

Multivariate Analysis and Statistical Learning

Implementazione PC Algorithm

Authors: Alex Foglia, Tommaso Puccetti

Università degli Studi di Firenze

21/12/2018

Accenni teorici (1)

- Le reti Bayesiane possono essere rappresentate attraverso l'utilizzo di un grafo aciclico diretto, o **DAG**.
- Un DAG un grafo diretto in cui non compaiono cicli, dove per ciclo si intende un qualunque cammino finito che, a partire da un nodo iniziale v termini in v .

Accenni teorici (2)

Sia $G = (V, E)$ un DAG su un insieme finito $X = \{X_v \mid v \in V\}$ di variabili casuali, allora:

$$\forall u, v \in V \text{ non adiacenti} \mid v \in nd(u) \Rightarrow u \perp\!\!\!\perp v \mid nd(u) - v$$

Dove $nd(u)$ l'insieme dei nodi *non discendenti* di u , ossia tutti quei nodi u' per cui non esiste un cammino da u a u' .

PC-Algorithm

Dato un insieme di variabili con distribuzione di probabilità congiunta gaussiana, possibile imparare il DAG sottostante al campione osservato attraverso l'utilizzo del **PC-Algorithm**.
Esso composto da due sotto-funzioni che risolvono due diversi problemi:

- 1 La costruzione dello scheletro (o grafo morale)
- 2 La costruzione del DAG a partire da un dato scheletro

Pseudocode: generazione dello scheletro

Algorithm 1: The PC-algorithm for the skeleton

Input: z-transform of estimated partial correlations, tuning parameter α

Output: Skeleton of CPDAG G , separation sets S (used later for directing the skeleton)

```
1 Form the complete undirected graph  $\tilde{G}$  on the set  $\{1, \dots, p\}$ ;
2  $l = -1$ ;  $G = \tilde{G}$ ;
3 repeat
4    $l = l + 1$ ;
5   repeat
6     Select an ordered pair of adjacent variables  $i, j$  in  $G$  such that
        $|adj(i, G) \setminus \{j\}| \geq l$ ;
7     repeat
8       Choose  $K \subseteq adj(i, G) \setminus \{j\}$  with  $|K| = l$ ;
9       if  $\sqrt{n - |K| - 3} |Z(i, j \mid K)| \leq \Phi^{-1}(1 - \alpha/2)$  then
10        Delete edge  $i, j$ ;
11        Denote this new graph by  $G$ ;
12        Save  $K$  in  $S(i, j)$  and  $S(j, i)$ ;
13    until edge  $i, j$  is deleted or all  $K \subseteq adj(i, G) \setminus \{j\}$  with  $|K| = l$ 
      have been chosen ;
14  until all ordered pairs of adjacent variables  $i$  and  $j$ , such that
     $|adj(i, G) \setminus \{j\}| \geq l$  and  $K \subseteq adj(i, G) \setminus \{j\}$  with  $|K| = l$ , have
    been tested for conditional independence ;
15 until for each ordered pair of adjacent nodes  $i, j$ :  $|adj(i, G) \setminus \{j\}| < l$ ;
```

Pseudocodice: costruzione del DAG

Algorithm 2: The PC-algorithm: extending the skeleton to a CPDAG

Input: Skeleton G of CPDAG, separation sets S

Output: CPDAG

```
1 forall pairs of nonadjacent variables  $i, j$  with common neighbor  $k$  do
2   if  $k \notin S(i, j)$  then
3      $\quad$  Replace  $i - k - j$  in Skeleton of  $G$  by  $i \rightarrow k \leftarrow j$ ;
4 repeat
5   R1 Orient  $j - k$  into  $j \rightarrow k$  whenever there is an arrow  $i \rightarrow j$  such
      that  $i$  and  $k$  are nonadjacent;
6   R2 Orient  $i - j$  into  $i \rightarrow j$  whenever there is a chain  $i \rightarrow k \rightarrow j$ ;
7   R3 Orient  $i - j$  into  $i \rightarrow j$  whenever there are two chains
       $i \rightarrow k \rightarrow j$  and  $i \rightarrow l \rightarrow j$  such that  $k$  and  $l$  are nonadjacent;
8 until no more orienting of undirected edges is possible by the rules R1
      to R3;
```
