Modelli Probabilistici per le Decisioni

UniShare

Davide Cozzi
@dlcgold

Indice

Capitolo 1

Introduzione

Questi appunti sono presi a lezione. Per quanto sia stata fatta una revisione è altamente probabile (praticamente certo) che possano contenere errori, sia di stampa che di vero e proprio contenuto. Per eventuali proposte di correzione effettuare una pull request. Link: `https://github.com/dlcgold/Appunti`.

Capitolo 2

Ripasso di Probabilità

Riprendiamo qualche definizione.

Definizione 1. *Definiamo* ***variabile casuale*** *come un'osservazione, un esito o un evento il cui valore è incerto.*

Definizione 2. *Definiamo* ***dominio o spazio degli eventi*** *come l'insieme dei possibili valore che può assumere una variabile casuale.*

Definizione 3. *Definiamo* ***spazio di probabilità o modello di probabilità*** *come uno spazio degli eventi corredato da un assegnamento:*

$$P(\omega), \omega \in \Omega$$

tale che:

- $0 \leq P(\omega) \leq 1$
- $\sum_{\omega} p(\omega) = 1$

con omega evento e Ω spazio degli eventi.

Definizione 4. *Definiamo* ***evento atomico o campione*** *una specificazione completa del calore delle variabili casuali di interesse.*
L'insieme di tutti i possibili eventi atomici è:

- *mutualmente esaustivo (non potendo accadere altro)*
- *mutualmente esclusivo (può accadere solo un evento atomico di quelli possibili)*

Definizione 5. *Definiamo un* ***evento*** *(non atomico) A può essere un qualunque sottoinsieme di Ω tale che:*

$$P(A) = \sum_{\omega \in A} P(\omega)$$

Definizione 6. *Definiamo una **variabile aleatoria** è una variabile che può assumere diversi valori in corrispondenza di altrettanti eventi che costituiscono una partizione dello spazio delle probabilità.*

Si ricorda che, per una variabile a e una b:

- $0 \leq P(a) \leq 1$
- $P(\top) = 1$ e $P(\bot) = 0$
- $P(a \vee b) = P(a) + p(b) - p(a \wedge b)$

Definizione 7. *Definiamo una **probabilità condizionata** rappresenta la verosimiglianza che un evento a si verifichi se b si verifica e si denota con:*

$$P(a|b)$$

Si ha quindi la specifica che alcuni eventi rendono altri eventi più o meno verosimili.
*Si parla quindi di eventi **dipendenti**.*

Definizione 8. *Due eventi sono **indipendenti** se un evento non influisce sulla realizzazione dell'altro:*

$$P(a|b) = P(a)$$

Si ha quindi la seguente regola.

Teorema 1 (Regola del prodotto)**.** *Possiamo calcolare che due eventi si verifichino contemporaneamente tramite la probabilità condizionata e quella dei singoli eventi:*

$$P(a, b) = P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$$

*Con $P(a, b) = P(a \wedge b)$ è detta **probabilità congiunta** ("=" perché sono due modi per scriverla).*

Posso fare la tabella dei vari eventi condizionati.

Teorema 2 (Regola della somma)**.** *Si ha che, avendo la tabella degli eventi:*

$$P(x) = \sum_{y} P(x, y)$$

*con $P(x)$ detta **probabilità marginale**.*

La somma di tutte le possibili combinazioni di eventi, quindi dei valori della tabella, deve dare 1.
Su slide esempio di uso di quanto detto, dove si arriva al teorema di Bayes.
Si vuole infatti passare dal conoscere $P(a|b)$ al consocere $P(b|a)$.

Teorema 3 (Teorema di Bayes)**.** *Il teorema enuncia che:*

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

Avendo:

- *$P(h)$ che è la probabilità conosciuta a priori di h. Tale probabilità riflette qualsiasi conoscenza di base sulla possibilità che h sia corretta*
- *$P(D)$ che è la probabilità conosciuta a priori di D, ovvero la probabilità che D sia osservato*
- *$P(D|h)$ che è la probabilità di osservare D in presenza dell'ipotesi h*
- *$P(D|h)$ che è la probabilità a posteriori di h. Tale probabilità riflette la "confidenza" di avere h dopo che D è stato osservato*

In altri termini, avendo:

$$P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$$

ho che:

$$P(a|b)P(b) = P(b|a)P(a)$$

arrivando a dire che:

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

notando la correlazione tra probabilità congiunta e Bayes.
Che è il punto fondamentale della moderna teoria dell'intelligenza artificiale in quanto permette di raccogliere l'evidenza senza poi usare le tabelle delle probabilità congiunte, che sarebbero difficilissimi da osservare. Se pensiamo ad alcuni eventi come cause "nascoste" non necessariamente osservabili se modelliamo la verosimiglianza degli eventi osservabili date le cause nascoste si ha:

$$P(causa|effetto) = \frac{P(effetto|causa)P(causa)}{P(effetto)}$$

Si ha quindi un modello per inferire/derivare la verosimiglianza della causa nascosta e quindi rispondendo a:

$$P(causa|effetto)$$

Avendo quindi la probabilità di una causa dato un effetto. *Dato l'effetto modello la causa.*
Se non si ha una delle due probabilità a priori posso stimare per poi normalizzare. In altri termini il denominatore $P(D)$ è spesso solo una quantità di normalizzazione, essendo spesso difficile da stimare.

