DOMANDE ICON 3°APPELLO

1°candidato:

Quando si implementa una classificazione tramite apprendimento le classi che dovranno essere predette necessitano un certo grado di indipendenza di fatto le classi devono essere insieme disgiunti. Quando l’indipendenza delle classi non è garantibile dovrebbe potersi garantire il multi label ovvero l’uso di metodi di scikitlearn che permettono di etichettare un esempio di dato come appartenente a più classi.

Quali sono le dimensioni del problema di apprendimento? Le features che permettono la classificazione.

Ingegnerizzazione delle feature come passare da delle features multivalore a feature booleane?(passare da un numero di variabili reali ad un numero di variabili booleane non è un mapping uno ad uno, per cui come si fa ad avere lo stesso quantitativo informativo passando da variabili reali a booleane)

Vuole sapere la complessità della base di conoscenza utilizzata quindi quante regole usate e la loro tipologie ovvero come sono fatte le regole in particolare vuole sapere quanti antecedenti ci sono in ogni regola( le regole sono diverse dai fatti)? C’è bisogno di spiegare la scelta fatta delle regole e della loro complessità.

Le matrici di confusione o le metriche di valutazione del modello di apprendimento necessitano di essere stampati nella documentazione di modo che siano riassuntive di tutti i test fatti.

La cross validation viene fatta per motivi statici del campionamento

DOMANDE TEORIA DOPO AVER VISTO IL PROGETTO:

Apprendimento non supervisionato in generale e differenze rispetto a quello supervisionato? Nell’apprendimento non supervisionato i dati vengono suddivisi in cluster raggruppamenti che vanno a formare dei patern tra i dati da classificare (permettendoci di scoprire nuove info che non abbiamo in partenza). I pattern sono basati sulla similarità degli esempi questa similarità può essere quella del coseno.

Quando facciamo una classificazione son apprendimento non supervisionato stiamo usando un clusterizzatore.

Il k-means visto come risolutore di un problema di ottimizzazione teso a ottimizzare l’iperparametro k ovvero il numero di cluster più adatto per dividere l’intero data set.

2° CANDIDATO NICOLAS PINTO PRESENTAZIONE PROGETTO:

Differenza tra apprendimento probabilistico e ragionamento probabilistico? Quando si parla di ragionamento si ha un modello che non è osservabile per intero e viene usato per fare predizioni.

Sulla classificazione non supervisionata ha usato l’algoritmo di hard clustering.

Ha misurato alla fine della clusterizzazione la valutazione del’algoritmo? Si è stato fatto con varie metriche classiche.

Cosa si intende per inerzia?(riguardo al processo di clusterizzazione chiedere a Nicolas)

DOMANDE TEORIA:

Come si fa ad apprendere un modello probabilistico?[leggermente più complicato di un classificatore naive bayes quali modelli abbiamo?(sono reti TAN(caso particolare di reti bayesiane) in cui una variabile al più è condizionata da un’altra solo variabile)]

Bic = Baysian information criterion,Bac

Il naive bayes ha una topologia ben precisa

Partendo da un dataset che cosa bisogna fare per costruire una rete bayesiana? Sapendo che non sappiamo quali siano le classi che vogliamo ottenere.[oltre alla struttura della rette ci servono le CPT=condition probability table ogni CPT inerente ad una variabile risiede nel nodo della belif network]

La chain rule ci permette di scrivere le probabilità della belif network in forma di produttoria

Owl è un frammento di logica del primo ordine oppure va oltre? è oltre

3°CANDIDATO FEDERICA & NICOLA DOMANDE SUL PROGETTO

Nella documentazione è inutile mettere dettagli su cosa si utilizza a livello di tecnologia ovvero linguaggi o IDE etc…

DOMANDE DI TEORIA

Il KNN si può usare indifferentemente per classificare in modo esclusivo oppure su multi classi, ma è possibile usare il KNN per effettuare regressione?

