

DOMANDE APPRENDIMENTO AUTOMATICO DIVISE PER CATEGORIE

Sommario

DOMANDE APPRENDIMENTO AUTOMATICO DIVISE PER CATEGORIE.....	1
Alberi di decisione.....	2
Random forest.....	2
AlexNet.....	3
apprendimento supervisionato.....	3
Autoencoders.....	3
backpropagation per reti neurali.....	3
Learning rate.....	4
campo recettivo (receptive field) di un neurone di una CNN.....	4
classificazione lineare.....	5
Dadi / monete / probabilità.....	5
dataset / recall del modello.....	5
deep features.....	6
distribuzione congiunta di probabilità.....	6
Probabilità condizionata.....	6
entropia / crossentropy.....	6
funzione logistica.....	7
funzione loss in rete neurale.....	7
GAN.....	8
layer convolutivo 2D.....	8
Long-short term memory models (LSTMs).....	9
MaxPooling (derivata).....	9
Minibatch.....	10
modelli generativi.....	10
mutua informazione (information gain).....	10
Naive Bayes.....	10
neuroni artificiali.....	11
Overfitting.....	11
regressione logistica.....	11
regressione multinomiale.....	12
Regressione lineare.....	12
reti per classificazione di immagini.....	12
ReLU(x) (Rectified linear unit).....	13

scomparsa del gradiente (vanishing gradient).....	13
Softmax.....	13
stride in un layer convolutivo.....	13
tecniche discriminative.....	13
tecnica a discesa del gradiente.....	14
transposed convolutions.....	14
U-net.....	14
Inception module.....	14
Minimi locali - fase training.....	15
Intersection over Union (IoU).....	15
Transformers.....	15

Alberi di decisione

1. Selezionare l'affermazione corretta sugli alberi di decisione:
 - a. Possono essere utilizzati solo con features discrete
 - b. Possono esprimere qualunque funzione di classificazione**
 - c. Non presentano problemi di overfitting
 - d. Il costo computazionale della produzione è piuttosto elevato
2. Selezionare la sentenza erranea riguardo gli alberi di decisione
 - a. Possono essere utilizzati solo con features discrete**
 - b. Il costo computazionale è molto basso
 - c. Hanno una forte tendenza all'overfitting
 - d. Possono esprimere qualunque funzione di classificazione
3. Nel caso di un albero di decisione con features discrete, cosa si può dire della profondità dell'albero?
 - a. È minore o uguale al numero delle features**
 - b. Non si può dire nulla
 - c. È minore o uguale al numero delle classe
 - d. È sicuramente maggiore del logaritmo in base due del numero dei dati

Random forest

1. Selezionare la sentenza errata relativa alle random forest (foreste di alberi decisionali)
 - a. Richiedono tecniche opportune per la creazione di alberi di decisione diversi relativi ad uno stesso dataset
 - b. Tendono a migliorare l'explainability (spiegabilità) degli alberi di decisione riducendo l'instabilità nella selezione degli attributi**
 - c. Tentano di mitigare il fenomeno dell'overfitting tipico degli alberi di decisione
 - d. È una tecnica di apprendimento ad "ensemble" basata su una combinazione di alberi di decisione

AlexNet

1. AlexNet, la prima rete convoluzionale profonda vincitrice della ImageNet competition è stata realizzata in quale anno:

- a. 2012
- b. 1993
- c. 1971
- d. 2019

apprendimento supervisionato

1. Selezionare la sentenza errata riguardo all'apprendimento supervisionato
 - a. **Richiede la costante supervisione di un esperto durante il training**
 - b. Può comprendere sia problemi di regressione che di classificazione
 - c. Si riferisce all'apprendimento di funzioni basato su esempi di training composti da coppie di input-output
 - d. La definizione della ground truth può richiedere l'intervento umano ed essere onerosa
2. Che cosa si intende con apprendimento supervisionato?
 - a. Apprendimento sotto la supervisione diretta di un esperto
 - b. Apprendimento che tende ad imitare il comportamento di un esperto
 - c. Apprendimento che non fa uso di tecniche statistiche o probabilistiche
 - d. **Apprendimento di funzioni basato su esempi di training composti da coppie input-output**
3. In che situazioni si parla di apprendimento auto-supervisionato? (self supervised)
 - a. Quando il modello è in grado di configurare in modo automatico la propria architettura
 - b. Quando il modello è supposto contribuire alla creazione di nuovi dati di training
 - c. **Qualora i dati di input possano essere considerati come annotazioni (labels) per guidare l'apprendimento, come nel caso degli autoencoders**
 - d. Quando l'apprendimento prevede una sinergia tra l'uomo e la macchina

