

## Classificazione, Modello di vicinanza, Mappaggio, Instance- Based Classifiers, K-nn, Diagrammi Voronoi, Quad-tree, classificatori bayesiani

Prof. Pierangelo Veltri – 23/10/2023- Autori: Maturo, Accetturo - Revisionatori: Accetturo

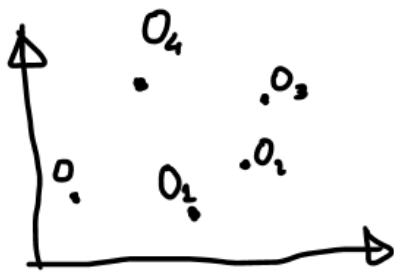
### Classificazione

La classificazione è una tecnica di apprendimento automatico che consiste nell'assegnare automaticamente un'etichetta o una classe a un oggetto o a un insieme di dati in base a un modello costruito da dati di addestramento. È uno dei problemi di apprendimento supervisionato più comuni ed è utilizzato in una vasta gamma di applicazioni come riconoscimento di oggetti, diagnostica medica, ricerca informazioni ecc.

Un altro modello che non è stato ancora discusso è il "modello di vicinanza"

### Modello di Vicinanza

Nel modello di vicinanza, è necessario prima posizionare gli oggetti e successivamente classificarli in base al concetto di prossimità.



#### Come si collocano negli spazi gli oggetti?

Se i vettori, i record o le righe della sua tabella vengono rappresentati in uno spazio a  $n$  dimensioni, per semplicità riportandoli in uno spazio bidimensionale, gli oggetti possono essere considerati come punti o unità, noti anche come punti più comuni, ciascuno caratterizzato da una serie di variabili che li distinguono nello spazio.

Supponendo di avere due dimensioni, se sono  $n$ , si troverebbero in uno spazio in  $R^n$ .

L'idea fondamentale è quella di rappresentarli in uno spazio in cui ogni dimensione corrisponda a un attributo.

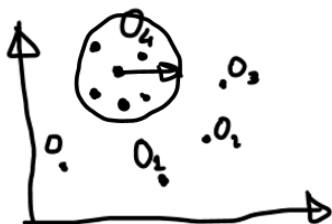
Ad esempio, una dimensione potrebbe rappresentare la familiarità e un'altra potrebbe rappresentare altre informazioni; ognuna di queste diventa una dimensione e gli oggetti vengono collocati nello spazio di conseguenza.

In questo contesto, non è necessario creare un modello basato sui dati per posizionare l'oggetto poiché l'oggetto stesso viene posizionato nello spazio in base alle informazioni relative alle variabili.

Tuttavia, la sfida consiste nel determinare come classificare questo nuovo oggetto.

Per risolverla, la classe di appartenenza viene stabilita basandosi sul concetto di prossimità.

### Concetto di Prossimità in Classifier Nearest Neighbor



Il concetto di prossimità nel Classificatore Nearest Neighbor (k-NN) ruota attorno all'idea di valutare la distanza tra i dati considerati e i punti circostanti. Prima di procedere con questa valutazione, è fondamentale determinare la dimensione dello spazio in cui operare, poiché questa scelta influenzerà notevolmente il numero di punti coinvolti.

Per comprendere come viene decisa la dimensione, è essenziale considerare la rarità del dataset, ovvero quanto sia ampio e variegato. Un dataset esteso e eterogeneo richiede particolare attenzione.

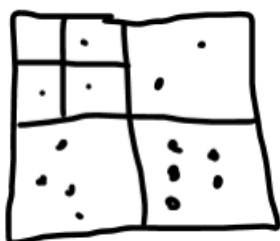
Se imposti la dimensione della sfera di ricerca **troppo grande**, potresti includere **oggetti che fanno parte di classi diverse**. D'altra parte, se la dimensione è **troppo piccola**, l'approccio diventa **sensibile al rumore**.

**Nel k-NN, il vantaggio fondamentale è che non è necessario costruire un modello a priori. Ciò che serve è definire una dimensione che consenta la classificazione degli oggetti basata sulla loro distanza dagli elementi circostanti.**

### Come si decide la dimensione da applicare?

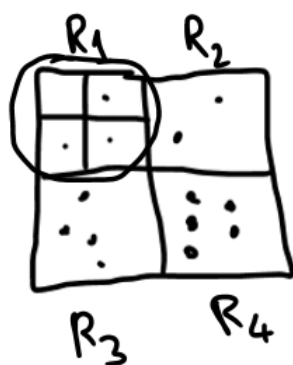
Per selezionare questa dimensione in modo appropriato, bisogna considerare la natura del dataset: quanto è esteso e vario. Inoltre, è importante stabilire attentamente quanti dei "k" oggetti più vicini nel set di confronto devono essere considerati. Quindi, anche se può sembrare che si sia trovato un oggetto alla distanza minima, per garantire l'appartenenza a una classe specifica, è necessario individuare ulteriori oggetti.

### Mappaggio (Metodo di suddivisione in quadranti)



Quando si tratta di mappare dati all'interno di uno spazio, la strategia iniziale è suddividere lo spazio in quadranti. **Questa suddivisione, fondata sui dati disponibili e stabilita in anticipo, è cruciale poiché determina la dimensione e la struttura dello spazio.** L'obiettivo principale è trovare un modo efficiente per calcolare la vicinanza di un oggetto nello spazio.

Se si conosce il quadrante in cui ricade un oggetto, non è necessario calcolare le distanze rispetto a tutti gli altri punti nello spazio. Le distanze vengono misurate rispetto al centro del quadrante, che funge da riferimento per una distanza minima di classe. È possibile anche utilizzare attributi diversi per definire i quadranti. In generale, la rappresentazione del quadrante rappresenta l'intero oggetto.



Questo metodo di suddivisione in quadranti **consente di evitare il confronto diretto di ciascun elemento del dataset**, riducendo così la necessità di copiare l'intero dataset in memoria per confronti dettagliati. Invece di effettuare confronti puntuali, è possibile determinare in quale quadrante si trova l'oggetto iniziale all'interno dello spazio tridimensionale. Questo approccio ha le sue radici nell'ambito dei database ed è stato originariamente sviluppato per le basi di dati spaziali o geografiche.

### Esempio

Quando si effettua una ricerca in un database spaziale per individuare ristoranti in una specifica area o i più vicini, **il software adotta una strategia che evita il confronto diretto con tutti i dati iniziali.** Invece, inizia con una **procedura di filtraggio**.

Immaginiamo di modificare la dimensione della ricerca, passando da un raggio di 10 km a 2 km. In questo caso, si crea una sotto-area all'interno della mappa generale, che può essere concepita come una suddivisione. La suddivisione iniziale si basa su una struttura a quadranti simile a un albero, con quattro quadranti principali denominati R1, R2, R3 ed R4.



Se, ad esempio, si suddivide ulteriormente il quadrante R1, si otterranno R11, R12, R13 e R14. Questo approccio è ricorsivo, poiché si continua a suddividere lo spazio in quattro parti finché non si individua l'elemento desiderato in un singolo quadrante. Si procede con ulteriori filtri, seguendo la struttura dei quadranti. Questa metodologia agevola l'individuazione efficiente degli elementi vicini all'oggetto cercato.

L'idea fondamentale è la possibilità di suddividere lo spazio in più dimensioni, semplificando notevolmente la ricerca all'interno di un database spaziale.

## Instance-Based Classifiers

Gli "Instance-Based Classifiers," noti anche come "Lazy Learners," sono un tipo di algoritmo di apprendimento automatico utilizzati nell'ambito della classificazione e della predizione. A differenza degli "Eager Learners," che costruiscono un modello esplicito durante la fase di addestramento, gli "Instance-Based Classifiers" non generano un modello globale. Invece, memorizzano direttamente gli esempi di addestramento nel dataset.

Questi classificatori prendono decisioni di classificazione basandosi sulla similarità tra un nuovo esempio e gli esempi di addestramento memorizzati. In pratica, quando si riceve un nuovo esempio da classificare, l'algoritmo calcola la sua somiglianza rispetto agli esempi di addestramento precedentemente memorizzati e assegna la stessa classe dell'esempio più simile o di un insieme di esempi più simili.

Due esempi noti di "Instance-Based Classifiers" sono:

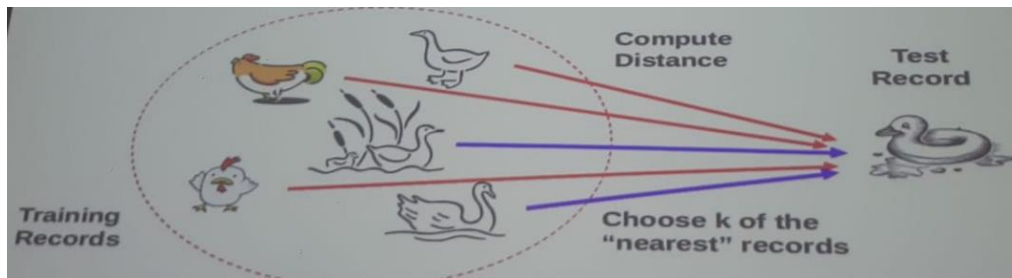
1. Rote-Learner: Questo tipo di classificatore memorizza tutti gli esempi di addestramento. Durante la fase di classificazione, se gli attributi di un nuovo esempio corrispondono esattamente a uno degli esempi di addestramento, viene assegnata la stessa classe di quell'esempio.
2. Nearest Neighbor

## Classificatori Nearest Neighbor

Nel processo di classificazione basata sul k-NN, si utilizzano i k punti più vicini per effettuare la classificazione. La fondamentale idea sottostante è che, se un oggetto presenta somiglianze significative con un gruppo di punti, è probabile che appartenga a quel gruppo.

Per applicare questa metodologia, si parte da un insieme di dati di addestramento, da cui si identificano gli elementi più vicini a un dato punto. Questi elementi più vicini vengono selezionati per comprendere quale categoria rappresentano. Tuttavia, è importante notare che **la classificazione**

**non si basa sulla somiglianza visiva o sull'immagine effettiva.** Invece, il concetto di "distanza" è un punto chiave.



La distanza in questo contesto non è legata all'aspetto fisico dell'immagine, ma piuttosto alle **differenze tra le**

**dimensioni dei dati.**

Per esempio, se un oggetto "starnazza come un'oca" e un altro "cammina come un'oca," la distanza tra le dimensioni di queste due caratteristiche determina se sono simili o diverse. Quindi, la distanza non riguarda l'immagine in sé, ma piuttosto le differenze tra attributi specifici.

Nella pratica, si definiscono le distanze tra i vari attributi. Ad esempio, per stabilire se il suono è "uguale" o se la "camminata" è simile, è necessario avere una definizione chiara della distanza tra queste caratteristiche. Questa è una fase cruciale del processo e implica la definizione di come i dati vengono misurati e comparati per giungere a una classificazione accurata.

In altre parole, si stabiliscono le distanze tra le dimensioni dei dati che sono rilevanti per la classificazione.

Inoltre, va notato che il dataset di addestramento non è suddiviso in parti come training, validation e test set, ma è trattato come un insieme unico. Questo rende il confronto diretto tra il nuovo valore e i dati del dataset possibile senza la necessità di suddivisioni. Se si desidera allenare il modello, è possibile applicare queste suddivisioni, ma non sono obbligatorie per la creazione di un modello k-NN efficace.

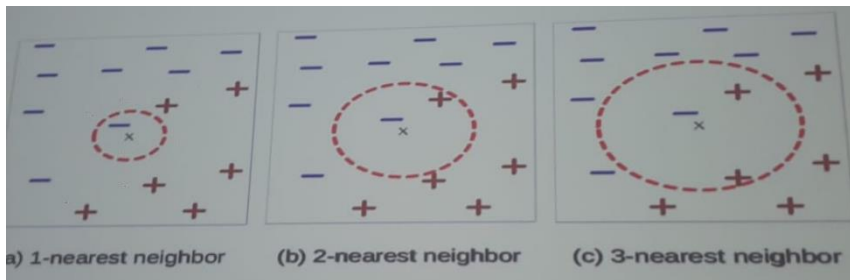
### Come utilizzare i classificatori?

Per utilizzare i classificatori basati su istanze, è fondamentale disporre di **tre elementi chiave**: un training set, una metrica per calcolare la distanza tra i record e un valore noto come "k," che rappresenta il numero di vicini da considerare.

Il processo di classificazione di un nuovo oggetto si articola nei seguenti passaggi:

1. Calcolo della distanza tra il nuovo oggetto e i record presenti nel training set.
2. Identificazione dei k vicini più prossimi al nuovo oggetto.
3. Utilizzo delle etichette di classe di questi vicini più prossimi per determinare a quale classe appartiene il nuovo oggetto. Questo può essere fatto, ad esempio, scegliendo la classe più frequente tra i vicini.

Da notare che **il training set è essenziale per valutare la dimensione ottimale del raggio e il valore di k necessario per ottenere risultati affidabili.** Ad esempio, se l'obiettivo è classificare un punto sconosciuto, si cerca di individuare almeno tre oggetti simili situati all'interno di un cerchio con un raggio specifico. Tuttavia, la scelta del valore "tre" come soglia di vicinanza deve essere testata per garantire l'affidabilità del sistema poiché non esiste una definizione universale per questo valore; dipende dalla situazione specifica.



Se non si riuscisse a trovare abbastanza punti entro il raggio specificato, potrebbe essere necessario estendere la ricerca utilizzando un raggio più ampio fino a individuare un numero sufficiente di vicini. Inoltre, il

training set contribuisce a stabilire il numero minimo di punti con cui un oggetto deve condividere somiglianze affinché il valore di  $k$  sia affidabile. Questo valore, noto come " $k$ ," deve essere significativo, poiché un valore troppo basso potrebbe portare a classificazioni inaffidabili. La scelta del numero di vicini da considerare, o delle "istanze," dipende dalla situazione specifica e dalla casistica.

In conclusione, il training set gioca un ruolo cruciale nella valutazione della dimensione ottimale del raggio e del valore di  $k$ , ma non è essenziale per la costruzione del modello iniziale. La scelta di questi parametri è strettamente legata all'applicazione specifica e all'analisi dei dati. L'autore sembra delineare il concetto di validazione, il quale richiede che il numero di casi utilizzati in una finestra temporale di un anno debba essere almeno  $X$  volte superiore al numero di casi di interesse per una specifica patologia.

### In campo Medico

È chiaro che, se si analizzano i punti nello spazio, questi possono rappresentare due classi, positive e negative, costituite da diverse variabili. Ad esempio, nel contesto di analisi medica, si tratta delle descrizioni delle patologie e delle prestazioni diagnostiche per immagini. Un elemento fondamentale in questa area è la ripetizione delle definizioni. Se si considera un ampio campione di reperti di diagnostica per immagini, come risonanze magnetiche o TAC, le definizioni si ripetono spesso. Anche se le descrizioni possono variare tra lingue, come italiano o inglese, la struttura rimane sostanzialmente la stessa. È simile a quando una persona da diversi luoghi del mondo, come Cina, Sudafrica o Nord Europa, arriva a una conclusione comune: siamo tutti esseri umani.

Pertanto, l'obiettivo è sviluppare sistemi di classificazione automatici che, partendo dalle immagini, propongano diagnosi basate su queste definizioni. Tuttavia, è importante notare che queste proposte di diagnosi richiedono sempre la convalida da parte di un medico. In questa fase, l'attenzione si concentra sulla composizione digitale delle terminologie e sulle distanze tra i termini. Ciò significa definire come una determinata immagine possa avere caratteristiche più o meno marcate, come la presenza di contorni definiti o meno. Questo è un processo che richiede l'utilizzo di tecniche semantiche e l'analisi delle descrizioni.

In sintesi, la fase iniziale di questo processo coinvolge l'analisi delle immagini e l'estrazione delle loro caratteristiche. Queste caratteristiche vengono convertite in un vettore e utilizzate per calcolare le distanze rispetto ai termini presenti in un dataset o base di conoscenza. È un processo iniziale, ma non è l'intero processo, poiché successivamente c'è la validazione da parte del medico.

