www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

DOMANDA

Il dataset che devi analizzare consiste in 1,000 istanze, gli attributi di input sono; tre continui, due simmetrici, quattro asimmetrici. Non sono presenti dati mancanti.

Il tuo obiettivo è studiare similarità e differenze che caratterizzano le 1,000 istanze in tuo possesso, allo scopo di riassumere in modo efficace potenziali gruppi di istanze simili.

Da una prima ispezione dei dati ti accorgi che le osservazioni continue tendono ad essere addensate, mentre gli attributi non continui sono caratterizzati da sparsità.

Decidi di impostare e condurre un'analisi di clustering, come definisci la funzione di similarità e prossimità? Quali sono le accortezze che decidi di mettere in atto? Perché? Puoi intervistare un esperto di dominio, che domande ritieni possano esseri utili a facilitare la tua analisi e a migliorarne i risultati?

Come ritieni si possa procedere per valutare la qualità dei risultati che hai ottenuto, quali misure, strategie e procedure pensi siano adatte a trattare il problema che ti si para dinnanzi?

Illustra lo schema concettuale del workflow di processamento tramite una figura che identifichi le componenti e poi commenta ogni singola componente.

NON E' CONSENTITO FORMULARE DOMANDE AL DOCENTE DURANTE LA PROVA D'ESAME.

RISPONDA ALLA DOMANDA CHIARENDO LE IPOTESI E ASSUNZIONI SULLE QUALI SI BASA LA SUA RISPOSTA.

DOMANDA

Il dataset che devi analizzare consiste in 1,000,000 istanze, gli attributi di input sono 10,000 e tutti continui, mentre l'attributo di output assume tre differenti valori Y={y1, y2, y3} che risultano mutuamente esclusivi. Il dataset contiene 30,000 istanze di classe y1 e 800,000 istanze di classe y3.

Il tuo obiettivo è prevedere il valore della variabile di output Y utilizzando tutti i 10,000 attributi di input.

Gli algoritmi che desideri addestrare sono computazionalmente pesanti, quali strumenti e procedure hai a disposizione per superare questo limite? Quali sono i rischi ai quali ti esponi se applichi procedure di tal tipo?

Una volta ricavata una stima dell'accuratezza del tuo modello devi compararla il livello di accuratezza A raggiunto da un tuo concorrente, purtroppo non disponi di dati aggiuntivi e il tuo competitore non rende disponibile il dataset che ha usato per addestrare il suo modello. Come puoi approcciare questo problema?

Illustra lo schema concettuale del workflow di processamento tramite una figura che ne identifichi le componenti di processamento. Successivaemnte descrivi la funzione di ogni componente di processamento.

DOMANDA

Il dataset che devi analizzare consiste in 10,000 istanze, tutti gli attributi sono continui, tranne l'attributo che devi apprendere che può assumere cinque differenti valori Y={pioggia, vento, neve, sole, nuvole}. Possiamo assumere che pioggia e vento possano accadere contemporaneamente, ma anche singolarmente, allo stesso modo sole e vento possono occorrere contemporaneamente o separatamente, mentre è ragionevole pensare che sia pioggia che neve richiedano la presenza di nuvole, anche se le nuvole possono presentarsi singolarmente. Il tuo compito è prevedere in base ai valori degli attributi di input quali condizioni meteo si avranno domani.

Desideri che il tuo modello sia parsimonioso ed eventualmente interpretabile. Pertanto, decidi di selezionare gli attributi di input piu' rilevanti tramite opportuna procedura, come è organizzata questa procedura? La procedura utilizza tutte le 10,000 istanze? Perché?

Come formuli il problema di apprendimento? Di che problema si tratta?

Illustra lo schema concettuale del workflow di processamento tramite una figura che ne identifichi le componenti di processamento. Successivaemnte descrivi la funzione di ogni componente di processamento.

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari