

- 1) Quali sono i parametri di una distribuzione multinormale? Per ciascun parametro specificare la dimensionalità.
- 2) Nell'ambito dei multi-classificatori che cosa è e come funziona il Borda count?
- 3) Che cosa si intende per clustering? Fare esempi di applicazioni.
- 4) Cosa si intende per K-fold cross-validation? Quali sono i vantaggi rispetto a un semplice split a due dei dati di training?
- 5) Nell'ambito dell'apprendimento automatico cosa si intende per generalizzazione e overfitting?
- 6) Cosa si intende per multi-classificatore? Quando un multi-classificatore è efficace?
- 7) Descrivere a grandi linee l'algoritmo di Clustering K-means.
- 8) Fare esempi pratici di ragionamento induttivo e deduttivo
Esempio di ragionamento induttivo: A è un uomo, A muore => gli uomini muoiono.
Ovvero ho C1 e R2, e posso quindi elaborare una teoria di R.
Esempio di ragionamento deduttivo: Gli uomini muoiono, A è un uomo => prima o poi A morirà. Cioè ho la regola generale R e un caso C1, e posso sapere il risultato R1.
- 9) Qual è l'idea di base dell'algoritmo backpropagation per l'addestramento di reti neurali?
- 10) Descrivere a grandi linee un classificatore Random Forest.
- 11) Qual è l'obiettivo di una tecnica di regressione?
- 12) Indicare le differenze tra le tecniche di riduzione di dimensionalità PCA e LDA.
- 13) Cosa si intende per future reward nell'ambito del reinforcement learning? Fare un esempio.
- 14) Nell'ambito dei multi-classificatori come si può ottenere indipendenza tra i singoli classificatori utilizzati?
- 15) Che cosa si intende con tokenizzazione negli LLM? Perché si preferisce usare token invece di singoli caratteri o intere parole?
- 16) Indicare le principali "stagioni" nello sviluppo dell'intelligenza artificiale e machine learning.
Tra il 1940-1974 si ha la nascita e gli anni d'oro, in cui si hanno i primi risultati nell'ambito del symbolic reasoning, del problem solving e del natural language processing.
Tra il 1974-1980 si ha il primo inverno, in cui c'è una riduzione dei finanziamenti dovuta alla mancanza di risultati all'altezza delle aspettative.
1980-1987 nuova primavera, nascono sistemi esperti, l'algoritmo di Backpropagation e finanziamenti per la quinta generazione di calcolatori.
1987-1993 secondo inverno, la quinta generazione floppa, vengono nuovamente interrotti i finanziamenti, le reti neurali non scalano a problemi complessi
1993-2011 tempi moderni, con hardware più potenti, vengono studiate le reti bayesiane, gli agenti intelligenti, classificatori come SVM, multi classificatori come random forest e adaboost, gli HMM, modelli come Deep Blue e Watson, successi dell'IA in molte discipline.
2011-oggi deep learning
- 17) Che cosa codifica la funzione Q nell'ambito dell'approccio Q-learning? Quali sono gli input che ne determinano il valore?
- 18) In classificazione cosa si intende per superficie decisionale o di separazione? Riportare anche un esempio grafico.

- 19) Definire cosa si intende per apprendimento supervisionato e non supervisionato.
- 20) Rispetto a K-means l'approccio di clustering EM con Gaussian mixture quali maggiori flessibilità consente?
- 21) Nell'ambito delle reti neurali che cosa si intende per problema del vanishing gradient? Come può essere risolto?
- 22) La posizione e forma dell'ellissoide di una distribuzione multinormale come è influenzata da μ e Σ ?
- 23) Come può essere matematicamente definita la loss function (su un singolo pattern) per l'addestramento di una rete neurale?
- 24) Con quali tecniche si può estendere SVM da 2 a più classi?
- 25) Nell'ambito di CNN, che cosa si intende con transfer learning? Quali sono le tecniche di transfer learning utilizzabili?
- 26) Qual è la funzione obiettivo in formato matriciale della multiple linear regression?
- 27) Cosa si intende per approccio parametrico e non-parametrico nell'ambito della classificazione? Fare un esempio di classificatore parametrico e non parametrico.
- 28) Che cosa sono i criteri di clustering? Fare un esempio.
- 29) Per l'addestramento di una rete neurale che cosa si intende con vettore di output desiderato? Come può essere definito? Come si può calcolare l'errore da retro-propagare a partire dal vettore desiderato e dal valore calcolato dalla rete per un pattern? Oltre alla spiegazione riportare un esempio.
- 30) Come è definita la funzione di attivazione Relu? Perché consente di addestrare reti neurali profonde limitando il problema del vanishing gradient?
- 31) Descrivere a grandi linee l'approccio di classificazione AdaBoost
- 32) Nell'ambito dei multi-classificatori quali sono le più comuni tecniche di fusione a livello di decisione e di confidenza?
- 33) Come si calcola l'attivazione (net) di un neurone artificiale? Indicare formula e commentarla.
- 34) Quali sono le condizioni necessarie affinché le tecniche di deep learning siano più efficaci di altri approcci di machine learning?
- 35) Cosa si intende per Clustering esclusivo e Clustering soft (o Fuzzy). Quest'ultimo che vantaggi può avere?
- 36) Indicare la formula di Bayes per la probabilità a posteriori, definendo i termini.
- 37) Cosa è possibile apprendere mediante tecniche di reinforcement learning? Fare un esempio.
- 38) Nell'ambito di classificazione con SVM cosa si intende per pattern linearmente separabili e non linearmente separabili? Fare esempio grafico dei due casi.
- 39) Qual è l'idea di base dell'algoritmo di clustering EM con Gaussian mixture?
- 40) Cosa si intende per iperparametri? Fornire esempi pratici di iperparametri. Come si ottimizzano?
- 41) Come opera un livello di pooling in una CNN?
- 42) Nel caso di pattern non-linearmente separabili, nella formulazione di SVM lineare come si approccia il problema?
- 43) Quali sono i più noti algoritmi di clustering?
- 44) Fare esempi pratici di pattern numerici, categorici e di sequenze.
- 45) Perché le recenti reti neurali deep sono più efficaci delle MLP a tre livelli?

- 46) La formula di distanza di un pattern dall'iperpiano risultante dal training di un SVM dipende da tutti i pattern del training set o solo da una parte di questi? Motivare la risposta.
- 47) Nel classificatore di Bayes cosa si intende per densità di probabilità condizionale e probabilità a priori
- 48) Dare la definizione di Training, Validation e Test set e discutere una possibile suddivisione dei dati nei tre insiemi.
- 49) Nell'ambito di CNN, che cosa si intende per connessioni locali e condivisione di pesi?
- 50) Cosa si intende con SVM lineari? Cosa sono le superfici di separazione nel caso $d=2$ e $d=3$?
- 51) Come si imposta un problema di multiple linear regression? Come sono popolati X , y e β ?
- 52) Qual è l'obiettivo delle tecniche di riduzione di dimensionalità?
- 53) Che cosa denota la matrice Σ nella definizione della distribuzione multinormale?
- 54) Nelle SVM non lineari cosa si intende per kernel? Quali sono i kernel più utilizzati?
- 55) Qual è la differenza sostanziale dell'approccio "On-line" rispetto a "SGD con mini-batch" per il training di reti neurali?
- 56) Descrivere le principali criticità e limitazioni dell'algoritmo di Clustering K-means.
- 57) Come classifica i pattern un classificatore K-NN? Discutere la complessità computazionale.
- 58) Nel classificatore SVM cosa sono i support vectors?
- 59) Cosa si intende per convergenza di un algoritmo di apprendimento iterativo? Accuratezza e loss come si comportano durante le iterazioni in caso di convergenza? Disegnare un semplice grafico.
- 60) Nella regressione cosa si intende per variabile indipendente e variabile dipendente?
- 61) Cosa si intende per funzione obiettivo e loss function?
- 62) Quando una rete neurale si definisce deep (profonda)?
- 63) Indicare le differenze tra reti neurali feedforward e le reti neurali ricorrenti, disegnando un esempio di entrambe.
- 64) Nella regressione lineare (sia rispetto ai parametri sia rispetto alla variabile indipendente) i dati con cosa sono approssimati nel caso 2D e 3D?
- 65) Nell'ambito dell'apprendimento automatico quali sono le principali cause di overfitting?
- 66) Quanti sono i parametri indipendenti di una distribuzione multinormale nel caso 3-dimensionale? Motivare la risposta.
- 67) Quali sono le più comuni funzioni di attivazione utilizzate per neuroni artificiali? Perché è necessario che siano non-lineari e differenziabili (esistenza derivata)?
- 68) Quali sono le più note tecniche di riduzione di dimensionalità? Quali i loro tipici utilizzi?
- 69) Dal punto di vista pratico nell'approccio di Parzen che differenza c'è se si usa una funzione finestra ipercubo piuttosto che multinormale?

