

Propostes de preguntes per als exàmens de Xarxes Neuronals i Aprenentatge Profund

Recordeu que les preguntes seran majoritàriament de resposta molt curta.

Si poseu preguntes noves indiqueu-ho aquí, només cal la secció:

La majoria de preguntes que heu posat necessiten desenvolupament. L'examen en principi no serà d'aquest tipus. Tot i que hi ha preguntes marcades en verd, no estaran articulades com ho heu fet vosaltres sinó que es faran per a que la resposta pugui ser curta o test

El power point de Attention (Transformers...) sortirà? NO

Podeu afegir els punts que considereu necessaris. Respecteu les aportacions dels companys. Uns dies abans de l'examen i si el nombre de preguntes és significatiu posaré comentaris sobre la idoneïtat de les preguntes

pregunta candidata per l'examen

podria sortir una versió modificada

no sortirà a l'examen

comentari de professor

TEMA 0: Preguntes miscelània (inclouen o relacionen conceptes de més d'un tema).

1. Una pregunta, pots dir-nos quantes preguntes podem esperar trobar-nos al examen? Si ho sapigues vos ho diria. Menys de 50. El 80 o 90 % de les preguntes estan inspirades en les que vosaltres heu posat
2. ¿Como se relacionan el overfitting y la regularización? Aquest és un concepte que es va veure el primer parcial. Aquest no hi hem entrat massa perquè l'hem donat per enterat
- 3.

TEMA 1: Intro NN

1. Notació de NN: identificar posició de node, o etiquetar nodes en una NN massa simple
2. Explicació/Esquema d'un Perceptron és molt probable que no surtin preguntes de text lliure llarg
3. ¿Qué pasa si usamos demasiadas capas ocultas sin regularización? //¿Cómo puede beneficiar la regularización a los modelos con muchas capas ocultas?
4. ¿Qué es el desvanecimiento del gradiente?
5. ¿Qué función de activación usarías para: on? Abans de calcular la loss? Però la idea és bona
 - A) Clasificación binaria
 - B) Clasificación multiclase
 - C) Capas ocultas
6. Quina és la funció principal de les capes oculta en una xarxa neuronal?

7. ¿Por qué añadimos un bias al vector de entrada del perceptrón?
8. Tenint una matriu jacobiana per a N dimensions, per què podem reduir la seva dimensionalitat a 2D? Aneu forts, m'agradaria preguntar-la
9. ¿Qué variación de perceptrón usarías si encuentras que tu modelo tiene poca estabilidad? Si la respuesta es MLP la pregunta no es correcta, sino, no se per on vas Em referia a AdaTron i MinOver
10. ¿Cómo podemos evitar la limitación del perceptrón que “solo puede clasificar clases linealmente separables”?
11. ¿Cual es el papel de la función de activación?
12. El perceptró funciona per problemes no lineals (T/F) simple
13. Què és el softmax?
14. La regla de Hebb disminueix el pes de la connexió en cada instant següent del camí. (T/F)
15. ¿Para qué problemas se usa Softmax y porqué?

TEMA 2: MLP, BackProp

1. Com es modifica l'algorisme de backpropagation per treballar amb replicated features?
2. Capes convolucionals detecten patrons locals i les capes de padding permeten que la xarxa reconegui els patrons independentment de la ubicació (T/F) ??? crec que està mal redactada. no l'entenc
3. Perquè fer transposed convolution és equivalent a fer una convolució amb el kernel transposat amb full padding. (T/F). això es definició, seria millor preguntar perquè
4. Què és una funció d'activació i per què és important? Pregunta 1-11
5. Què és una xarxa neuronal convolucional (CNN) i com es diferencia d'un MLP? Centrar la pregunta, massa extensa
6. Si la derivada parcial del error respecto al peso que queremos actualizar es positiva (/negativa), debemos aumentar o disminuir el peso? Por qué? Això es descens del gradient que correspon al primer semestre
7. ¿Qué nos indica que la derivada sea grande en relación con la actualización del peso? ???
8. ¿Qué regla utilizamos para el backward pass?
9. ¿Como/ Porque soluciona ReLU el problema de Vanishing Gradient de Sigmoid?
10. Utilitat de tenir més hidden layers concretar
11. Per què inicialitzem els gradients a zero abans d'una època? Segur? No és avans de cada minibatch?
12. La funció sigmoide provoca cèl·lules mortes. (Sí/no)

TEMA 3: Activations, losses

1. Què són les unitats mortes en el context de ReLU? No nomes en RELU
2. Els models aprenen menys de les mostres més inesperades. (T/F)
3. El MSE penalitza més els errors grans i MAE és menys sensible a valors atípics. (T/F)
4. Quina és la diferència principal entre una classificació multi-class i una classificació multi-label? Aquest dilema segur que surt, no en aquests termes
5. ¿Por qué/cómo SGD puede hacerte escapar de esos mínimos locales en los que te puede hacer caer BGD?

