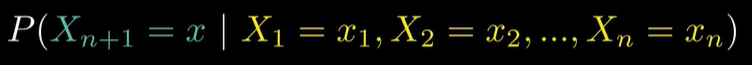
**MARKOV CHAIN**

**CATENE DI MARKOV**

Possiamo vedere la catena di markov come un grafo di probabilità, ad esempio supponiamo di avere a che fare con una mensa che in un dato giorno abbiamo servito un pasto a scelta tra hamburger, pizza o hotdog le probabilità di vedere servito uno tra questi pasti il giorno seguente è dato dal seguente grafo:



una proprietà importante per le catene di Markov è quella che lo stato futuro dipenderà solo e soltanto dal corrente stato nel quale ci troviamo, in termini matematici usando le probabilità possiamo dire la probabilità di un evento che dovrebbe essere legato agli avvenimenti precedenti come segue:



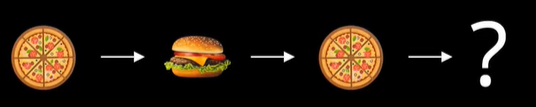
Dipende solo dall’attuale stato ovvero l’ultimo nello storico e precedente a quello futuro ovvero:

Text

Description automatically generated

Questa proprietà rende più semplice riuscire a ragionare e risolvere determinati problemi della realtà.

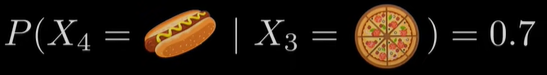
Quindi supponiamo che il ristorante serva la seguente sequenza di pasti tra i 3 a disposizione:



E vogliamo sapere quale che sia la probabilità di ricevere un hotdog al 4° giorno, per la teoria classica delle probabilità per condizioni date dal teorema di Bayes avremo che tale probabilità sia data come la condizionata ai giorni precedenti come segue:



Ma sotto le ipotesi di markov abbiamo che essa sarà esclusivamente condizionata dal precedente stato ovvero quello che è stato servito nel 3° giorno:

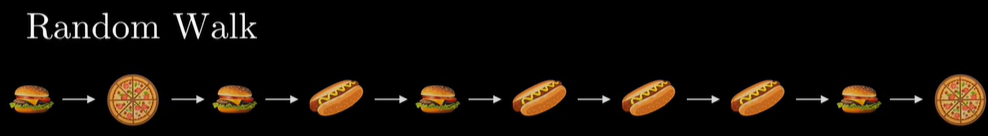


La seconda proprietà importante delle catene di Markov è che la somma dei pesi associata alle frecce in uscita da uno stato (nodo del grafo precedente visto) deve essere pari ad 1, poiché tali pesi rappresentano probabilità :

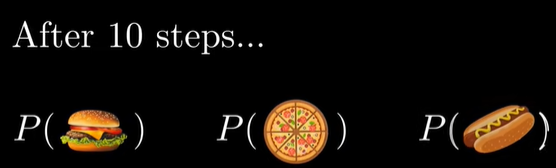
**EX: se lo stato è quello dell’hamburger allora la somma dei suoi archi uscenti è**

**0.2 + 0.6 + 0.2 = 1**

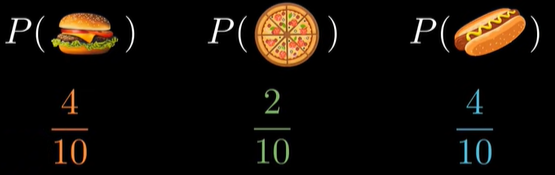
Ora supponiamo di effettuare un random walk nel grafo e compiamo dieci step causali ottenendo la seguente sequenza di pasti fornita per 10 giorni:

****

Ora invece ci chiediamo quali siano le probabilità dei 3 pasti nella sequenza dei 10 pasti presentati:

****

La probabilità dei pasti nella sequenza è data dall’approccio frequentista della probabilità quindi avremo:

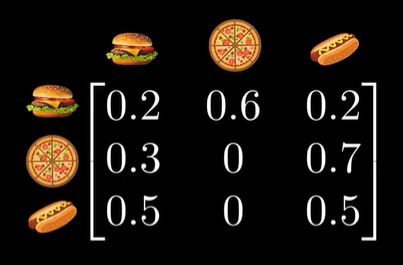
****

si può facilmente capire che a seconda della sequenza prese in esame le frequenze cambieranno dando vita a probabilità diverse, ma cosa succede se il processo dovesse tendere ad infinito ovvero se applichiamo il concetto di limite a tali sequenze? Avremo che le probabilità dei 3 pasti convergeranno verso i seguenti valori:

Graphical user interface, application

Description automatically generated

Questa distribuzione che assumono prende il nome di distribuzione STAZIONARIA chiamata anche stato di equilibrio. Nonostante abbiamo compreso la convergenza a determinati valori risulta complicati calcolarli in questa maniera per cui pensiamo ad una rappresentazione del grafo tramite l’algebra lineare e in particolare tramite una MATRICE DI ADIACENZA:

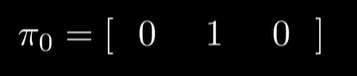


Questa MATRICE DI ADIACENZA è perfettamente equivalente alla rappresentazione a grafo vista precedentemente.

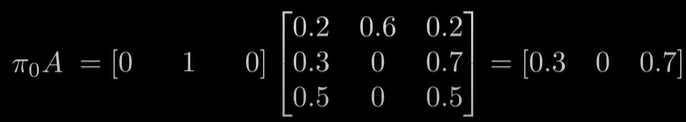
Ora chiamiamo la matrice di adiacenza A:



E consideriamo un vettore iniziale le cui entrate stabiliscano il punto di partenza nella camminata casuale per il grafo ovvero se l’entrata è posta ad 1 ci troveremo nello stato corrispondente a 1 se un entrata è a 0 allora vorrà dire che non siamo in tale stato e supponendo che l’ordine delle entrate sia quello delle colone della matrice ovvero [Hamburger, pizza, hotdog] allora il seguente vettore indicherà che ci troviamo nello stato pizza:



Vettore andiamo adesso a moltiplicare il vettore con la matrice di adiacenza ottenendo un nuovo vettore come segue:



E continuiamo a iterare lo stesso processo sostituendo al vettore iniziale il successivo vettore che troveremo ad ogni moltiplicazione (vettore x matrice):

A picture containing application

Description automatically generated

Text

Description automatically generated with medium confidence

Ora se la distribuzione è una distribuzione stazionaria allora ci sarà un particolare vettore il quale una volta ottenuto e moltiplicato con la matrice di adiacenza di darà sempre lo stesso vettore e questo vettore sarà nel nostro caso proprio il vettore delle convergenze visto in precedenza.

A picture containing text

Description automatically generated

La precedente equazione di algebra lineare ci porta a pensare agli auto-vettori e autovalori

A picture containing text, gauge

Description automatically generated

Dove lambda = 1, questo significa che pi è un auto-vettore e se così è esso deve rispettare una condizione che hanno tutti gli auto-vettori ovvero che la somma delle entrate sia pari ad 1:

Text

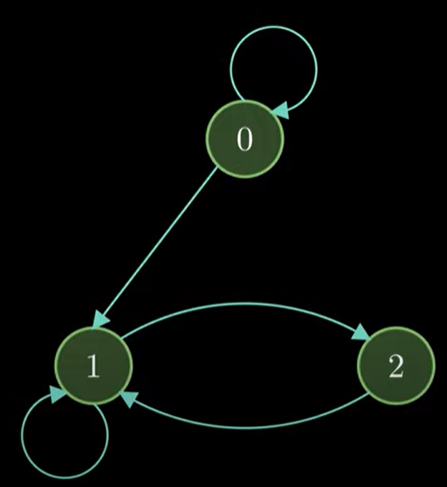
Description automatically generated with medium confidence

E come è possibile constatare il precedente vettore dato dalle convergenze su una sequenza infinita porta a darci proprio l’auto-vettore cercato:

Text

Description automatically generated

In alcuni grafici come nel seguente la probabilità tramite un random walk partendo dallo stato zero di rivisitarlo una volta che lo abbiamo lasciato è pari a zero ovvero non accadrà mai più



La probabilità pari a zero di rivisitare lo stato zero fa si che esso sia un particolare stato detto TRANSIENT STATE:

Text

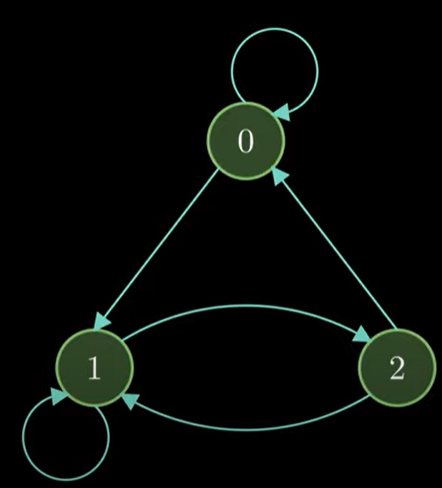
Description automatically generated

Mentre per gli altri 2 stati 1 e 2 la probabilità di rivisitarli in futuro partendo da essi dopo averle lasciati è una probabilità pari ad 1 ovvero è una certezza che prima o poi rivisiterò questi stati nuovamente e per questo vengono detti STATI RICORRENTI:

Text

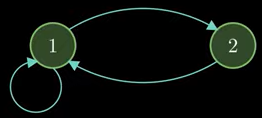
Description automatically generated with medium confidence

Nei casi come questo in cui nel grafo che illustra la catena di markov abbiamo un stato che una volta lasciato non è più raggiungibile dagli altri stati diciamo che la catena di markov è RIDUCIBILE, una catena di markov di questo tipo può essere trasformata in una IRRIDUCIBILE ovvero una in cui da ogni stato è possibile prima o poi arrivare negli altri, per fare questa trasformazione occorrerà aggiungere l’arco mancante che trasformerà il grafo nel seguente:



Il nome riducibile associato alla catena è dato poiché è possibile dividere la catena attuale in sotto catene più piccole le quali risulteranno irriducibili

Icon

Description automatically generated with low confidence 

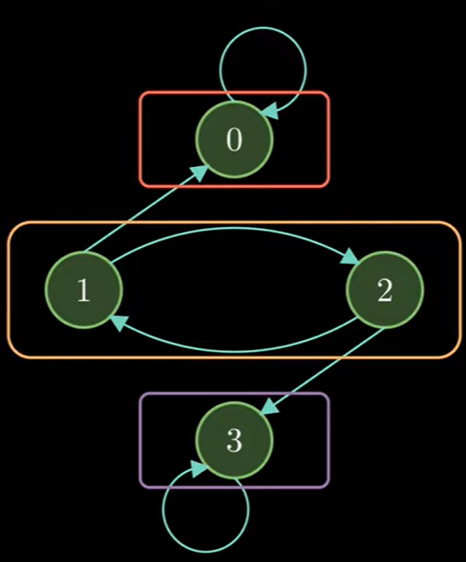
Ora affrontiamo una catena di Markov con più stati e cerchiamo di capire quali tra questi comunicano e quali no, la seguente catena di Markov è nota come the gambler’s ruin:

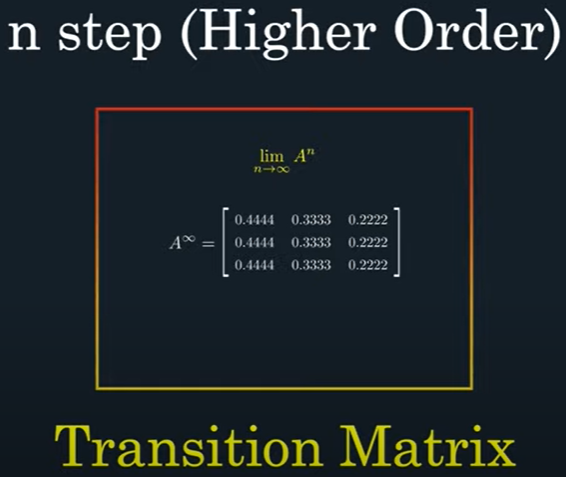
Shape

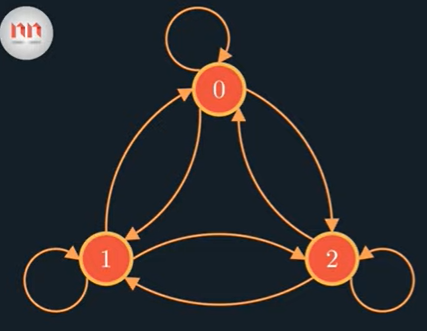
Description automatically generated

In questa catena gli stati 0 e 3 risultano isolati e non comunicano con gli altri, mentre gli stati 1 e 2 comunicano tra di loro in modo infinito, mentre nel momento in cui comunicano rispettivamente 1 con 0 oppure 2 con 3 la comunicazione si interrompe e rimane bloccata in uno di questi stati non collegati.

Possiamo dividere gli stati di questa catena di Markov in classi come segue la classe 0, la classe 3 e la classe data da (1 e 2) come in figura:







Dato il seguente grafo rappresentante una catena di Markov vediamo la matrice di adiacenza associata A:

Text, application, chat or text message

Description automatically generated

Il nostro obiettivo sarà quello di riuscire a trovare una risposta per calcolare la probabilità di raggiungere lo stato j dallo stato i dopo esattamente un numero n di step. Definiremo tale probabilità

A close-up of a logo

Description automatically generated with low confidence

Ad esempio la probabilità di raggiungere lo stato 2 dallo stato 0 dopo esattamente 1 passo è la seguente:

Logo

Description automatically generated

Questa probabilità sarà uguale a: A picture containing text, gauge, device, meter

Description automatically generated

Per cui potremo dire che per ogni probabilità che vogliamo calcolare che preveda il raggiungimento da uno stato ad un altro in esattamente un passo sarà uguale a:

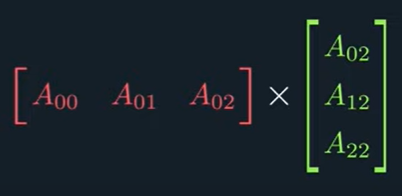


Se invece consideriamo un n + grande ad esempio n = 2 e considerando sempre come stato di partenza 0 e come stato di arrivo 2 abbiamo che tale probabilità è data dai seguenti cammini che si traducono per le proprietà della catena di Markov in prodotti tra le probabilità date dagli archi:

Text

Description automatically generated

Siamo interessati a trovare i prodotti che diano le probabilità sfruttando l’algebra lineare, infatti possiamo pensare ad essi come prodotto scalare tra 2 vettori ove questi vettori risultano essere righe e colonne della matrice di adicenza:

 Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

Come visto dai 2 esempi sembra esserci un pattern che ci permette di sfruttare la moltiplicazione tra matrici dell’algebra lineare per ottenere le probabilità alle quali siamo interessati tali probabilità saranno date dall’elevamento a potenza della matrice di adiacenza:

Graphical user interface, text, application

Description automatically generated

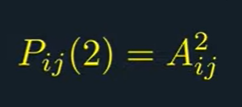
Possiamo notare molto bene da questo esempio come l’elevamento al quadrato della matrice di adiacenza ci fornisca tutte le probabilità inerenti a trovare la probabilità di partire da uno stato j e arrivare in uno i in esattamente 2 steps

A screenshot of a computer

Description automatically generated with low confidence

Graphical user interface, text, application

Description automatically generated



Possiamo generalizzare il processo per n passi esatti ricorrendo a un elevamento a potenza n-esima della matrice di adiacenza per trovare le probabilità di nostro interesse come segue:



Il teorema di CHAPMAN-KOLMOGOROV ci permette di dire che la probabilità di raggiungere lo stato j partendo dallo stato i in n passi è data dalla sommatoria delle probabilità che ci permettono di raggiungere partendo da i uno stato di mezzo chiamato stato di sosta k in esattamente r passi con r < n per la probabilità di raggiungere lo stato j partendo dallo stato sosta in esattamente n – r passi:

Text

Description automatically generated

A seguire troviamo una breve dimostrazione del teorema di Chapman-Kolmogorov come segue:



Ora vediamo cosa succede affrontando il limite dell’elevamento a potenza della matrice di adiacenza



Graphical user interface, text, application

Description automatically generated

La matrice di adiacenza elevata a potenza infinita arriva a convergere a tre vettori identici che sono il vettore stazionario (ovvero l’autovettore)



Ovvero questa probabilità è la stessa per un j fissato, ovvero potremmo partire da qualunque stato i e raggiungere un j-esimo stato in un numero infinito di passi che conduce ad avere una stessa probabilità



La matrice di adiacenza elevata ad infinito converge solo nel caso in cui la matrice (o la catena di markov) rispetti 3 proprietà:

1. Irriducibile
2. Aperiodica
3. X

**HIDDEN MARKOV MODEL**

Per cercare di capire come applicare un modello di hidden markov chain cerchiamo di rappresentare un possibile problema di incertezza che andremo a gestire con le catene di markov e il teorema di Bayes.

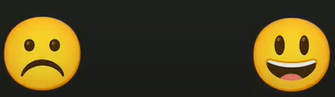
Per l’esempio supponiamo che un nostro amico Benny sia in una città lontana dalla nostra ove il meteo ha solo le seguenti tre condizioni meteorologiche:



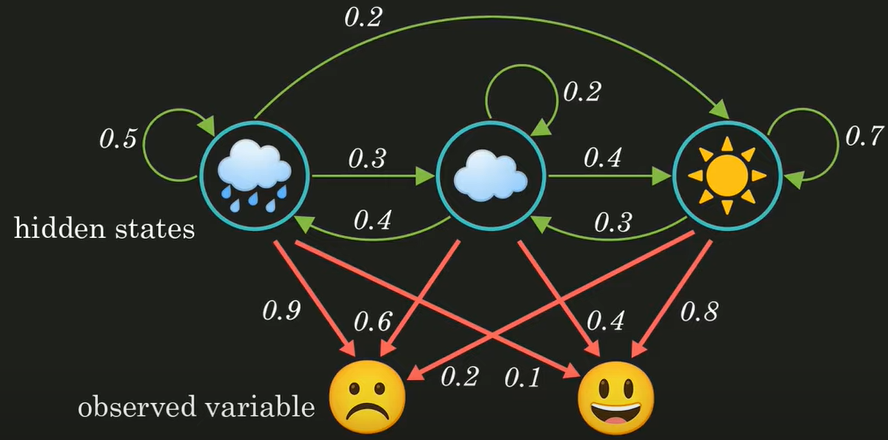
Modelliamo questi 3 stati meteorologici secondo una distribuzione e andiamo a disegnarne il grafo di transizione che ne rappresenterà la catena di markov con la quale andremo a lavorare:



Ora assumiamo che il nostro amico Benny possa assumere solamente 2 tra i seguenti stati d’animo triste o allegro:



E che questi due stati dipendano dal tempo che c’è nella sua città per cui il tempo andrà ad influenzare l’umore di Benny:



Le dipendenza dell’umore di Benny sono espresse in probabilità dagli archi rossi. Rispetto a tale problema essendo noi non residenti nella città di Benny non conosciamo il meteo del giorno, ma videochiamando Benny possiamo osservare il suo stato d’animo e sapere se è triste o allegro. Di fatto Questo sistema dato da i 3 stati a noi nascosti più i 2 osservabili del nostro amico prende il nome di modello nascosto di markov e ci servirà per predire degli avvenimenti inerenti il tempo nella città di Benny.

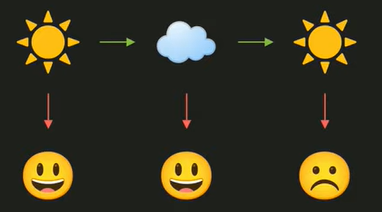
In breve, possiamo dire che un HMM ovvero un Hidden Markov Model è dato come:



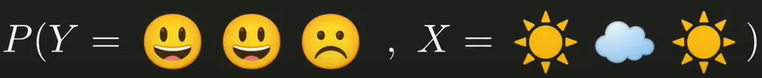
Ora usiamo le matrici e l’algebra lineare per rappresentare questo modello e le deduzione che ne possiamo fare:



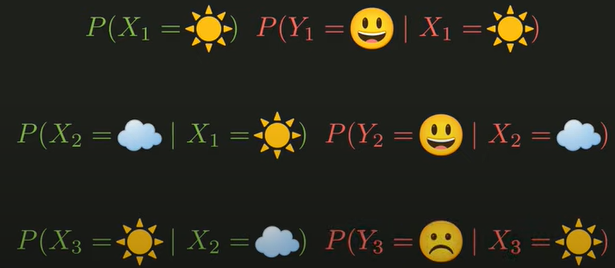
Ora prendiamo una sequenza di giorni con in considerazione con i rispettivi stati di umore di Benny:



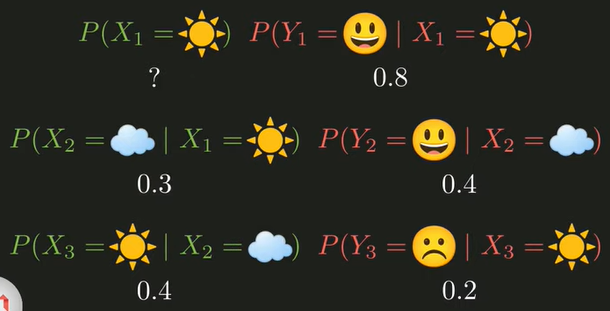
E ora vogliamo sapere la probabilità che tale evento sia accaduta ovvero che si sia verificata questa sequenza in termini di probabilità vogliamo calcolare:



Ora utilizzando le proprietà delle catene di Markov possiamo pensare a tale probabilità come una combinazione di 6 termini:

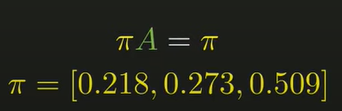


Una volta definito il calcolo non ci resta che calcolare la formula usando le probabilità che prenderemo direttamente dalle 2 precedenti matrice ovvero quella verde che rappresenta la matrice di adiacenza per la catena di markov nascosta e quella rossa che rappresenta gli stati di Benny al variare del tempo (la 1°riga si riferisce alla pioggia, la 2° al nuvoloso e la 3° al sole):



Resta scoperta la prima probabilità ovvero che il primo giorno della sequenza ci sia il sole, per calcolarla utilizziamo proprio le proprietà della catena di Markov e andiamo a calcolare il vettore stazionario della distribuzione stazionaria data dalla catena di Markov di modo che avremo il primo termine cercato ovvero la probabilità a priori che sia un giorno di sole. Per effettuare il calcolo usiamo o il metodo iterato delle potenze, oppure moltiplichiamo ripetutamente la matrice di adiacenza con se stessa.

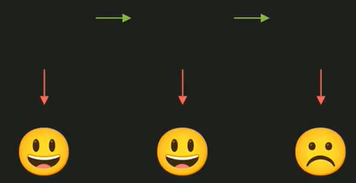
Otterremo quindi il nostro autovettore:



E quindi potremo usare la priorità a priori del sole ovvero P(sole) = 0.509

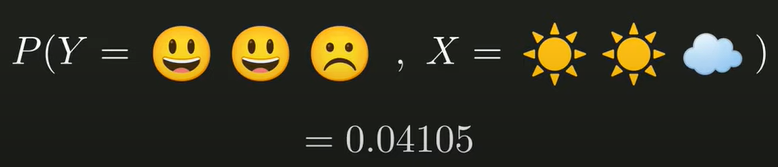
Ora poniamoci un altro interrogativo per sfruttare a pieno il potere del HMM, e quindi quale è la più probabile sequenza metereologica che potremmo osservare data la precedente sequenza di umore di Benny?

Quindi abbiamo gli umori in sequenza vogliamo conoscere la sequenza più probabile di tempo:



Questa risulta essere la sequenza che massimizza la probabilità di vedere la sequenza di umore precedentemente osservata:



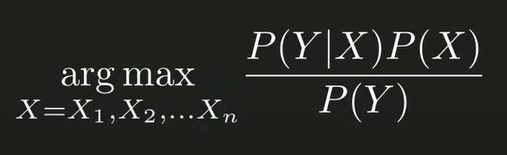


Ora cerchiamo di capire l’algebra lineare utilizzata assieme alle proprietà di Markov e il teorema di Bayes i quali strumenti rendono incredibilmente utile e potente il modello delle catene nascoste di markov e ci permettono di effettuare delle previsioni come quelle descritte:

partiamo dal formulare la nostra precedente domanda avendo fissato che gli stati d’animo sono rappresentati come Y e che il tempo come X, allora in forma di probabilità quelle che precedentemente abbiamo fatto è trovare la probabilità massima che dato Y avvenga X:



Ora esprimendo i questo modo la probabilità sembra non esserci soluzione al problema ed è qui che interviene il teorema di Bayes che ci permette di riscrivere la precedente formula come:

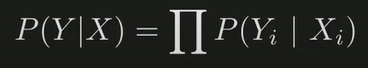


Poiché vogliamo massimizzare tale probabilità e per semplicità di calcoli possiamo eliminare il denominatore P(Y) e proseguire con la restante parte.

E prendiamo in considerazione la prima parte della formula P(Y|X) sulla quale possiamo dire che tenendo in conto dell’assunzione fatta che uno stato d’animo dipende esclusivamente dal meteo del giorno in cui Benny ha quell’umore, possiamo utilizzare l’assunzione di indipendenza degli eventi e riscrivere tale probabilità come segue:



E quindi portando sotto produttoria otteniamo:

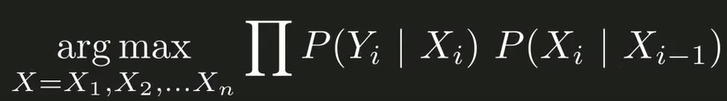


Una volta effettuato questo passaggio utilizziamo la proprietà della catena di Markov che ci dice che lo stato futuro dipende solo dallo stato attuale e non dagli altri precedenti all’attuale e otteniamo che Xi dipende solo da Xi – 1 e quindi la probabilità di P(X) nella formula del teorema di Bayes può essere calcolata tenendo conto di tale proprietà della catena di Markov come:



Ove utilizzeremo ancora una volta il vettore stazionario per calcolare la probabilità di x0 .

Ora mettendo assieme le espressione trovate otteniamo che la probabilità che vogliamo massimizzare è data da:



**HIDDEN MARKOV MODEL : FORWARD ALGORITHM**

Partiamo da un semplice modello a 2 stati pioggia e sole come mostrato in figura:



I 2 stati rappresentano il tempo in un dato giorno in una citta (in maniera del tutto similare all’esempio visto per HMM), poi abbiamo come variabile osservabile l’umore di Benny che vive in questa città. I 2 stati di umore di Benny dipendono dal tempo del giorno e seguono delle probabilità date come segue:

Diagram, schematic

Description automatically generated

In base a questa rappresentazione che risulta un modello nascosto di Markov semplificato rispetto a quello visto precedentemente, andiamo a definire la matrice di Adiacenza A dello strato nascosto quindi quella inerente al tempo e su di essa con i metodi visti ( Monte carlo, elevamento a potenza della matrice e metodo delle potenze iterate) andiamo a calcolare l’autovettore [pi] e infine diamo una rappresentazione degli umori in base al tempo (ove la 1° riga della matrice da le probabilità rispettivamente per triste dato il giorno di pioggia = 0.8 e allegro dato il giorno di pioggia = 0.2, mentre la 2°riga fornisce triste dato sole = 0.4 e allegro dato sole = 0.6):

Text

Description automatically generated with low confidence

Ora supponiamo di aver osservato una data sequenza di umore di Benny e vogliamo sapere qual è la probabilità di tale sequenza osservata noti i parametri del HMM(ovvero le probabilità):



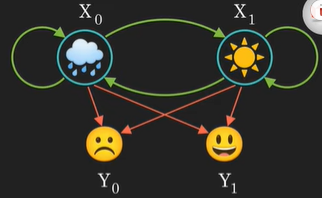
Supponiamo che la sequenza sia questa sovrastante la quale in forma matematica diventa:

A picture containing text, device, meter, gauge

Description automatically generated

Dobbiamo prendere in considerazione tutte le possibili combinazioni meteorologiche che danno vita alla sequenza da noi desiderata.

Dati i valori di x0, x1 agli stati del tempo e y0 e y1 agli stati di umore come segue:



Andiamo a calcolare ad esempio che la precedente osservazione sia data dalla seguente tripla meteorologica:

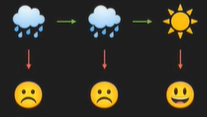


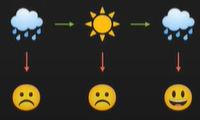
La quale per poter rispondere alla nostra domanda ovvero la probabilità che sia stata tale tripla a generare le osservazione è data dalla formula:



Ora calcoliamo tutte le possibili sequenze che portano alla sequenza osservata, in particolare abbiamo 2 scelte e 3 giorni nei quali compiere la scelta per cui avremo 23 = 8 possibili combinazioni

1:A picture containing background pattern

Description automatically generated 2: 

3: 4: A picture containing schematic

Description automatically generated

5:A screenshot of a video game

Description automatically generated with medium confidence 6: A screenshot of a video game

Description automatically generated

7:A picture containing text

Description automatically generated 8: A picture containing text, device, gauge

Description automatically generated

Di seguito riportiamo le probabilità i forma matematica come elencate dagli stati dal 1° al 8°:

1°



2°



3°



4°



5°



6°



7°



8°

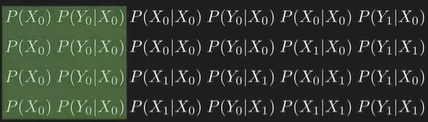


Adesso viene la parte interessante e nuova nel trattare la computazione di queste probabilità tramite un HMM, in particolare se volessimo scoprire la probabilità presa in esempio dovremmo andare a calcolare tutte le possibili combinazione le quale se pur ora non risultino eccessive sono solo 8 in realtà nei modelli usati a scopi scientifici le variabili nascoste e osservate potrebbero essere di gran lunga maggiori in numero e questo calcolo essendo esponenziale porta a notevole complessità.



Complessità che si paga in termini di tempo richiesto per riuscire a calcolare tale risultato, per agevolare il processo e renderlo più efficiente riguardo alla velocità usiamo l’algoritmo forward il quale permette di diminuire i calcoli salvandone alcuni e riutilizzandoli.

Possiamo notare infatti che alcuni calcoli di probabilità sono ripetuti più volte come i seguenti nelle precedenti formule:

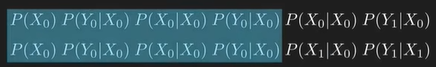


Oppure

Text

Description automatically generated

O ancora queste ripetizioni di calcoli:



Oppure

Graphical user interface, text

Description automatically generated

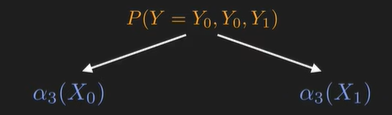
Quindi l’idea è quella di salvare questi risultati e utilizzarli tutte le volte che si necessitano, per farlo abbiamo bisogno di un meccanismo che generalizzi e trovi tali calcoli ripetuti e per farlo ci basiamo su delle relazioni di ricorrenza. Questa relazione può essere in qualche modo ricavata utilizzando le proprietà delle catene di Markov e andando a calcolare la probabilità di un evento in base esclusivamente al suo stato precedente.

Basandosi su tale principio di Markov e utilizzando il concetto di ricorsione diamo vita alla programmazione dinamica che permette il calcolo per ricorsione. Andiamo a vedere graficamente quello che intendiamo:

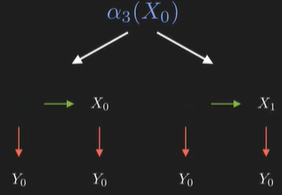
Diagram, schematic

Description automatically generated

I primi 2 step precedenti sono vuoti questo significa che stiamo considerando a sx tutte le triple che finiscono con lo stato X0 e a dx tutte le triple che terminano con X1. Ora supponiamo di conoscere le probabilità date dalle triple precedentemente descritte e sostituiamo la precedente scrittura con alfa3(X0) e alfa3(X1), dove il pedice do alfa sta ad indicare la lunghezza della tripla presa in considerazione e la quale finisce con lo stato passato come argomento della funzione alfa e otteniamo che il precedente grafo può essere riscritto come segue:



In maniera ricorsiva andiamo a calcolare le funzioni alfa:



Questo calcolo può essere fatto per le proprietà della catene di Markov



Continuando con la decomposizione della funzione alfa arriviamo al ground della ricorsione in questo modo:



La formula del calcolo della funzione alfa2 è data dal seguente calcolo di probabilità

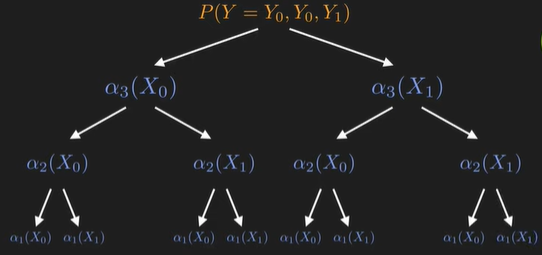


Ma per il calcolo di alfa1 non possiamo effettuare più ricorsione e le probabilità le prediamo direttamente dall’autovettore stazionario pi e dalla matrice delle osservazioni andando a calcolare le seguenti formule:



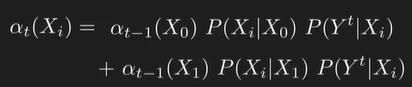


Alla fine del processo di ricorsione avremo il seguente albero :



Attraverso questo albero possiamo calcolare ogni probabilità e conservarla per usi futuri nelle successive predizioni.

Le formule usate derivano dalla generica formula:



A picture containing text, device, meter, gauge

Description automatically generated

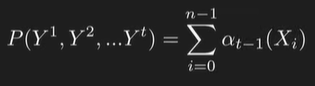
Dove il passo base di ricorsione è dato da:



Che in base all’autovettore stazionario sarà



In fine per calcolare la probabilità sulla sequenza osservata avremo tale formula:



Queste formule viste costituiscono il FORWARD ALGORITHM per efficientare il tempo della computazione nella predizioni dei HMM, vediamo in azione l’algoritmo con il calcolo delle formule per la domanda posta all’inizio e vediamo quale sia la probabilità di osservare la tripla che avevamo definito:

