

Bayesian constrained-based structure - appunti chiamata conoscitiva

Apprendimento di grafi: strutturare modelli che catturano la dipendenza tra le variabili: utilizzeremo i grafi, dove ogni arco rappresenta una relazione.

Finora il metodo più usato è quello score-based. Noi andremo ad usare in cui gli archi determinano una relazione di dipendenza o indipendenza, che diventa condizionale se è presente una freccia (cioè l'arco è direzionato). Il secondo metodo è meno preciso del primo ma è molto più conveniente dal punto di vista computazionale.

Uno dei metodi per ricostruire un grafo è tramite i test di indipendenza condizionale: qui entra in gioco il PC algorithm (il funzionamento è spiegato in Slides2, JMLR (2007) e Pcalg (che spiega il pacchetto R)).

Il nostro obiettivo è quello di modificare i test frequentasti e renderli bayesiani (che performi in caso di alta dimensionalità).

L'idea di partenza è di calcolare il Bayes-factor tra due modelli (assumendo gaussianità, almeno inizialmente), quindi trovare la marginal likelihood partendo da due ipotesi differenti:

- modello con variabili dipendenti condizionatamente ad altre,
 - Modello con variabili indipendenti condizionatamente ad altre;
- così otterremo anche una misura di incertezza sulle ipotesi iniziali (rappresentata dalla presenza di una freccia tra due nodi oppure no).

L'output dell'algoritmo sarà un grafo parzialmente direzionato (alcuni archi saranno direzionati, altri no).

Attualmente ci sono metodi per fare inferenza di grafi che funzionano quando sono presenti poche variabili, manca tutta la parte che riguarda inferenza ad alta dimensionalità.

La letteratura sugli approcci bayesiani si trova in Slides4, da guardare la parte che riguarda il Gaussiano (no categorica); una implementazione del metodo è presentata nel paper BCDAG che descrive il relativo pacchetto.