

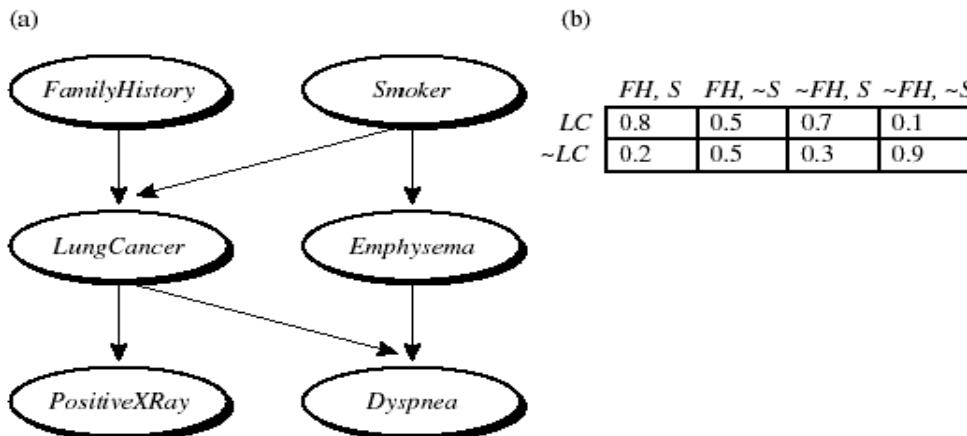
# BAYESIAN CLASSIFICATION

- Clasificatori statistici
- Pot face o predictie asupra probabilitatii apartenentei la o clasa
- Se bazeaza pe teorema lui Bayes
- Studiile au aratat ca un clasificator Bayes naiv este comparabil ca si performanta cu arborii de decizie si cu unele retele neuronale
- Clasificatorul naiv presupune ca efectul unei valori a unui atribut asupra unei clase este independent de valoarea altor attribute (abordarea naiva)

- Teorema lui Bayes

- X – persoana de 35 ani, cu un venit de 40000\$
  - H – ipoteza ca aceasta persoana va cumpara un calculator
  - $P(H|X)$  = probabilitatea ca X sa cumpere un calculator, cunoscandu-i varsta si venitul
  - $P(H)$  = probabilitatea ca un client (orice client) sa cumpere un calculator
  - $P(X|H)$  = probabilitatea ca X sa aiba 35 ani si un venit de 40000, cunoscandu-se faptul ca va cumpara un calculator
  - $P(X)$  = probabilitatea ca o persoana sa aiba 35 ani si un venit de 40000

- $P(X|H)$ ,  $P(X)$ ,  $P(H)$  – pot fi estimate din setul de date
- Teorema: 
$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}.$$
- Clasificatorul bayesian calculeaza  $P(C_i|X)$  pentru fiecare clasa  $C_i$ , alegand clasa cu cea mai mare probabilitate
- Aceste clasificatori sunt foarte eficienti atunci cand exista acea independenta presupusa
- In practica insa exista dependente intre variabile, dependente de care tin cont retelele bayesiene
- Retea bayesiana (vezi figura din slide-ul urmator)
  - model grafic al relatiilor cauzale pe care poate fi facuta invatarea
  - Are doua componente: un graf aciclic si un tabel de probabilitati conditionate pentru fiecare variabila



**Figure 6.11** A simple Bayesian belief network: (a) A proposed causal model, represented by a directed acyclic graph. (b) The conditional probability table for the values of the variable *LungCancer* ( $LC$ ) showing each possible combination of the values of its parent nodes, *FamilyHistory* ( $FH$ ) and *Smoker* ( $S$ ). Figure is adapted from [RBKK95].

**Referinta figura:** J. Han, M. Kamber, *Data Mining: Concepts and Techniques*, 2nd Edition, Morgan Kaufmann, 2006.