Autoencoders

1. Quale delle seguenti sentenze relative agli autoencoders è scorretta?
 - a. **Gli autoencoders richiedono l'uso di livelli densi**
 - b. Possono essere utilizzate per la rimozione di rumore (denoising)
 - c. L'encoder e il decoder non devono essere necessariamente simmetrici
 - d. La rappresentazione interne prodotta dall'encoder abitualmente ha una dimensione ridotta rispetto a quella di partenza
2. Quale delle seguenti sentenze relative agli autoencoders è corretta?
 - a. Gli autoencoders richiedono l'uso di livelli densi
 - b. L'encoder e il decoder devono essere strettamente simmetrici
 - c. **La rappresentazione interna prodotta abitualmente ha una dimensione ridotta rispetto a quella di partenza**
 - d. È una rete neurale che codifica se stessa
3. Quale delle seguenti non è una applicazione tipica degli autoencoders?
 - a. **Segmentazione di immagini (semantic segmentation)**
 - b. Rimozione del rumore (denoising)
 - c. Riduzione delle dimensioni (dimensionality reduction)
 - d. Rilevamento di anomalie (anomaly detection)

backpropagation per reti neurali

1. Selezionare la sentenza scorretta relativa alla backpropagation per reti neurali
 - a. Richiede la memorizzazione delle attivazioni di tutti i neuroni della rete durante la forward pass

- b. Ha un costo computazionale paragonabile a quello del calcolo “in avanti” (inference) lungo la rete
 - c. Si basa tipicamente su algoritmi di tipo generico
 - d. Tipicamente, si effettua solo durante la fase di “training” della rete
2. Quale delle seguenti affermazioni relative alla backpropagation è corretta?
- a. Si effettua solo durante il “training”
 - b. Viene fatta sia durante la fase di “inference” (calcolo in avanti) che in quella di “training”
 - c. È molto più costosa, in termini di tempo, del calcolo in avanti lungo la rete
 - d. Viene effettuata unicamente lungo le skip connections delle reti residuali, per evitare perdita del gradiente
3. Selezionare la sentenza scorretta relativa alla backpropagation per reti neurali
- a. È l’algoritmo per il calcolo della derivata parziale della loss rispetto a ogni parametro della rete
 - b. Tipicamente, il gradiente viene artificialmente rinforzato ad ogni layer attraversato per contrastare il fenomeno della sua scomparsa (vanishing)
 - c. Si riduce a semplici calcoli algebrici facilmente parallelizzabili in strutture di calcolo tipo GPU
 - d. L’algoritmo calcola il gradiente un layer alla volta, sfruttando la regola matematica per la derivazione di funzioni composte

Learning rate

1. Selezionare la sentenza errata relativa al learning rate
- a. È una metrica che misura la capacità di apprendimento del modello
 - b. Un learning rate alto tipicamente velocizza il training ma potrebbe saltare sopra al minimo
 - c. È un iper-parametro che definisce la lunghezza del passo durante la discesa del gradiente
 - d. Il learning rate può variare durante il training

campo ricettivo (receptive field) di un neurone di una CNN

1. Selezionare la sentenza scorretta relativa al campo ricettivo di un neurone di una CNN:
- a. Definisce la porzione dell’input che influenza l’attivazione di un determinato neurone
 - b. Dipende dalla profondità del layer in cui si trova il neurone e dalle dimensioni e gli striders dei kernel dei layers precedenti
 - c. È sempre almeno pari alla dimensione spaziale del dato di input
 - d. Aumenta rapidamente con l’attraversamento di livelli con downsampling
2. Il campo ricettivo (receptive field) di un neurone di una CNN dipende da:
- a. La profondità del layer in cui si trova il neurone e le dimensioni e gli striders dei kernel dei layers precedenti
 - b. La profondità del layer in cui si trova il neurone e le dimensioni dei kernel dei layers precedenti, ma non dai loro striders
 - c. La dimensione del kernel e il numero dei canali del layer in cui si trova il neurone
 - d. Unicamente dalla profondità del layer a cui si trova il neurone
3. Componendo due layer Conv2D con stride 1, il primo con kernel 5x5 e il secondo con kernel 3x3 quale sarà il campo ricettivo dei neuroni finali?
- a. 7
 - b. 8

- c. Dipende dal padding
- d. 3

classificazione lineare

1. In quale di questi casi una tecnica di classificazione lineare potrebbe non fornire risultati soddisfacenti?
 - a. Quando le features sono indipendenti tra loro, data la classe
 - b. Quando non tutte le features di input sono rilevanti ai fini della classificazione
 - c. Quando la classificazione dipende da un confronto tra features
 - d. Quando esiste una elevata correlazione tra le features

Dadi / monete / probabilità

1. Ci sono due dadi, uno normale e uno truccato che restituisce 6 con probabilità 0.5 e gli altri valori con probabilità 0.1. Faccio tre lanci con lo stesso dado e osservo un 3, un 6 e un 2, cosa posso concludere?
 - a. La probabilità di usare uno o l'altro dei dadi è esattamente la stessa
 - b. Nulla
 - c. È più probabile che il dado sia normale
 - d. È più probabile che il dado sia truccato
2. Ci sono due dadi, uno normale e uno truccato che restituisce 6 con probabilità 0.5 e gli altri valori con probabilità 0.1. Faccio tre lanci con lo stesso dado e osservo un 3 e un 6 cosa posso concludere?
 - a. La probabilità di usare uno o l'altro dei dadi è esattamente la stessa
 - b. Nulla
 - c. È più probabile che il dado sia normale
 - d. È più probabile che il dado sia truccato
3. Ci sono due monete, una normale e una che restituisce testa con probabilità $\frac{3}{4}$ e croce con probabilità $\frac{1}{4}$. Faccio due lanci con la stessa moneta e osservo una testa e una croce. Che cosa posso concludere?
 - a. Nulla
 - b. È più probabile che la moneta sia normale
 - c. È più probabile che la moneta sia truccata
 - d. La probabilità di usare uno o l'altra moneta è esattamente la stessa

dataset / recall del modello

$$\text{RECALL} = \frac{TP}{TP+FN}$$

1. Un dataset contiene $\frac{1}{3}$ di positivi e $\frac{2}{3}$ di negativi. La recall del modello è di $\frac{2}{3}$. Che percentuale dei dati sono i falsi positivi?
 - a. $\frac{2}{9}$
 - b. $\frac{1}{9}$
 - c. $\frac{1}{3}$
 - d. Non può essere stabilito
2. Un dataset contiene $\frac{1}{3}$ di positivi e $\frac{2}{3}$ di negativi. La recall del modello è di $\frac{2}{3}$. Che percentuale dei dati sono i falsi negativi?
 - a. $\frac{2}{9}$
 - b. $\frac{1}{9}$
 - c. $\frac{1}{3}$
 - d. Non può essere stabilito

3. Un dataset contiene $2/3$ di positivi e $1/3$ di negativi. La precisione del modello è $9/10$. Che percentuale dei dati totali sono falsi positivi?
 - a. $1/9$
 - b. Non può essere stabilito
 - c. $2/27$
 - d. $1/10$

deep features

1. Cosa si intende con "deep" features?
 - a. Features sintetizzate in modo automatico a partire da altre features
 - b. Features ottenute mediante utilizzo di sensori ottici di profondità
 - c. Features relative a dati in 2 o più dimensioni
 - d. Features soggette a una approfondita supervisione da parte umana

distribuzione congiunta di probabilità

1. Selezionare la sentenza corretta relativa alla distribuzione congiunta di probabilità
 - a. Non permette il calcolo di eventi condizionali
 - b. Non permette di fare nessun tipo di predizione
 - c. Non consente una visione distinta delle singole features
 - d. Il suo calcolo presenta problemi di scalabilità all'aumentare delle features
2. Selezionare la sentenza errata relativa alla distribuzione congiunta di probabilità per N variabili aleatorie discrete
 - a. È la distribuzione di probabilità di tutte le possibili tuple di valori per le variabili
 - b. Richiede il calcolo di un numero esponenziale di parametri
 - c. Non permette il calcolo di probabilità condizionali tra le features
 - d. Consente il calcolo delle probabilità marginali delle singole features

Probabilità condizionata

1. Selezionare la sentenza corretta relativa alla probabilità condizionata $P(A|B)$ tra due eventi A e B
 - a. $P(A|B)$ è sicuramente maggiore o uguale di $P(A \text{ and } B)$
 - b. $P(A|B)$ è sicuramente maggiore o uguale di $P(A)$
 - c. $P(A|B)$ è sicuramente minore o uguale di $P(A \text{ and } B)$
 - d. $P(A|B)$ è sicuramente minore o uguale a $P(A)$

entropia / crossentropy

1. Se un modello calcola una distribuzione di probabilità, aggiungere alla funzione obiettivo una componente tesa a diminuire l'entropia avrà l'effetto di:
 - a. Nessun effetto concreto
 - b. Focalizzare le scelte sui casi più probabili
 - c. Ridistribuire le probabilità in modo più bilanciato tra tutti i casi
 - d. Favorire l'uscita da minimi locali
2. Se un modello calcola una distribuzione di probabilità, aggiungere alla funzione obiettivo una componente tesa ad aumentare l'entropia avrà l'effetto di:
 - a. Nessun effetto concreto
 - b. Focalizzare le scelte sui casi più probabili
 - c. Ridistribuire le probabilità in modo più bilanciato tra tutti i casi
 - d. Favorire l'uscita da minimi locali
3. Selezionare la sentenza corretta relativa alla crossentropy $H(P,Q)$ tra P e Q

- a. È uguale alla divergenza di kullback-Leibler $KL(P,Q)$ più l'entropia $H(P)$ di P
 - b. Misura la loglikelihood di P data la distribuzione di Q
 - c. Ha un valore massimo quando $P=Q$
 - d. È una funzione simmetrica $H(P,Q)=H(Q,P)$
4. Selezionare la sentenza erronea relativa alla crossentropy $H(P,Q)$ tra P e Q
- a. È uguale alla divergenza di kullback-Leibler $KL(P,Q)$ più l'entropia $H(P)$ di P
 - b. Misura la loglikelihood di Q data la distribuzione di P
 - c. Ha un valore minimo quando $P=Q$
 - d. È una funzione simmetrica $H(P,Q)=H(Q,P)$
5. Selezionare la sentenza errata relativa all'entropia per la distribuzione di probabilità di una variabile aleatoria discreta
- a. Il suo valore è minimo (e uguale a 0) quando la probabilità è tutta concentrata in una classe
 - b. Il range del suo valore è tra 0 e $\log n$ dove n sono i possibili valori di X
 - c. È una misura del grado di disordine della variabile aleatoria
 - d. Il suo valore è minimo (e uguale a 0) quando la probabilità è equamente distribuita tra tutte le classi
6. Il range dell'entropia per la distribuzione di probabilità di una variabile aleatoria discreta è
- a. Tra 0 e 1
 - b. Tra 0 e $\log n$ dove n sono i possibili valori di x
 - c. Tra 0 e infinito
 - d. Tra -1 e 1
7. Una variabile aleatoria discreta con valori a,b e c ha la seguente distribuzione di probabilità: $P(a)=1/4$, $P(b)=1/2$, $P(c)=1/4$. Qual è la sua entropia?
- a. $3/2$
 - b. $\log(3)$
 - c. $4/5$
 - d. $5/4$
8. Siano date le seguenti distribuzioni di probabilità P e Q: $P(0)=3/8$, $P(1)=1/2$, $P(2)=1/8$ and $Q(0)=1/2$, $Q(1)=1/4$, $Q(2)=1/4$. Quanto vale la crossentropy $H(P|Q)$ tra P e Q?
- a. $13/8$
 - b. $3/2 + \log(3)$
 - c. $5/2 - \log(3)/2$
 - d. 2

funzione logistica

1. Selezionare la sentenza errata relativa alla derivata della funzione logistica
 - a. È una funzione monotona
 - b. Tende a 0 quando x tende a -inf
 - c. Ha il suo massimo in corrispondenza dello 0
 - d. È una funzione simmetrica
2. La derivata della funzione logistica $\delta(x)$ è
 - a. $\delta(x) / \delta(1-x)$
 - b. $\delta(x) * (1 - \delta(x))$
 - c. $\delta(x) / (1 - \delta(x))$
 - d. $\delta(x) * \delta(1-x)$

funzione loss in rete neurale

1. Quale funzione di loss è tipicamente utilizzata in una rete neurale per classificazione binaria che utilizza una sigmoid come attivazione finale?
 - a. Categorical crossentropy
 - b. Absolute error
 - c. **Binary crossentropy**
 - d. Mean squared error
2. Quale funzione di loss è tipicamente utilizzata in una rete neurale per classificazione a categorie multiple che utilizza softmax come attivazione finale?
 - a. Binary crossentropy
 - b. **Categorical crossentropy**
 - c. Absolute error
 - d. Mean squared error

GAN

1. Selezionare la sentenza corretta:
 - a. Una GAN è una rete che permette di generare attacchi per un qualunque modello predittivo
 - b. Le GAN hanno una struttura encoder-decoder simile a quella di un autoencoder
 - c. **Le GAN possono soffrire del fenomeno di "mode collapse" cioè la tendenza a focalizzare la generazione su un unico o pochi esempi**
 - d. Le GAN basano il loro training su una funzione di loglikelihood relativa ai dati generali

layer convolutivo 2D

(Width x Height x Depth)

- Number of filters K (kernels)
- Filter size (spatial) F (3,3)
- Stride at which filters move at S
- Zero padding P

The formula for the output shape is given as

1. $W_{new} = (W - F + 2*P)/S + 1$
2. $H_{new} = (H - F + 2*P)/S + 1$
3. $D_{new} = K$

$$(16-3)/2 + 1 = 6.5 \rightarrow 7$$

1. Il tensore di input di un layer convolutivo 2D ha dimensione (16,16,32). Sintetizzo 8 kernel con dimensione spaziale (3,3), stride 2, nessun padding (valid mode). Quale sarà la dimensione dell'output?
 - a. (7,7,15)
 - b. (8,8,8)
 - c. **(7,7,8)**
 - d. (8,8,32)
2. Il tensore di input di un layer convolutivo 2D ha dimensione (32,32,8). Sintetizzo un unico kernel con dimensione spaziale (4,4), stride 2, nessun padding (valid mode). Quale sarà la dimensione dell'output?

- a. (16,16,1)
 - b. (16,16,8)
 - c. (15,15,8)
 - d. (15,15,1)
3. Il tensore di input di un layer convolutivo 2D ha dimensione (16,16,8). Sintetizzo 4 kernel con dimensione spaziale (5,5), stride 2, nessun padding (valid mode). Quale sarà la dimensione dell'output?
- a. (6,6,4)
 - b. (8,8,8)
 - c. (7,7,4)
 - d. (7,7,8)

Number of parameters in Keras Conv2D layer is calculated using the following equation:

$$\text{number_parameters} = \text{out_channels} * (\text{in_channels} * \text{kernel_h} * \text{kernel_w} + 1) \#$$

1 for bias

So, in your case,

```
in_channels = 3
out_channels = 16
kernel_h = kernel_w = 3
number_parameters = 16(3*3*3 + 1) = 448
```

4. Un layer convolutivo 2D con stride 1, kernel size 3x3, e senza padding prende in input un layer con dimensioni (32,32,3) e restituisce un layer di dimensione (32,32,16). Quanti sono i suoi parametri?
- a. 160
 - b. 28
 - c. 432
 - d. 448
5. Il numero dei parametri di un layer convolutivo dipende da:
- a. Unicamente dalle dimensioni dei layer di input e di output
 - b. Lo stride dei kernel e di tutte le dimensioni di input e output, compresi i canali
 - c. La dimensione spaziale dei kernel e il numero dei canali di input e output
 - d. Lo stride dei kernel e le dimensioni spaziali di input e output
6. Un layer convolutivo 2D con stride 1, kernel size 1x1, e senza padding prende in input un layer con dimensioni (32,32,16) e restituisce un layer di dimensione (32,32,4). Quanti sono i suoi parametri?
- a. 2
 - b. 68
 - c. 8
 - d. 64

Long-short term memory models (LSTMs)

1. Selezionare la sentenza scorretta relativa ai Long-Short Term Memory Models (LSTMs):
- a. Utilizzano delle particolari porte (gates) per gestire l'evoluzione della cella di memoria durante l'elaborazione di una sequenza di dati
 - b. Sono una particolare tipologia di Rete Ricorrente
 - c. Sono prevalentemente utilizzati per l'elaborazione di sequenze di dati
 - d. Sono prevalentemente utilizzati per la segmentazione di immagini mediche

2. I long-short term memory models (LSTMs) sono modelli utilizzati prevalentemente per:
 - a. Segmentazione per immagini mediche
 - b. **Elaborazione di sequenze di dati**
 - c. Predire traiettorie per agenti a guida autonoma
 - d. Elaborazione di immagini

MaxPooling (derivata)

1. Qual è la derivata della funzione di MaxPooling?
 - a. L'identità
 - b. Non è una funzione derivabile
 - c. **1 in corrispondenza del massimo e 0 altrove**
 - d. 1 ovunque

Minibatch

1. Qual è l'effetto tipico dell'aumento della dimensione del minibatch durante il training?
 - a) La backpropagation è effettuata più frequentemente e l'aggiornamento dei parametri è più accurato
 - b) La backpropagation è effettuata più frequentemente ma l'aggiornamento dei parametri è meno accurato
 - c) La backpropagation è effettuata meno frequentemente e l'aggiornamento dei parametri è meno accurato
 - d) **La backpropagation è effettuata meno frequentemente ma l'aggiornamento dei parametri è più accurato**
2. Qual è l'effetto tipico della riduzione della dimensione del minibatch durante il training?
 - a) La backpropagation è effettuata più frequentemente e l'aggiornamento dei parametri è più accurato
 - b) **La backpropagation è effettuata più frequentemente ma l'aggiornamento dei parametri è meno accurato**
 - c) La backpropagation è effettuata meno frequentemente e l'aggiornamento dei parametri è meno accurato
 - d) La backpropagation è effettuata meno frequentemente ma l'aggiornamento dei parametri è più accurato

modelli generativi

1. Selezionare la sentenza scorretta riguardo i modelli generativi
 - a. Un tipico esempio di tecnica generativa è Naive Bayes
 - b. Sono modelli che cercano di apprendere la distribuzione di probabilità dei dati
 - c. Generative adversarial networks, variational autoencoders e diffusion models sono esempi di tecniche generative profonde
 - d. **Sono modelli meta-teorici rivolti alla automatizzazione della generazione di reti neurali**
2. Con modelli generativi si intende:
 - a. Il processo di automatizzazione della generazione di reti neurali
 - b. **Modelli che cercano di apprendere la distribuzione di probabilità dei dati**
 - c. L'uso di attacchi avversariali allo scopo di aumentare la robustezza di modelli
 - d. L'applicazione di tecniche genetiche al deep learning

mutua informazione (information gain)

1. Selezionare la risposta scorretta relativa alla mutua informazione $I(X,Y)$ tra due variabili aleatorie X e Y (anche detta Information Gain nel contesto degli alberi di decisione)
 - a. È una funzione simmetrica $I(X,Y)=I(Y,X)$
 - b. **Coincide con l'entropia $H(Y|X)$ di Y dato X**
 - c. Può essere utilizzata per guidare la selezione degli attributi durante la costituzione di un albero di decisione
 - d. Misura il guadagno di informazione su Y dopo aver osservato X