### DEFINIZIONE DI DISTANZA

La definizione della distanza è una parte essenziale del processo di classificazione. Per comprenderla meglio, prendiamo ad esempio strutture proteiche. La definizione di distanza in questo contesto è simile a valutare le differenze tra individui provenienti da diverse parti del mondo. Sebbene le caratteristiche possano variare, l'essenza rimane la stessa, ossia l'appartenenza alla razza umana.

Tuttavia, definire una distanza tra attributi all'interno di un dominio costituisce una sfida complessa. Questa definizione deve essere integrata con altre informazioni presenti nei record, che possono comprendere dati testuali o altre variabili influenti sulla distanza. Modificare le definizioni può complicare il problema, aumentando le dimensioni dello spazio degli attributi. Ad esempio, è possibile aggiungere attributi per migliorare la definizione della distanza.

Iniziamo spesso con l'automazione e il calcolo automatico delle distanze. Tuttavia, quando i risultati non soddisfano, è possibile ridurre il rumore eliminando attributi che si ritengono inutili o confusionari.

In conclusione, la definizione della distanza è un aspetto cruciale del processo di classificazione. La scelta della definizione specifica dipende dall'applicazione e dalla situazione, e potremmo dover sperimentare diverse definizioni per ottenere risultati migliori. Questo spiega perché continuiamo a lavorare costantemente sulle definizioni per migliorare il processo.

## PRO E CONTRO DI K-NEAREST NEIGHBOR

Vantaggi dei KNN:

- ✓ **Non richiedono la costruzione di un modello**
- ✓ Rispetto ai sistemi basati su regole o decision tree permettono di costruire “contorni” delle classi non lineari e sono quindi **più flessibili**

Svantaggi dei KNN:

- ✓ **Richiedono una misura di distanza per valutare la vicinanza**
- ✓ Richiedono una fase di pre-processing per normalizzare il range di variazione degli attributi
- ✓ La classe è determinata localmente e quindi è **suscettibile al rumore** dei dati
- ✓ Sono molto sensibili alla presenza di attributi irrilevanti o correlati che falseranno le distanze tra gli oggetti
- ✓ Il **costo** di classificazione può essere **elevato** e dipende linearmente dalla dimensione del training set in mancanza di opportune strutture ad indice

La definizione della *bontà della misura* è strettamente legata alle preferenze dell'utilizzatore del classificatore. Il costo associato alla classificazione può aumentare notevolmente quando si rende necessario confrontare attributi provenienti da domini completamente diversi. Il concetto di distanza varia notevolmente in base agli attributi che costituiscono il record, il che implica che in certi casi potrebbe non esservi una definizione chiara della distanza.

Per illustrare, quando ci si occupa della definizione delle strutture proteiche, è fondamentale determinare quanto una previsione sia simile o diversa da un'altra, valutando i cambiamenti nei composti chimici o nelle distanze tra le parti esterne e interne delle strutture. Questa complessità è dovuta al fatto che le metriche di definizione della distanza devono essere integrate con altre informazioni. I record comprendono non solo le caratteristiche delle strutture proteiche, ma anche informazioni sulla reattività dei farmaci, sul collegamento delle molecole e dati di natura testuale. Di conseguenza, la definizione della distanza diventa una combinazione di questi fattori.

Quando il KNN non riesce a fornire risultati soddisfacenti con un dataset esistente, è possibile apportare modifiche al dataset stesso e cercare informazioni aggiuntive. Per esempio, in un dataset di pazienti, potrebbe essere utile esaminare le cartelle cliniche a disposizione per individuare attributi che indichino se un paziente ha subito un infarto. Questo nuovo attributo può essere considerato nella definizione della dimensione della distanza.

In situazioni in cui il KNN è utilizzato ma i risultati non sono soddisfacenti, si può prendere in considerazione la rimozione di attributi dalla tabella dei dati. Per esempio, se una tabella ha dieci dimensioni e una di esse rappresenta la classe, è possibile ridurre il numero di dimensioni da nove a otto eliminando alcuni attributi che potrebbero causare rumore. Questi attributi non verranno considerati nella definizione della distanza, evitando così di influenzare negativamente le decisioni del classificatore. Per esempio, si possono calcolare le distanze tra due strutture proteiche senza considerare la struttura secondaria, che fornisce informazioni sull'alfa elica e il beta foglietto. In alternativa, si può includere la condivisione di alfa eliche e beta foglietti nel calcolo della distanza in sostituzione della distanza tra gli amminoacidi.