- 70) Nella formulazione dell'SVM lineare la funzione obiettivo richiede di massimizzare il margine. L'ottimizzazione è però vincolata; in cosa consistono i vincoli? quanti sono?
- 71) Cosa si intende per risoluzione dei problemi con approccio "forza bruta". Si tratta di intelligenza artificiale?
La "Brute Force" o "ricerca esaustiva" è una tecnica di risoluzione di problemi che consiste nell'enumerare e valutare tutte le possibili alternative nel problema, scartando quelle non accettabili. Ha un costo computazionale molto elevato, pertanto spesso non è computazionalmente gestibile. Non si tratta di AI, anche se alcune tecniche di ricerca esaustiva, non aventi capacità di ragionamento e comprensione, prendono il nome di "weak AI".
- 72) Come opera un livello di convoluzione di una CNN?
- 73) La densità locale di pattern con quali metodi (parametrici e non-parametrici) può essere stimata? Fare un esempio.
- 74) Definire i problemi di Classificazione e Regressione evidenziandone le differenze e fornendo per ciascuno esempi reali della loro applicazione.
- 75) Come può essere scelto nella pratica il numero di cluster in un algoritmo di clustering come K-means?
- 76) Quali sono le limitazioni di Q learning per risolvere problemi complessi? Le tecniche di deep learning possono essere di aiuto in questo caso? Come?
- 77) Nell'ambito dei multi-classificatori che cosa è e come funziona il Borda count?
- 78) Cosa si intende per multi-classificatore? Quando un multi-classificatore è efficace?
- 79) Che cos'è il learning rate nell'ambito dell'apprendimento di reti neurali? Cosa succede se viene scelto un learning rate troppo piccolo o troppo grande?
- 80) Quali sono i parametri di una distribuzione multinormale?
- 81) Descrivere a grandi linee un classificatore Random Forest
- 82) La posizione e forma dell'ellissoide di una distribuzione multinormale come è influenzata da μ e Σ ?
- 83) Nell'ambito dei multi-classificatori come si può ottenere indipendenza tra i singoli classificatori utilizzati?
- 84) Nell'ambito delle reti neurali, quali sono le principali differenze di CNN rispetto a MLP?
- 85) Qual è il principio su cui si basa il classificatore SVM? Cosa si intende per margine?
- 86) Cosa si intende per convergenza di un algoritmo di apprendimento iterativo? Accuratezza e loss come si comportano durante le iterazioni in caso di convergenza. Disegnare un semplice grafico.
- 87) Per l'addestramento di una rete neurale che cosa si intende con vettore di output desiderato? Come può essere definito? Come si può calcolare l'errore da retro-propagare a partire dal vettore desiderato e dal valore calcolato dalla rete per un pattern?
- 88) Nel classificatore di Bayes cosa si intende per densità di probabilità condizionale e probabilità a priori.
- 89) Come si misurano le prestazioni di un classificatore?
- 90) Qual'è la funzione obiettivo in formato matriciale della multiple linear regression?

- 91) Cosa si intende con SVM lineari? Cosa sono le superfici di separazione nel caso $d=2$ e $d=3$?
- 92) Qual è l'idea di base dell'algoritmo di clustering EM con Gaussian mixture?
- 93) Quali sono le più comuni funzioni di attivazione utilizzate per neuroni artificiali? Perché è necessario che siano non-lineari e differenziabili (esistenza derivata) ?
- 94) Che cos'è il learning rate nell'ambito dell'apprendimento di reti neurali? Cosa succede se viene scelto un learning rate troppo piccolo o troppo grande?
- 95) Nella formulazione dell'SVM lineare la funzione obiettivo richiede di massimizzare il margine. L'ottimizzazione è però vincolata; in cosa consistono i vincoli? quanti sono