Naive Bayes

1. Selezionare la sentenza errata relativa alla tecnica Naive Bayes
 - a. È una tecnica di tipo generativo in quanto cerca di determinare la distribuzione delle varie categorie dei dati
 - b. Deriva dall'ipotesi teorica semplificativa che le features sono indipendenti tra loro, date le classi
 - c. **Non può essere utilizzata se le features non sono tra loro indipendenti, date le classi**
 - d. Fornisce un modo computazionalmente efficiente per approssimare la distribuzione congiunta di probabilità delle features
2. Perché la tecnica Naive Bayes è detta "Naive" (ingenua)?
 - a. **Perché suppone ingenuamente che le features siano indipendenti tra loro, date le classi**
 - b. Perché suppone ingenuamente che i dati di training rispecchino i dati reali
 - c. Perché fornisce un modo semplice ma preciso di calcolare la distribuzione congiunta di probabilità delle features
 - d. Perché suppone ingenuamente che la teoria possa avere applicazioni pratiche
3. Avendo 5 categorie di dati e 3 features di input booleane, quanti parametri indipendenti devono essere stimati secondo la tecnica Naive Bayes (compresi i priors)
 - a. 15
 - b. 16
 - c. **19**
 - d. 20

neuroni artificiali

1. Selezionare la sentenza scorretta relativa ai neuroni artificiali
 - a. Un neurone artificiale tipicamente calcola una combinazione lineare dei suoi input, seguita dalla applicazione di una funzione di attivazione non lineare
 - b. Il numero dei parametri di un neurone artificiale è lineare nel numero dei suoi input
 - c. Un neurone artificiale definisce un semplice modello matematico che simula il neurone biologico
 - d. **Un neurone artificiale può apprendere qualunque funzione dei suoi input**
2. Selezionare la sentenza corretta
 - a. Il numero dei parametri di un neurone artificiale è quadratico nella dimensione dei suoi input

- b. Un neurone artificiale tipicamente calcola una combinazione lineare dei suoi input, seguita dalla applicazione di una funzione di attivazione non lineare
- c. Un neurone artificiale può apprendere qualunque funzione dei suoi input
- d. Un neurone artificiale può apprendere solo funzioni lineari

Overfitting

1. Quale delle seguenti tecniche non può essere utilizzata per contrastare l'overfitting?
 - a. Early stopping
 - b. Data augmentation
 - c. Introduzione di dropout layers
 - d. Aggiunta di skip connections
2. Quale delle seguenti situazioni non è particolarmente problematica dal punto di vista dell'overfitting?
 - a. Avere pochi dati di training
 - b. Avere dati molto rumorosi
 - c. Disporre di un modello molto espressivo
 - d. Effettuare un training molto prolungato

regressione logistica

1. Selezionare la sentenza scorretta riguardo alla regressione logistica
 - a. Permette di associare una probabilità alla predizione della classe
 - b. Il calcolo della predizione si basa sulla loglikelihood dei dati di training
 - c. La predizione dipende dal bilanciamento dei dati di training rispetto alle classi
 - d. I parametri del modello possono essere tipicamente calcolati in forma chiusa, mediante una forma esplicita
2. Selezionare la sentenza errata riguardo alla regressione logistica
 - a. Si basa su una combinazione lineare delle features in input
 - b. La probabilità della predizione cresce se ci si allontana dalla superficie di confine tra le classi
 - c. Non dipende dal bilanciamento dei dati di training rispetto alle classi
 - d. Nel caso di classificazione binaria la superficie di confine tra le classi è un iperpiano
3. In quali di questi casi la regressione logistica potrebbe essere in difficoltà?
 - a. Quando non tutte le features di input sono rilevanti ai fini della classificazione
 - b. Quando esiste una elevata correlazione tra le features
 - c. Quando la classificazione dipende da un confronto tra le features
 - d. Quando le features sono indipendenti tra loro, data la classe
4. Selezionare la sentenza corretta riguardo la regressione logistica
 - a. I parametri del modello sono tipicamente calcolati mediante discesa del gradiente
 - b. La predizione non dipende dal bilanciamento dei dati di training rispetto alle classi
 - c. I parametri del modello possono essere tipicamente calcolati in forma chiusa, mediante una formula esplicita
 - d. Il calcolo della predizione non si basa sulla loglikelihood dei dati di training, in quanto si tratta di una tecnica discriminativa

regressione multinomiale

1. Riguardo alla regressione multinomiale , selezionare la sentenza corretta tra le seguenti:
 - a. Per n features di input e m classi, il numero dei parametri del modello cresce come $O(n+m)$
 - b. **Il peso con cui è valutata ogni feature è tipicamente diverso per ogni classe**
 - c. I pesi delle features sono sempre tutti positivi, i bias possono essere negativi
 - d. Per ogni input, esiste almeno una classe con probabilità >0.5
2. Riguardo alla regressione multinomiale , selezionare la sentenza errata tra le seguenti:
 - a. Per n features di input e m classi, il numero dei parametri del modello è $n \times m + m$
 - b. Il peso con cui è valutata ogni feature è tipicamente diverso per ogni classe
 - c. Il peso delle features indica la loro importanza ai fini della classificazione
 - d. **Per ogni input, esiste almeno una classe con probabilità >0.5**