Differenza tra alberi di decisione usati per fare regressione oppure classificazione? La differenza sta nei nodi foglia

Definire il concetto di conseguenza logica?

Quale è la semantica di false nelle regole false 🡨 a? false è un atomo speciale sempre falso in qualsiasi interpretazione e grazie a tale atomo false

4°2 CANDIDATI SCONOSCIUTI

Mettere la varianza dell’algoritmo su ogni misurazione fatta ove c’è bisogno di specificare il fold della cross validation o quante volte è stato rieseguito l’algoritmo.

Se il data set è troppo piccoli si usa come cross validation la leave-one-out cross validation

Se dovessi avere a che fare con una base semantica come nel web per la parte di dati che tipo di linguaggio avrei bisogni? Lo schema è definito con il framework RDF, attraverso delle triple soggetto, predicato e oggetto.

torta del sematic web??? ricerca

La parte dei dati è divisa da quella delle assersioni

L’over fitting si può riscontrare nella fase di test quado sul training set si ottengono ottimi risultati, mentre sul test set si ottengono cattivi risultati.[ è una cosa da specificare in documentazion]

Un algoritmo che si usa per effettuare il rasoio di Okchamm e abbattere la complessità? C’è bisogno di bassare tali parametri al algoritmo stesso affinché si ottimizzi la complessità dell’albero e questo avviene attraverso l’algoritmo di REGOLARIZZAZIONE il quale impone prima di provare gli alberi + semplici e in seguito quelli più complessi

!!!!Rivedere LA REGOLARIZZAZIONE!!!!! Si penalizzare un albero con un certo peso moltiplicato per il numero dei nodi dell’albero.

Confronto tra random forest e modelli di boosting cosa cambia nel loro apprendimento?

Quali sono + generici gli algoritmi locali o gli algoritmi locali stocastici? Gli stocastici

Nel algoritmo di ricerca locale chiede sia la parte di TRY 1°ciclo che la parte di RANDOM WALK 2°ciclo interno

SECONDO GIORNO 3°APPELLO 1°CANDIDATO

Se si usa una rete Bayesiana bisogna mettere in evidenza la struttura della rete e le tabelle CPT (“quelle dette dal prof”) e poi ci sono le CPD(“altre tabelle usate l’acronimo sta per distribuzione di probabilità condizionata”).

Con il prolog si ragiona con il mondo chiuso e non aperto. Se si dovesse ragionare con mondo aperto dovremmo usare come strumento AI\_LOG che tramite un interprete di prolog aumenta la potenza di ragionamento del prolog per poter ragionare sul mondo aperto.

Nell’apprendimento automatico spesso conviene usare modelli semplici, ci sono modelli semplici che dividono lo spazio in due parte ovvero affrontano la separazione lineare quali sono tali modelli che affrontano il problema della separazione lineare?

SVM appartiene a dei metodi più potenti che riescono ad affrontare il problema della separazione lineare anche quando il problema non è linearmente separabile. [chiede come è definita la funzione del kernel]

Quali sono i modelli che si basano su di un iperpiano? Vuole sentire parlare del percettrone e delle reti neurali

2°CANDIDATO

Nel progetto con le librerie che si scelgono bisogna specificare se si ragiona con mondo chiuso o mondo aperto. La domanda di complessità della KB viene sempre fatta vuole sapere ordine di grandezza degli assiomi e delle regole.

CSP ricerca locale e in particolare ha chiesto gli algoritmi greedy.

Interpretazione di una KB in logica proposizionale?

Che cosa è un problema di regressione?

3°CANDIDATO

K-means di scikitlearn utilizza la misura di INERZIA come bontà del processo di classificazione eseguito. Il K-means l’abbiamo visto come algoritmo di ottimizzazione e in particolare va a minimizzare una determinata misura.

Che differenza c’è tra RDFschema ed OWL? Quale dei 2 linguaggi è più complesso?

Libreria OWLready per trattare l’ontologia della KB. Normalmente sia con RDFschema che con OWL si lavora in mondo aperto

RDF schema è al di sopra di RDF ed OWL è al di sopra di RDF schema.

RDF è un modello di dati basato su triple che descrivono un grafo di conoscenza. Le triple sono costituite da un soggetto che viene identificato come un URI, poi il predicato è dato da una proprietà ovvero da un URI che risulta essere anche essa una risorsa identificabile nel web in modo univoco(“questo serve a definire la proprietà tramite un ontologia del web semantico”), il valore al terzo posto si può esprimere in due modi differenti a seconda della proprietà usata(“tipi di proprietà una relazione che collega individui [object type] oppure quelle che collegano un individuo ad un valore [data type]”)

XML schema fornisce la sintassi per ogni tipologia di dati da utilizzare.

4° CANDIDATO

Durante la presentazione del corso vuole sapere quale soni gli argomenti del corso trattati nel progetto sviluppato.

La struttura della rete bayesiana può essere costruita tramite un modello oppure costruita in modo che sia ottimizzata.

Con quale algoritmo vengono apprese le probabilità della rete Bayesiana?

Libreria bl learn

3° APPELLO

Nel K-means c’è un meccanismo per scegliere il K qual è tale meccanismo che ci permetterà di scegliere il modello con il migliore K? Per uscire dall’algoritmo di partizionamento in classe del K-means devo ottenere il minimo della misura di nostro interesse è l’INERZIA ovvero la distanza tra gli esempi e il centroide di appartenenza [bisogna massimizzare la distanza tra i vari cluster di modo che i centroide siano il più lontano possibili tra di loro].

Il K-means è deterministico o non deterministico? E’ un algoritmo che ha 2 cicli e parte da un assegnazione casuale del K ovvero il numero di centroidi e quindi di classi che vogliamo ottenere(questa è la prima scelta casuale che viene fatta). Successivamente tramite la misura di inerzia si cerca di modificare il K per migliorare la misura di inerzia.

Una misura è la distanza degli esempi rispetto al centroide della classe di appartenenza, questa misura può essere ottimizzata.

Il centro del K-means o del raggio dell’algoritmo può essere visto come il centroide calcolato tra i vettori centroidi di ogni classe oppure può essere calcolato come unico centroide di tutti gli esempi sottoposti a classificazione.

Soft-clustering(ovvero il clustering probabilistico) nell’ apprendimento non supervisionato qual è la misura che viene ad essere ottimizzata è quella della vero-somiglianza. [vuole la formula della verosomiglianza] in particolare si ottimizza la log-likehood.

Quali algoritmi si basano sulla massimizzazione della likehood?

Quali logiche sono quelle che descrivono bene le basi di conoscenza ontologiche? Sono le logiche descrittive

2°CANDIDATO ANDREA

Quando la rete neurale potrebbe non essere adatta come modello? Quanti non ha sufficienti dati poiché tale modello ne richiede una mole notevole.

Cosa è e a cosa serve un meta interprete? Un meta interprete serve per fare una riflessione sul linguaggio stesso e affinché ci sia un interprete serve un meta linguaggio [].

La ricerca locale con algoritmi genetici può essere vista come un processo di ottimizzazione. Raggiunge un punto fisso che ci permette di capire quali individui proseguiranno e tale misura si chiama fitness.

3°CANDIDATO

Il KNN è un algoritmo deterministico.

Quali sono i modelli probabilistici semplici che trattano il concetto di serie e conseguenzialità?Uno è HMM

Modelli Ensemble ?

Differenze tra Random forest e modelli Boosting?

Con una base di conoscenza settata con un mondo chiuso si ragiona di default con prolog o linguaggi simili.

Il passo di inferenza logica a livello tecnico? il passo è una ripetizione è la tecnica di RISOLUZIONE Porta al algoritmo SLD

Un nodo di una belif network è una variabile più la distribuzione di probabilità che la definisce.

Il nodo interno di un albero può essere un singolo test o condizione booleana basata su di una sola variabile oppure su più variabili in caso di alberi decisionali multivariati.