6. Que ens aporta el stochastic loss? Suposo que us refeririu al stochastic GD, es una pregunta molt similar a l'anterior
7. ¿Qué estrategias (2) podemos implementar para impedir el ruido excesivo (inestabilidad) de SGD? No veig per on vas
8. ¿Por qué en BGD el orden de presentación de los datos no es relevante? es millor preguntar porque en minibatch o el stocastic si que es rellevant
9. ¿Qué significa que una unidad/neurona tenga una saturación peligrosa? Que vol dir?
10. ¿Cómo puede ayudar las conexiones residuales a el vanishing gradient?
11. Què és la matriu jacobiana? La matriu jacobiana és una matriu de derivades parcials que descriu com canvien les variables d'un sistema en funció de les altres, i la pregunta seria, quines variables son les que canvien i quines les que fan canviar? I quina dimensio te una matriu jacobiana i porque?
12. Diferencia entre one-hot y multi-hot y aplicación de cada uno
13. La derivada de tanh pasa per el 0?
14. ¿Cómo podemos ayudar a softmax a tener la estabilidad que le falta cuando el vector z contiene numeros grandes? ¿Y porque esa ayuda funciona? Uff, és de nota!

TEMA 4: Training deep nets

1. Què és el descens de gradient estocàstic (SGD)?
2. Quina és la importància de la inicialització dels pesos en una xarxa neuronal?
3. Com es poden evitar els problemes de sobreajustament en una xarxa neuronal? Molt llarga
4. Quina és la funció del dropout en xarxes neuronals?
5. Diagnosticar learning curves
6. Què és el problema del vanishing gradient? Totes les que son descripció s'han de re-redactar i buscatn alguna cosa més que la definicio
7. Què provoca un exploding gradient?
8. Per què no es poden inicialitzar tots els pesos a 0?
9. Què és el Dropout i per què s'utilitza?
10. Què vol dir early stopping i quin problema evita? No evita overfitting
11. (de la 4.2) ¿Cómo puede afectar inicializar todos los pesos con el mismo valor?
12. ¿Cómo puede afectar una inicialización de pesos demasiado grandes o pequeños?

TEMA 5: CNN

1. Com es processen les imatges en una xarxa neuronal convolucional? No veig per on va la pregunta
2. Què és l'augment de dades? No és només de CNN pero alguna pregunta caura de data augmentation
3. Què fa una xarxa neuronal convolucional? Massa larga
4. Què és una convolució en una xarxa neuronal convolucional?
5. Què és la visió per computadora? No és d'aquesta assignatura
6. Quina és la funció dels filtres en una xarxa neuronal convolucional?
7. Quina és la funció de la capa de pooling en una CNN?
8. Perquè és necessaria la data augmentation? (què aporta) les respostes han de ser curtes

9. Quin benefici aporta aplicar el mètode Pooling? (a) Augmenta la resolució de la imatge b) **Redueix la dimensionalitat de la imatge** c) Duplica la mida dels filtres convolucionals d) S'utilitza per normalitzar els valors de píxels)
10. Com s'escolleixen els pesos dels filtres que se li aplica a cada convolució? (a) Es defineixen manualment abans de l'entrenament b) Són seleccionats aleatòriament en cada època c) **S'aprenen automàticament durant l'entrenament** d) Són iguals a totes les capes de la xarxa)
11. En imatges, com es perd la spacial information?
12. Per quines d'aquestes aplicacions no podem fer servir CNN (a) object detection b) segmentation c) probability classification d) **Totes elles es poden utilitzar amb CNN**)
13. Per què es fa flatten?
14. Mlp conserva estructura espacial. (True/false)
15. No se puede usar Stride sin padding (True /false)
16. Una imagen puede tener más de tres canales (true /false) seria de psiv no?
17. Un solo filtro / kernel puede detectar más de una característica. (True/false).
18. En las pooling layers aprende parametros (True/false) es un regal

TEMA 6: Classificació

1. Quines diferències existeixen entre multiclass i multilabel classification? Pq multiclass no usa sigmoid?
2. Finetuning vs feature extraction(diferencias o aplicaciones) massa llarga, pero necessaria
3. Què és Stochastic Depth?
4. El cross-entropy mesura la precisió. (T/F)
5. Un feature map és una imatge. (T/N) que és una imatge?
6. El gradient pot desaparèixer en xarxes profundes. (T/N)
7. Cada filtre d'una CNN què representa? Una característica específica que ha après
8. A cada filtre li correspon una feature map (T/F)
9. El Dropout només s'aplica durant la fase de test per estabilitzar la sortida(T/F)

TEMA 8: Autoencoders.

1. Utilitat convolucio tranposta(explicacio o pregunta multiple)
2. Diferències principals entre els autoencoders tradicionals, VAE i GAN
3. Quin és l'objectiu principal d'un autoencoder?
4. Quin tipus de pèrdua (loss function) s'utilitza habitualment per entrenar un autoencoder?
5. Com sabem si l'embedding és útil? Excesivament ambigua
6. Per què pot ser útil introduir soroll a la imatge d'entrada en un autoencoder?

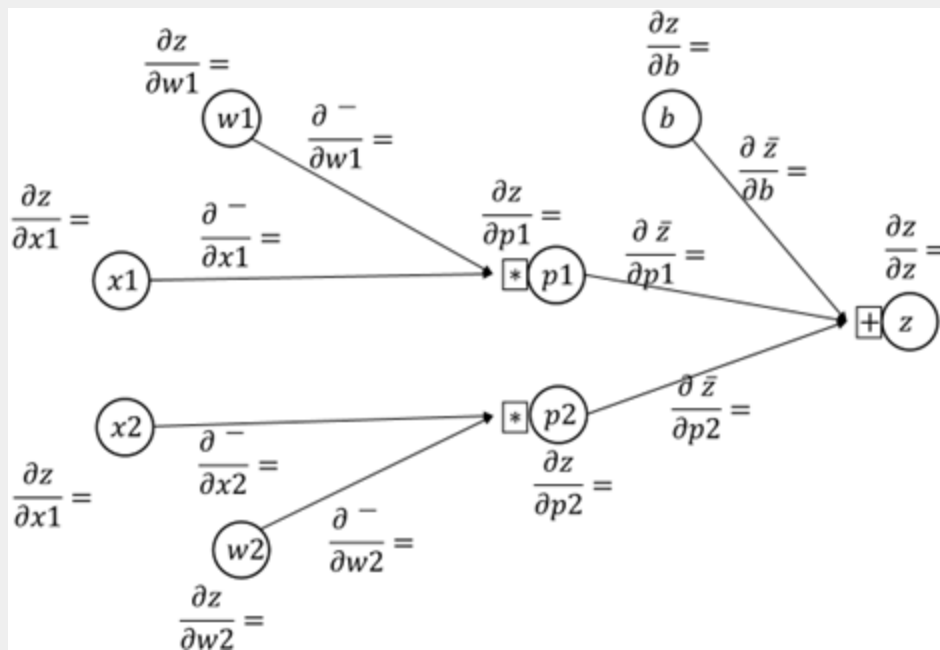
TEMA 9: Sequential reasoning: RNN.

1. Quines Gates te LSTM? I quines Gates te GRU?
Forget input i calculate, update i reset
2. Quantes matrius de pesos s'han d'actualitzar a l'hora de fer Backpropagation en una RNN? En quin model de RNN?
3. Què vol dir tokenizing? I vectorizing?
4. Quins beneficis aporta una LSTM? (a) ~~afegir complexitat al model~~ RNN b) reduir la complexitat del model c) càlculs més senzills d) no cap aporta benefici)

5. Quina informació s'utilitza a l'hora d'actualitzar l'estat intern? (Múltiples opcions: a) **input information** b) **past state information** c) the previous loss d) the vectorization of the first input) **pero les opcions dependra del tipus de cella (Elman, lst, gru)**
6. En que consisteix el Forward path? (a) algo relacionat amb l'augment de loss b) consisteix en escollir els passos en que es dura a terme el RNN c) the computed loss at each state **d) the sum of all the computed loss at each state** **pero aquesta pregunta no seria valida per a totes les xarxes?**
7. Quina d'aquestes opcions es un tipus d'embedding (a) one-hot embedding b) Learned embedding **c) a i b** d) Cap de les anteriors) **massa simple**
8. Quina es la funcio principal d'un embedding? **(a) Codificar instancies d'entrada paraules amb vectors** b) Millorar la memòria de la GPU durant l'entrenament c) Traduir textos entre idiomes de forma automàtica d) Assignar una categoria gramatical a cada paraula) **Totes les xarxes neurals generen embeddings**

TEMA 10: Metric learning.

1. **En quins escenaris ens convindrà fer servir few-shot learning?**
 2. Digues una diferència entre few-shot learning, one-shot learning i zero-shot learning. **Prefereixo l'estil de pregunta tipus l'anterior.**
 3. Com funciona una xarxa Siamese? **descriptiva**
 4. Què s'entén per hard negative i semi-hard negative? **descriptiva**
 5. Què és una pretext task? **descriptva**
 6. Què és l'aprenentatge multimodal? **descripiu**
 7. Dona dos exemples de modalitats diferents que es poden combinar en un sistema multimodal.
 8. Què significa que dues modalitats tinguin embedding compartit? **Que signigica**
 - 9.
-



per a que us feu una idea del tipus de problemes que pot entrar com a exemple la pregunta d'un any anterior:

A) faltaria indicar les derivades globals i locals,

Mes info en

<https://simple-english-machine-learning.readthedocs.io/en/latest/neural-networks/computational-graphs.html>

B) calcular el numero de paràmetres que una determinada xarxa ha d'aprendre

C) calcular les mides dels filtres i el tamany dels feature maps en una CNN. Com a exemple la pregunta de l'any passat