Capitolo 3

Modelli Probabilistici

Nel passato si sono usati **sistemi a regole**, dove codificando tutto quello che può succedere si cercava di giungere ad una decisione. Questo però era molto dispendioso, si arrivava o a vero a falso, senza via di mezzo, e si dovevano avere dati ipoteticamente completi e sicuri in partenza. Si parla in questo caso di **modelli logici**.
Viviamo in un'era dove si hanno molti dati, sia in ambito sociale, che di business che scientifico. Questi dati devono essere analizzati al fine di poter prendere **decisioni** e per farlo si deve per capire la situazione in cui ci si trova e spesso posso capirlo solo osservando i dati, non osservando la variabile specifica. Dalle osservazioni dobbiamo inferire il valore di variabili "nascoste". Spesso si ha a che fare con dati non completi, non consistenti, spesso errati, con rumore di trasmissione etc. . .
Tali dati sono comunque evidenze per percepire la situazione in cui ci si trova. L'obiettivo del corso è fornire strumenti modellistici per rappresentare l'incertezza nel modello, incertezza per struttura e parametri, e per rappresentare in termini probabilistici gli errori nei dati. Si vuole quindi implementare algoritmi di "ragionamento", automatizzati e adattivi, oltre che robusti e scalabili.
I **modelli probabilistici** sono anche detti **modelli generativi**. Si usa la teoria delle probabilità per esprimere incertezza e rumore associati al modello e ai dati, soprattutto usando la teoria Bayesana, per fare previsione e adattare i modelli. Questi modelli permettono di partire da una "credenza" iniziale, anche soggettiva, per poi raccogliere evidenze aggiustando tale modello.
I modelli probabilistici sono anche modelli di machine learning, in quanto apprendono.
Bisogna quindi partire dalle osservazioni generate rispetto ad un valore di variabile per poi inferire tale variabile (ad esempio parto dai risultati di un gioco per capire quando è bravo il giocatore, che non è una variabile che posso

sapere a priori). Si parte dai dati e si arriva al valore della variabile che ha generato questi dati (per questo *modello generativo*). Man mano che raccolgo informazioni raffino il modello, più o meno come fa un essere umano ("più rispondi alle domande all'orale e più il docente capisce il tuo voto, anche se alla fine non si ha la certezza che il voto rispecchi la preparazione"). I dati possono non portare alla certezza, ma più dati si hanno e più ci si avvicina, riducendo l'incertezza.
Un esempio pratico è il modello **Elo** (nato per gli scacchi) da cui deriva quello usato da *Xbox* per capire come appaiare giocatori online in base alle skill. Il valore di bravura viene rappresentato come una distribuzione, in primis con una Gaussiana, con una certa media e una certa varianza/deviazione standard, quindi solo due numeri. Cambiare il modello significa solo cambiare quei due valori. Per confrontare due giocatori capisco la distribuzione a partire dai dati del giocatore che si hanno, diminuendo l'incertezza all'aumentare dei dati. Con il modello probabilistico poi, a partire dal risultato modificherò le distribuzioni di partenza, cambiando la percezione su essi. Nel tempo posso tenere aggiornato i modelli probabilistici che rappresentano una certa variabile e usarli per fare confronti (ad esempio confrontando due giocatori per poi fare l'appaiamento).
Con i modelli probabilistici si ha una capacità espressiva maggiore di quella di un modello logico, avendo le distribuzioni di probabilità e potendo anche usare varie soglie.
Un *modello generativ*o parte dalle probabilità a priori e può "generare" possibili eventi, generando campioni verosimili con una certa distribuzione statistica.
Nella vita reale si osservano degli accadimenti e studiandoli si può risalire alla probabilità degli eventi, tramite l'approccio frequentista.

3.1 Incertezza

Si introduce quindi l'**incertezza**. Non sempre si ha a che fare con dati "certi" e precisi, che possono portare con più facilità ad una certa decisione, potendo giungere ad una decisione **ottimale** senza alcun dubbio su quale essa sia.
Con l'**incertezza** sui dati bisogna modificare l'idea di **soluzione ottimale**. Si arriva a dover capire quale sia la **soluzione ottima** in un contesto dove "non si sa cosa succederà", partendo da dati incerti.
Si ha che:

- un evento osservato può avere molte cause

- la verosimiglianza di un'ipotesi sulla causa cambia man mano che si raccolgono pezzi di evidenza
- usando modelli probabilistici di ragionamento possiamo calcolare quanto probabile è una certa ipotesi. Si ipotizza che le fonti di incertezza siano quantificabili

Vari esempi di vita in merito sulle slide.
Spesso si ha un approccio "frequentista", valutando la frequenza di un evento per capire la probabilità che tale evento accada, inferendo così una distribuzione di probabilità dalla frequenza con la quale si osservano i dati. Questo è piuù o meno come funziona il cervello umano ma bisogna fare la stessa cosa con un calcolatore e per questo ci verrà incontro il **teorema di Bayes**.
Si ha inoltre che un sistema che considera anche l'incertezza, che è presente in moltissime situazioni, dovrebbe funzionare meglio di uno che non lo fa ma ci serve in primis un modo per rappresentare l'incertezza stessa.
Più parametri ha il modello e più è difficile rappresentarlo.
Si ha il **Degree of Belief** che è una probabilità a priori sono ricavate da:

- osservazioni statistiche
- regole generali e conosciute
- combinazioni di sorgenti di evidenza

In ogni caso si hanno quindi evidenze empiriche.