

Il numero a fianco della domanda rappresenta il numero di volte che è stata posta. Quando esso non è presente, significa che le domande sono state poste solo una volta.

0.1 Reinforcement Learning

0.1.1 Cosa si intende per Apprendimento con Rinforzo?

0.1.2 Quali sono gli attori?

0.1.3 Cosa rappresenta la critica?

0.1.4 Che tipo di architettura si può ipotizzare nell'apprendimento con rinforzo?

0.1.5 Condizionamento classico e condizionamento operante

0.1.6 Quale relazione c'è con l'intelligenza?

0.1.7 Come potreste illustrare: Exploration vs Exploitation?

0.1.8 Cos'è il credit assignment?

0.1.9 Cosa si intende per traccia e quale è il suo ruolo? [2]

0.1.10 Definire l'algoritmo di Q-learning, descrivendo le equazioni opportune.

0.1.11 Scrivere le equazioni dell'algoritmo Q-learning in cui si consideri anche la traccia. [2]

0.1.12 Cosa si intende per politica epsilon-greedy? Come entra nell'algoritmo di Q-learning?

0.1.13 Che differenza c'è tra Q-learning e SARSA?

0.1.14 Dato un problema a piacere si descriva uno degli algoritmi e mostrare due passaggi di addestramento

0.1.15 Quale criterio si sceglie per definire i Reward?

0.1.16 A quali elementi sono associati i reward? Allo stato? All'azione? Allo stato prossimo? Perché? [2]

0.1.17 Impostare un problema su griglia (apprendimento del percorso di un agente, con partenza ed arrivo prescelti + ostacoli). La griglia fornisce un reward, diverso da zero, in ogni transizione.

Definire chiaramente il problema, farne un modello definendo le variabili e le funzioni che le legano.

Scrivere un risultato possibile dei primi 2 passi di apprendimento del problema definito al punto precedente.

0.2 Fuzzy System

0.2.1 Definire i passi per costruire un sistema fuzzy. [2]

0.2.2 Cosa si intende per FAM? [2]

0.2.3 Una FAM memorizza numeri o preposizione logiche? Come? [2]

0.2.4 Definire un problema per FAM a piacere che involva almeno due variabili in ingresso e due in uscita. [2]

0.2.5 Definire tutti i componenti e calcolare l'uscita passo a passo per un valore di input a piacere [2]

0.3 Macchine e intelligenza

0.3.1 Descrivere il test di Turing [2]

0.3.2 Descrivere l'esperimento della stanza cinese [2]

0.3.3 Come mai è stato proposto il test di Turing? [2]

0.3.4 Come mai è stato proposto l'esperimento della scatola cinese? [2]

0.3.5 Cosa voleva dimostrare il test di Turing? [2]

0.3.6 Cosa voleva dimostrare l'esperimento della scatola cinese? [2]

0.3.7 Cosa si intende per ipotesi forte ed ipotesi debole dell'AI? [2]

0.3.8 Riportare almeno due elementi del contraddittorio sulle ipotesi su cui è basata l'ipotesi debole sull'AI [2]

0.3.9 Descrivere il "Brain prosthesis thought experiment" di Moravec e commentarlo. [2]

0.4 Statistica

0.4.1 Esercizio 1

In una città lavorano due compagnie di taxi: blue e verde, la maggior parte dei taxisti lavorano per la compagnia verde per cui si ha la seguente distribuzione di taxi in città: 85% di taxi verdi e 15% di taxi blu. Succede un incidente in cui è coinvolto un taxi. Un testimone dichiara che il taxi era blu. Era sera e buio, c'era anche un po' di nebbia ma il testimone ha una vista acuta, la sua affidabilità è stata valutata del 70%. Qual è la probabilità che il taxi fosse effettivamente blu? Quale deve essere l'affidabilità del testimone perché la probabilità che il taxi fosse effettivamente blu sia del 99%?

0.4.2 Esercizio 2

Lo strumento principe per lo screening per il tumore al seno è la radiografia (mammografia). Definiamo X la situazione della donna: X=sana, malata, che non conosciamo. Definiamo Y l'esito della mammografia: Y=positiva, negativa, che viene misurato. Sappiamo che la sensibilità della mammografia è intorno al 90% ($P(Y=\text{positiva} | X=\text{malata})$) e che la specificità sia anch'essa del 90% ($P(Y=\text{negativa} | X=\text{sana})$). Qual è la probabilità che l'esame dia risultato positivo ($P(Y=\text{positivo})$), sapendo che le donne malate sono lo 0,01% ($P(X=\text{malata}) = 0,01\%$)? Qual è la percentuale di donne che hanno uno screening positivo, di essere effettivamente malate?

0.4.3 Enunciare il teorema di Bayes

Data una partizione dello spazio degli eventi $A_1 \dots A_n$, vale che:

$$P(A_i|E) = \frac{P(E|A_i)P(A_i)}{\sum_{j=1}^n P(E|A_j)P(A_j)}$$

0.4.4 Discutere l'analisi di varianza per un sistema lineare

Svolgere un'analisi di varianza per un sistema lineare significa analizzare quanto la stima di un parametro possa variare nelle diverse misure dei dati relativi al problema. L'analisi consente di esaminare a matrice dei covarianti, misurare quanto varia una misura di una variabile al variare del rumore e misurare quanto covariano due misure di due variabilità. **L'indice di correlazione** di due variabili viene calcolato proprio per misurare quanto le variabili si trovino lungo una funzione.

0.4.5 Dimostrare che la stima ai minimi quadrati è equivalente alla stima a massima verosimiglianza nel caso di errore Gaussiano sui dati. Cosa fornisce? Come?

Scriviamo il logaritmo negativo della verosimiglianza:

$$P(y_1, \dots, y_n; n, b; x_1, \dots, x_n) = - \sum_{i=1}^n \ln \left\{ \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \left(\frac{y_i - mx_i - b}{\sigma} \right)^2 \right] \right\} \quad (1)$$

$$= - \sum_{i=1}^n \ln \left(\frac{1}{\sqrt{2\pi}\sigma} \right) - \sum_{i=1}^n \left[-\frac{1}{2} \left(\frac{y_i - mx_i - b}{\sigma} \right)^2 \right] \quad (2)$$

$$= - \sum_{i=1}^n \ln \left(\frac{1}{\sqrt{2\pi}\sigma} \right) + \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - mx_i - b)^2 \quad (3)$$

Massimizziamo la **likelihood** ponendo a zero le derivate prime rispetto a m :

$$\frac{\partial P(y_1, \dots, y_n; n, b; x_1, \dots, x_n)}{\partial m} = \frac{\partial}{\partial m} \left[- \sum_{i=1}^n \ln \left(\frac{1}{\sqrt{2\pi}\sigma} \right) + \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - mx_i - b)^2 \right] \quad (4)$$

$$= 0 + \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - mx_i - b)^2 2(-x_i) \quad (5)$$

$$= -\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - mx_i - b)^2 x_i \quad (6)$$

$$-\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - mx_i - b)^2 x_i = 0$$

$$\sum_{i=1}^n (y_i - mx_i - b)^2 x_i = 0$$

$$m \left[\sum_{i=1}^n (x_i^2) \right] + q \left[\sum_{i=1}^n (x_i) \right] = \left[\sum_{i=1}^n (y_i x_i) \right]$$

Figure 1: Prima equazione

Massimizziamo la **likelihood** ponendo a zero le derivate prime rispetto a q :

$$\frac{\partial P(y_1, \dots, y_n; n, b; x_1, \dots, x_n)}{\partial q} = \frac{\partial}{\partial q} \left[- \sum_{i=1}^n \ln \left(\frac{1}{\sqrt{2\pi}\sigma} \right) + \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - mx_i - b)^2 \right] \quad (7)$$

$$= 0 + \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - mx_i - b)^2 2(-1) \quad (8)$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - mx_i - b)^2 \quad (9)$$

$$\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - mx_i - b)^2 = 0$$

$$\sum_{i=1}^n (y_i - mx_i - b)^2 = 0$$

$$m \left[\sum_{i=1}^n (x_i) \right] + q \left[\sum_{i=1}^n (1) \right] = \left[\sum_{i=1}^n (y_i) \right]$$

Figure 2: Seconda equazione

Ponendo a sistema le equazioni così ottenute ottengo:

$$\begin{bmatrix} [\sum_{i=1}^n (x_i^2)] & [\sum_{i=1}^n (x_i)] \\ [\sum_{i=1}^n (x_i)] & n \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} [\sum_{i=1}^n (y_i x_i)] \\ [\sum_{i=1}^n (y_i)] \end{bmatrix}$$

Lo stesso problema visto dal punto di vista dei **minimi quadrati** è impostato nel seguente modo.

$$\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & n \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

L'obiettivo è trovare una x tale che $(Ax - b)^T(Ax - b)$ è minima (minimizzazione di residui). La soluzione si ottiene calcolando $A^T Ax = A^T b$.

$$A^T A = \begin{bmatrix} x_1 & \dots & x_n \\ 1 & \dots & n \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & n \end{bmatrix} = \begin{bmatrix} [\sum_{i=1}^n (x_i^2)] & [\sum_{i=1}^n (x_i)] \\ [\sum_{i=1}^n (x_i)] & n \end{bmatrix}$$

$$A^T b = \begin{bmatrix} x_1 & \dots & x_n \\ 1 & \dots & n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} [\sum_{i=1}^n (y_i x_i)] \\ [\sum_{i=1}^n (y_i)] \end{bmatrix}$$

Ovvero:

$$\begin{bmatrix} [\sum_{i=1}^n (x_i^2)] & [\sum_{i=1}^n (x_i)] \\ [\sum_{i=1}^n (x_i)] & n \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} [\sum_{i=1}^n (y_i x_i)] \\ [\sum_{i=1}^n (y_i)] \end{bmatrix}$$

Che è la stessa soluzione ottenuta per la stima a massima verosimiglianza. Queste metodologie offrono una stima dei parametri di una funzione tramite la minimizzazione dei residui. La soluzione è quella che minimizza lo scarto quadratico medio dei residui, ovvero è a minima varianza.

0.5 Apprendimento supervisionato

- 0.5.1 Definire l'algoritmo di apprendimento di una rete neurale con unità arbitrarie. [2]**
- 0.5.2 Definire la funzione obbiettivo utilizzata. [2]**
- 0.5.3 Come si utilizza la funzione obbiettivo nell'algoritmo di apprendimento. [2]**
- 0.5.4 Cosa si intende per apprendimento per epoche e per trial?**
- 0.5.5 Qual è il vantaggio di ciascuna delle modalità di apprendimento?**
- 0.5.6 Cosa si intende per training e test set? Perché mai vengono utilizzati? Quali problemi si vogliono evitare?**
- 0.5.7 Una rete neurale con unità sigmoidali e un modello parametrico? È lineare? Perché?**
- 0.5.8 Se i dati sono acquisiti senza errori, è una buona scelta aumentare di molto i parametri del modello in modo da garantirsi che l'errore sul training set vada a zero? Perché?**
- 0.5.9 Cosa si intende per un problema di regressione ed illustrare una possibile soluzione.**
- 0.5.10 Come funziona l'approssimazione incrementale multi-scala, cosa garantisce e quali vantaggi può avere?**

0.6 AI

0.6.1 Si descriva il funzionamento della Forward Search. Perché è considerato un template e non un algoritmo?

L'idea è quella di esplorare il grafo partendo dal nodo iniziale, provando a trovare la strada per arrivare ad uno stato di goal. Ad ogni step della ricerca un nodo può essere etichettato in 3 modi:

1. **Unvisited:** deve essere ancora visitato dall'algoritmo.
2. **Alive:** visitato, ma l'algoritmo deve ancora visitare i nodi direttamente raggiungibili da esso. I nodi alive sono raccolti in una coda di priorità a .
3. **Dead:** visitato, ed anche ogni nodo vicino è stato visitato.

È considerato un template e non un algoritmo perché non è specificato il criterio con cui ordinare a .

0.6.2 Si elenchino due possibili implementazioni di Forward Search elencandone proprietà, vantaggi e svantaggi.

Breadth first search

La coda a è gestita in modo FIFO (First-in First-out). I percorsi con $k+1$ azioni vengono valutati dopo che ogni percorso con k azioni è stato esplorato. Se viene trovato il percorso, è garantito che questo avrà il minor numero di azioni (percorso più breve). Funziona in un tempo $O(|V| + |E|)$. Per costruire l'albero di ricerca individua in tutti i figli con la stessa profondità e garantisce il percorso più breve. Inoltre è simmetrico.

Depth first search

La coda a è uno stack gestito con una politica LIFO (Last-in First-out). È un algoritmo più "aggressivo" e cerca prima soluzioni nei percorsi più lunghi. Funziona con un tempo $O(|V| + |E|)$. Ha il vantaggio di ignorare cammini sbagliati nel caso in cui si trovi immediatamente una soluzione. Non è sistematico su spazi infiniti.

0.7 Clustering

0.7.1 Cosa si intende per clustering? In quali famiglie vengono divisi?

0.7.2 Che relazione c'è tra clustering e classificazione e quali sono le criticità?

0.8 Biologia

- 0.8.1 Definire il neurone biologico evidenziandone le parti più significative per la trasmissione dell'informazione ed il loro comportamento.**
- 0.8.2 Descrivere il funzionamento complessivo del neurone biologico.**
- 0.8.3 Che differenza c'è tra neuroni motori, neuroni sensoriali ed inter-neuroni?**
- 0.8.4 Come viene trasmessa ed elaborata l'informazione da un neurone?**
- 0.8.5 Cos'è uno spike?**
- 0.8.6 Quali sono le aree corticali principali?**
- 0.8.7 Cos'è il codice di popolazione?**
- 0.8.8 Data un'area cerebrale è univoca la funzione implementata in quell'area?**
- 0.8.9 Cosa sono i mirror neurons? Quali implicazioni hanno per i sistemi intelligenti e l'apprendimento?**