## MOTIVAZIONE

Nella classificazione dei termini, alcuni di essi sono considerati descrittivi, mentre altri sono caratterizzanti. Ad esempio, il termine "nome" ha un peso relativamente basso in una metrica, poiché è considerato un termine caratterizzante. Il suo significato semantico varia a seconda del contesto, e quindi la sua distanza rispetto ad altri termini può essere misurata su una scala da 0 a 1.

Nella classificazione, si può anche affinare la distanza tra termini. Ad esempio, se si combina il termine "nome" con un altro termine come "quadro" o "Van Gogh," la distanza tra di essi può essere ridotta. Questo affinamento può aiutare a catturare meglio il significato dei termini quando vengono utilizzati insieme.

Inoltre, la misura della distanza è utilizzata nelle ricerche di sotto-sequenze all'interno di dataset. Queste sotto-sequenze possono essere numeriche e forniscono informazioni dettagliate su una parte dell'immagine o dei dati. Ad esempio, cercare sotto-sequenze in un insieme di dati di sequenziamento del DNA può rivelare specifiche regioni di interesse.

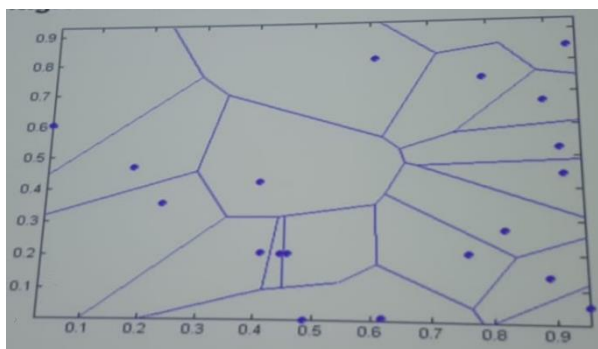
Infine, la misura della distanza è importante anche nelle verifiche di originalità, come quando si scrivono articoli scientifici. Prima di sottoporre un articolo, spesso si esegue una scansione antiplagio per rilevare eventuali similitudini non autorizzate con altre fonti.

## ANTI-PLAGIO

Questo strumento viene utilizzato per verificare la similarità tra testi, anche se non c'è un testo di riferimento diretto. La sfida consiste nel definire quanto due testi siano simili, considerando anche le possibili modifiche o spostamenti di parti di testo. Per farlo, si utilizza una metrica che valuta la similitudine tra i testi, evidenziando in giallo le parti copiate.

L'anti plagio sfrutta parole chiave specifiche per effettuare queste verifiche, utilizzando anche tecniche di selezione e confronto per individuare similitudini tra i testi.

## DIAGRAMMI DI VORONOI E TRIANGOLAZIONE DI DELAUNAY

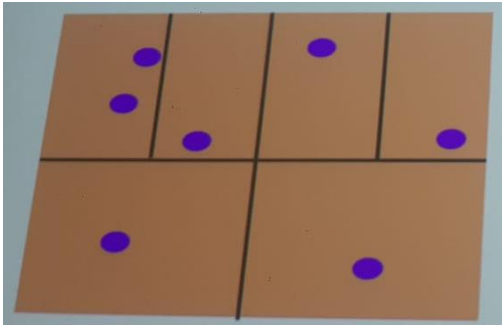


Nei diagrammi di Voronoi, lo spazio viene suddiviso in regioni geometriche tali che i punti all'interno di ciascuno spazio di riferimento siano equidistanti dai lati del poligono che lo circonda. Questa tecnica trova applicazioni nella definizione delle aree di copertura delle antenne telefoniche, in particolare per le reti cellulari.

