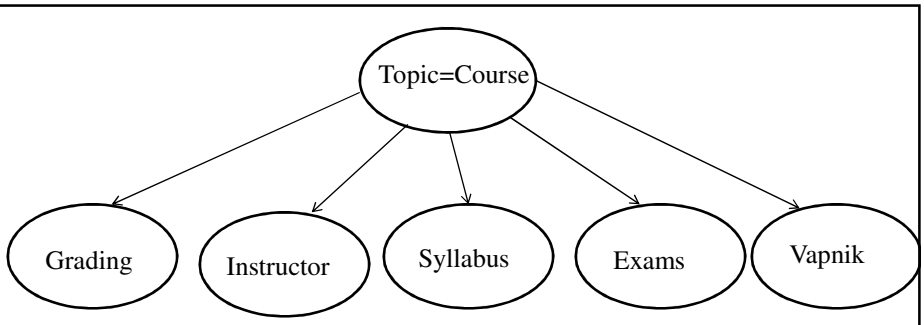
Bayesian Networks

Reading: Chapter 14, Sections 14.1-14.6 of Russell & Norvig, *Artificial Intelligence: A Modern Approach* (on e-reserve)

Graphical Models

- Bayesian networks
- Conditional random fields
- etc.



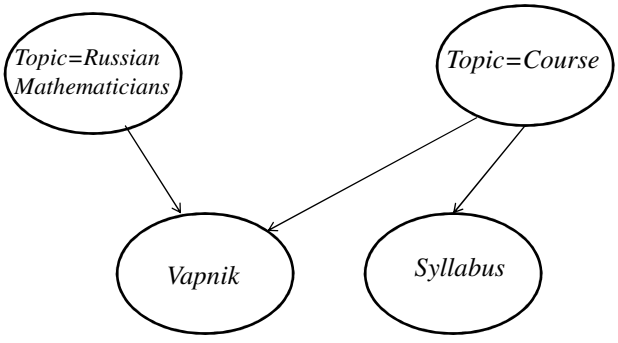
$$P(Grading \wedge Instructor \wedge Syllabus \wedge Exams \wedge Vapnik)$$
$$= P(Course)P(Grading \mid Course)P(Instructor \mid Course)P(Syllabus \mid Course)P(Exams \mid course)$$
$$P(Vapnik \mid Course)$$

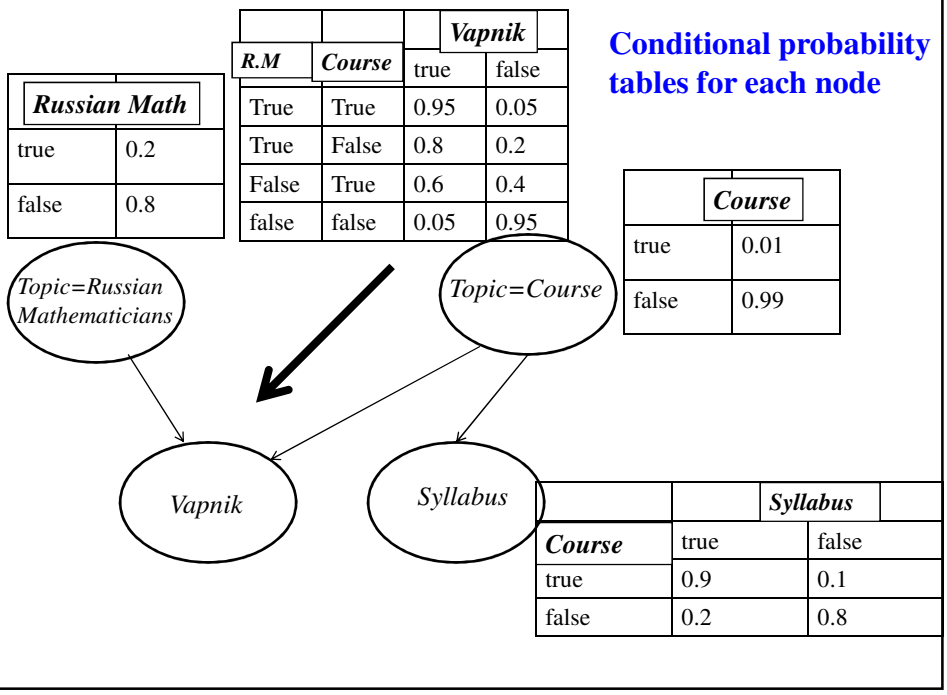
Naive Bayes:

$$P(C = c_j \mid X_1 = x_1, ..., X_1 = x_n) = P(C = c_j)\prod_i P(X_i = x_i \mid C = c_j)$$

Bayesian networks

- Idea is to represent dependencies (or causal relations) for all the variables so that space and computation-time requirements are minimized.





Conditional probability tables for each node

Semantics of Bayesian networks

- If network is correct, can calculate full joint probability distribution from network.

$$P((X_1 = x_1) \wedge (X_2 = x_2) \ldots \wedge (X_n = x_n))$$

$$= \prod_{i=1}^{n} P(X_i = x_i \mid parents(X_i))$$

where parents(X_i) denotes specific values of parents of X_i.

Example

- Calculate

$$P[(RussianMath = t) \wedge (Course = f) \wedge (Vapnik = f) \wedge (Syllabus = f)]$$

Examples

What is unconditional (marginal) probability that *Vapnik* is true?

What is the unconditional (marginal) probability that *Russian Mathematicians* is true?

Different types of inference in Bayesian Networks

Causal inference

Evidence is cause, inference is probability of effect

Example:
Instantiate evidence *Course= true.* What is *P*(*Syllabus* | *Course*)?

Diagnostic inference

Evidence is effect, inference is probability of cause

Example: Instantiate evidence *Syllabus* = true. What is *P*(*Course* | *Syllabus*)?

Example: What is P(*Course*|*Vapnik*)?

Inter-causal inference

Explain away different possible causes of effect

Example: What is P(*Course*|*Vapnik*,*RussianMath*)?

Why is P(*Course*|*Vapnik*,*RussianMath*) < P(*Course*|*Vapnik*)?

Complexity of Bayesian Networks

For n random Boolean variables:

- Full joint probability distribution: 2^n entries
- Bayesian network with at most k parents per node:
  - Each conditional probability table: at most 2^k entries
  - Entire network: $n\ 2^k$ entries

What are the advantages of Bayesian networks?

- Intuitive, concise representation of joint probability distribution (i.e., conditional dependencies) of a set of random variables.
- Represents "beliefs and knowledge" about a particular class of situations.
- Efficient (?) (approximate) inference algorithms
- Efficient, effective learning algorithms

Issues in Bayesian Networks

- Building / learning network topology
- Assigning / learning conditional probability tables
- Approximate inference via sampling