Regressione lineare

1. Selezionare la sentenza errata relativa alla regressione lineare
 - a. **Cerca di determinare un iperpiano di separazione tra due categorie di dati**
 - b. Il problema di ottimizzazione ammette una soluzione in forma chiusa
 - c. La funzione di loss è tipicamente una distanza quadratica tra i valori predetti e quelli osservati
 - d. cerca di stabilire una relazione tra i valori di una variabile di output e i valori di una o più features di input

reti per classificazione di immagini

1. Quale di queste reti non è stata progettata per classificare immagini?
 - a. Inception - v3
 - b. **U-Net**
 - c. VGG19
 - d. ResNet
2. Quale è la tipica struttura per una rete neurale di classificazione delle immagini?
 - a. Solo livelli densi
 - b. Un encoder seguito da un decoder
 - c. **Una sequenza alternata di convoluzioni e downsampling seguita da flattening e pochi livelli densi finali**
 - d. Una sequenza di convoluzioni che preservano la dimensione spaziale dell'input

ReLU(x) (Rectified linear unit)

1. Selezionare la sentenza errata relativa alla funzione ReLU(x) (rectified linear unit)
 - a. Lei o le sue varianti sono tipicamente utilizzate per i livelli interni delle reti neurali profonde
 - b. La sua derivata è una funzione a gradino
 - c. È una funzione monotona non decrescente
 - d. **Non può essere usata per layer convoluzionali**

scomparsa del gradiente (vanishing gradient)

1. Selezionare la sentenza scorretta relativa al problema della scomparsa del gradiente (vanishing gradient)
 - a. **Se il gradiente tende a zero anche i parametri e le attivazioni dei neuroni tendono a zero**
 - b. Se il gradiente tende a zero i parametri non sono più aggiornati e la rete smette di apprendere
 - c. Il problema è mitigato dall'uso di link residuali all'interno della rete
 - d. Il problema è fortemente attenuato dall'uso di ReLU (o sue varianti) come funzione di attivazione per i livelli nascosti della rete
2. Il problema della scomparsa del gradiente (vanishing gradient) si riferisce ad una progressiva diminuzione dell'intensità del gradiente dovuta a
 - a. **Backpropagation in reti profonde**
 - b. Dati troppo rumorosi o malamente processati
 - c. Troppi pochi dati di training a disposizione
 - d. Training eccessivamente lungo

Softmax

1. Selezionare la sentenza corretta relativa alla funzione softmax
 - a. Non può essere utilizzata nel caso di una classificazione binaria
 - b. **Restituisce una distribuzione di probabilità sulle classi**
 - c. Per una data classe, la somma dei valori su tutti gli input di un minibatch è sempre 1
 - d. Produce valori compresi nell'intervallo $[-1,1]$
2. Selezionare la sentenza errata relativa alla funzione softmax
 - a. Generalizza la funzione logistica al caso multiclasse
 - b. Permette di calcolare una distribuzione di probabilità sulle classi
 - c. Per un dato input, la somma dei suoi valori su tutte le classi è sempre 1
 - d. **Produce valori compresi nell'intervallo $[-1,1]$**

stride in un layer convolutivo

1. Qual è l'effetto di uno stride non unitario (>1) in un layer convolutivo?
 - a. **La dimensione spaziale diminuisce**
 - b. Nessun effetto spaziale, il numero dei canali decresce
 - c. La dimensione spaziale aumenta
 - d. Nessun effetto spaziale, il numero dei canali aumenta

tecniche discriminative

1. Selezionare la sentenza corretta relativa alle tecniche discriminative
 - a. Cercano di determinare le distribuzioni di probabilità delle varie classi di dati
 - b. **Si focalizzano sulla definizione delle frontiere di decisione (decision boundaries)**
 - c. Si applicano per lo più in ambito di apprendimento non supervisionato
 - d. Sono tipicamente meno espressive delle tecniche generative
2. Cosa si intende con tecniche discriminative?
 - a. Tecniche tipiche di unsupervised learning che tentano di separare dati in cluster distinti
 - b. **Tecniche di classificazione che si focalizzano sulla definizione delle frontiere di decisione (decision boundaries)**
 - c. Tecniche che cercano di discriminare i dati in base alle diverse distribuzioni di probabilità delle varie classi

- d. tecniche che cercano di identificare gli outliers all'interno dei data set

tecnica a discesa del gradiente

1. selezionare la sentenza corretta relativa alla tecnica a discesa del gradiente
 - a. permette sempre di individuare il minimo globale, se questo esiste
 - b. il risultato non dipende dalla inizializzazione dei parametri del modello
 - c. può essere applicata solo se la funzione da minimizzare ha una superficie concava
 - d. potrebbe convergere ad un minimo locale
2. Selezionare la sentenza scorretta relativa alla tecnica a discesa del gradiente
 - a. Potrebbe convergere ad un minimo locale
 - b. Può essere applicata solo se la funzione da minimizzare ha una superficie concava
 - c. Il risultato può dipendere dalla inizializzazione dei parametri del modello
 - d. È opportuno decrementare il learning rate verso la fine dell'apprendimento

transposed convolutions

1. Selezionare la sentenza errata relativa alle transposed convolutions
 - a. Possono essere interpretate come convoluzioni normali con stride sub-unitario
 - b. Sono prevalentemente utilizzate in architetture per Image-to-Image processing, come autoencoders o U-Nets
 - c. Sono essenzialmente equivalenti alla applicazione di un livello di upsampling seguito da una convoluzione normale
 - d. Richiedono la trasposizione dell'input prima di calcolare la convoluzione dei kernel

U-net

1. Selezionare la sentenza scorretta relativa alla U-Net
 - a. È un componente tipico dei modelli generativi a diffusione
 - b. È spesso impiegata per problemi di segmentazione semantica di immagini
 - c. Può essere usata per la rimozione del rumore (denoising) di immagini
 - d. Viene spesso utilizzata nell'ambito della classificazione dei generi musicali
2. Quale tra i seguenti è un tipico campo di applicazione della U-Net?
 - a. Segmentazione semantica
 - b. Generazione musicale
 - c. Object detection
 - d. Natural Language Processing

Inception module

1. Selezionare la sentenza errata relativa all' "inception module"
 - a. Sfrutta kernel di dimensione diversa
 - b. Tende a ridurre il costo computazionale sfruttando convoluzioni unitarie per diminuire il numero dei canali
 - c. Utilizza al proprio interno delle skip-connections per bypassare l'applicazione di parte dei kernel
 - d. È un componente tipico della rete Inception-v3

Minimi locali – fase training

1. Quale delle seguenti tecniche non può aiutare ad uscire dai minimi locali durante la fase di training?
 - a. Ridurre la dimensione del minibatch
 - b. Fare clipping del gradiente in un range prefissato
 - c. Aumentare il learning rate
 - d. Aggiungere un “momento” al gradiente, cioè parte del gradiente del passo precedente

Intersection over Union (IoU)

1. Selezionare la sentenza SCORRETTA relativa alla Intersection over Union (IoU)
 - a. E' frequentemente utilizzata come misura di similitudine tra bounding boxes
 - b. Restituisce un valore nel range [0,1]
 - c. Non è una funzione simmetrica dei suoi input
 - d. E' una metrica principalmente utilizzata nel campo della Object Detection

Transformers

Quale è l'obiettivo principale dell'algoritmo di clustering K-means?

- a. Ridurre il numero di clusters al minimo
- b. Trovare il punto medio del dataset
- c. Raggruppare i punti di un cluster attorno al loro centroide ✓
- d. ottimizzare il numero dei clusters basandosi sulla distribuzione Gaussiana dei dati

Quale è lo scopo dell'optimizer in Tensorflow/Keras?

- a. definire l'algoritmo che calcola i gradienti della loss e aggiorna i pesi del modello ✓
- b. salvare i migliori pesi del modello durante il processo di training
- c. contrastare l'overfitting
- d. aggiungere una penalità ai pesi del layer su cui viene istanziato

Selezionare la sentenza errata relativa ai transformers

- a. Hanno una tipica struttura encoder-decoder, ognuno formato da uno stack di sotto-componenti modulari
- b. Sono alla base delle reti della famiglia BERT e GPT
- c. Utilizzano pesantemente il meccanismo di attenzione
- d. Aggiungono ad ogni livello della rete un encoding posizionale per enfatizzare la posizione relativa dei tokens

Selezionare la sentenza SCORRETTA relativa all'overfitting

- a. Può essere particolarmente pericolosa per modelli altamente espressivi
- b. L'acquisizione di nuovi dati di training non può che peggiorare la situazione ✓
- c. Può essere contrastata con tecniche di regolarizzazione
- d. Può essere contrastata con la tecnica di early stopping durante la fase di training

Un training set è composto da 10000 dati. Se la batchsize è 50, quante volte verrà effettuata la backpropagation durante una singola epoca?

- a. 50
- b. 10000
- c. 200 ✓
- d. Nessuna delle altre risposte è corretta

Selezionare la sentenza errata relativa ai modelli a diffusione

- a. Il risultato è ottenuto attraverso passaggi multipli lungo una singola rete, tipicamente una Unet.
- b. Generano i risultati attraverso un processo di diffusione della informazione
- c. I modelli a diffusione iterano una operazione di denoising
- d. Tipicamente, lo spazio latente ha la stessa dimensione dello spazio visibile ✘

Risposta errata.

La risposta corretta è:

Generano i risultati attraverso un processo di diffusione della informazione

Nella Principal Component Analysis, cosa indica la varianza spiegata da ciascuna componente?

- a. indica la correlazione tra ciascuna componente e la variabile target
- b. indica la deviazione standard della componente rispetto alla media dei dati
- c. indica quanto una singola componente contribuisce alla varianza complessiva dei dati ✔
- d. misura la somma cumulativa delle differenze quadratiche tra le osservazioni e i valori predetti dalla componente