Le antenne cellulari sono collocate su torri o edifici e vengono divise in regioni, note come celle, per coprire il territorio. Tuttavia, poiché il territorio non è pianeggiante ma presenta variazioni di

altitudine, è necessario utilizzare una rete di triangolazione irregolare (Triangular Irregular Network) per collegare le antenne tra loro. Questo permette di mappare la morfologia del territorio in modo da calcolare non solo le distanze, ma anche la visibilità tra le antenne. In una città, questo assicura che le antenne coprano efficacemente il territorio e si possano vedere reciprocamente.

## QUAD-TREE



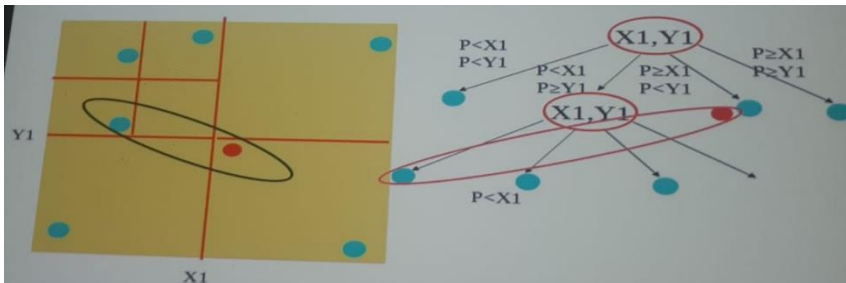
Nella suddivisione degli indici spaziali, esiste la possibilità di dividere lo spazio in quadranti in modo che ciascun quadrante contenga un valore specifico. Questi quadranti possono quindi essere rappresentati in una struttura ad albero. In questa struttura, la ricerca dei vicini diventa una questione di recuperare l'elemento dalla foglia dell'albero.

La struttura ad albero ci aiuta a trovare rapidamente il quadrante in cui si trova il nostro punto di interesse, poiché il test iniziale effettuato sulla radice dell'albero ci consente di navigare verso il quadrante corretto. Questo processo richiede solo un numero di passaggi logaritmico.

Tuttavia, è importante notare che la creazione dell'albero e la suddivisione in quadranti sono operazioni offline, cioè vengono eseguite inizialmente quando si ricevono i dati. Se il dataset cambia o si aggiunge nuova informazione, è necessario ricalcolare la struttura dell'albero e la suddivisione dei quadranti, il che può diventare oneroso.

## Problema

Il rovescio della medaglia dei quad tree include due aspetti principali. Uno di essi riguarda il calcolo dello spazio, mentre l'altro è associato al fatto che l'appartenenza di un oggetto a un quadrante non



implica necessariamente che questi oggetti siano vicini.

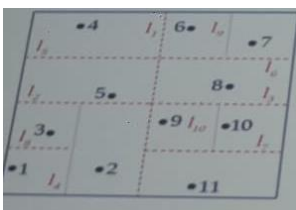
Ad esempio, consideriamo un caso in cui cerchiamo un oggetto rosso in uno spazio suddiviso in quadranti.

Potremmo trovarlo in un quadrante specifico, ma in realtà

potrebbe essere molto più vicino a un altro quadrante. Questo scenario solleva la questione di come suddividere un insieme di punti in quadranti e come costruire un albero bidimensionale per rappresentarli.

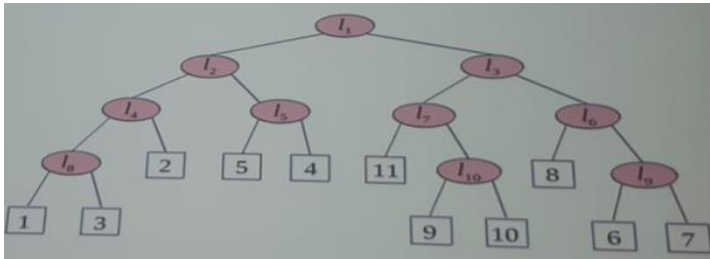
Oltre a ciò, possono sorgere domande riguardo alle regole date per la suddivisione e se queste regole siano esaustive o meno. In caso contrario, potrebbe essere necessario sviluppare ulteriori regole o metodi per la creazione del modello. POSSIBILE DOMANDA D'ESAME

## Kd-tree query



La divisione delle strutture inquadrabili presenta diverse conformazioni, e una possibile illustrazione è la seguente. Inizialmente, si parte con una struttura simile a una rete di nodi interconnessi. Ogni nodo, rappresentato come un punto celeste, è un nodo foglia, contenente già un valore di riferimento.





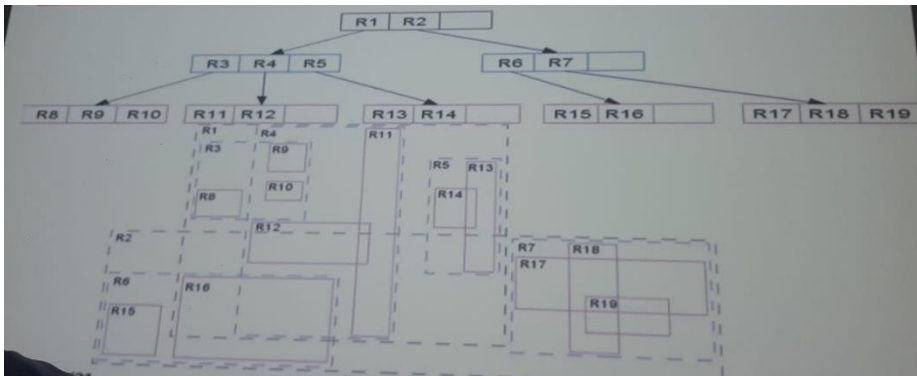
Quando si giunge a un nuovo nodo foglia, si ha accesso al valore di riferimento associato a quel punto. Tuttavia, è importante notare che i nodi intermedi dell'albero non contengono dati reali, ma piuttosto informazioni che guidano la navigazione nel sistema. Questi nodi intermedi servono a stabilire la struttura

dei quadranti senza contenere dati effettivi.

È essenziale mantenere chiara questa distinzione tra nodi foglia (contenenti dati di riferimento) e nodi intermedi (utilizzati solo per la navigazione). La decisione di includere informazioni relative ai dati nei nodi intermedi modificherebbe fundamentalmente la struttura dell'intero albero.

### R-TREE

Nel processo di creazione di una rappresentazione, ad esempio, utilizzo una griglia rettangolare suddivisa in zone. Queste suddivisioni rettangolari possono variare in dimensioni, il che significa che alcune zone potrebbero essere più piccole di altre, a seconda del contesto e dei dati considerati.



Il mio obiettivo principale è comprendere come muoversi nello spazio multidimensionale. Avendo a disposizione questa griglia, posso determinare se un oggetto appartiene a una determinata zona o regione della

rappresentazione.

### CLASSIFICATORI BAYESIANI

In medicina, è noto che la meningite causa il torcicollo nel 50% dei casi. La probabilità di un paziente di avere la meningite è 1 su 50.000, mentre la probabilità di avere il torcicollo è 1 su 20. Il teorema di Bayes consente di calcolare la probabilità inversa, ovvero la probabilità di avere la meningite dato il sintomo del torcicollo. Per effettuare questo calcolo, è necessario combinare le informazioni sulla probabilità condizionale con le probabilità dei casi singoli, seguendo una procedura convenzionale.

Nel calcolo delle probabilità, si possono utilizzare i valori di probabilità noti per effettuare previsioni o classificazioni. Ad esempio, se si conosce la probabilità che una squadra di casa vinca e la probabilità che una squadra ospite vinca, insieme alla probabilità di una squadra che ha vinto fuori casa rispetto a quella che ha vinto in casa, è possibile calcolare la probabilità che una squadra X vinca la prossima partita fuori casa. Questo calcolo si basa sull'applicazione delle probabilità condizionate e della combinazione delle probabilità.

**Questa stessa logica può essere utilizzata per scopi di classificazione.** Nel calcolo delle classi di appartenenza, si considerano valori numerici e si cerca di determinare a quale classe un nuovo oggetto appartiene. Questo processo si basa sulla valutazione dei pesi o delle correlazioni che indicano in quale direzione un discriminante spinge l'oggetto, aiutando così a effettuare la classificazione.