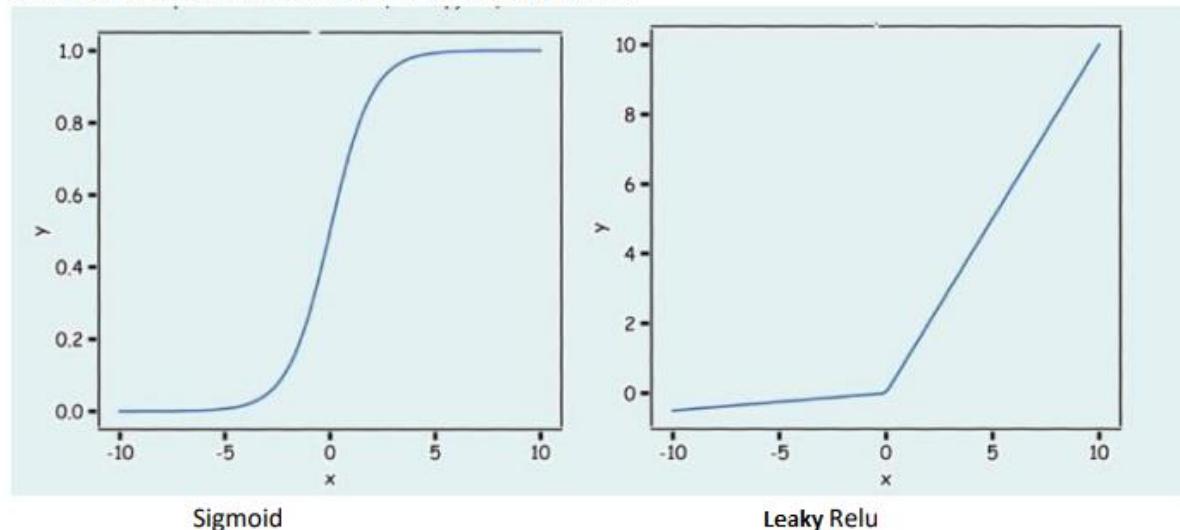
На кои од наведените модели за кластерирање потребно е да се наведе бројот на кластери?

- a. K-means Clustering
- b. AffinityPropagation Clustering
- c. DBCAN Clustering
- d. Agglomerative Clustering

Word2vec како основа за креирање на Embeddings користи:

- n-grams
- part of speech tagging
- skip-grams – ТОЧНО
- one-hot embeddings

На сликата се прикажани кои активацијски функции ?



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

### Decoder

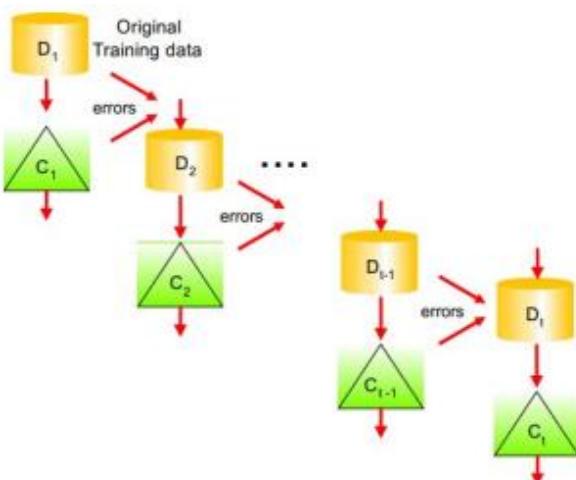
\*\*\*BERT -> Encoder

8. Во кој случај 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

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

### Boosting



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

Што претставува инерција (momentum) при оптимизацијата на невронските мрежи?

Select one:

- 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

Што претставува хиперпараметарот n\_estimators = 5 во XGBoost моделот?

- 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

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

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

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

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

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.