Introduction to Artificial Intelligence II

Short Lecture Notes Part II — Search, Logic and Probabilistic Reasoning

Ninad Joshi

Notation. $P(A \mid B)$ denotes conditional probability, $\models$ semantic entailment, and g, h, f denote search costs.

Contents

1 Lecture 1: Agents and Maze Navigation **2**

2 Lecture 2: State-Space Search and Heuristics **3**

3 Lecture 3: Propositional Logic and Entailment **4**

4 Lecture 4: First-Order Logic and Knowledge Representation **5**

5 Lecture 5: Probability, Bayes Rule and Independence **6**

6 Lecture 6: Bayesian Networks and Conditional Tables **7**

7 Lecture 7: Inference and Approximation in Bayesian Networks **8**

8 Lecture 8: Decision Networks and Rational Choice **9**

1 Lecture 1: Agents and Maze Navigation

An *agent* observes an environment through sensors and selects actions. A reflex agent acts on current percepts; a model-based agent retains information about previous states. In a maze, an agent may choose among four directions, subject to walls. A deterministic wall-following rule can be efficient in some mazes but fail in others. Randomised agents can escape repeated local behaviour, although their path length varies between runs.

$$\pi(a \mid s) = P(A_t = a \mid S_t = s), \qquad \bar{N} = \frac{1}{R}\sum_{r=1}^{R} N_r.$$

Compare agents using the mean, minimum and maximum steps, and report failures separately. Avoid interpreting a single lucky run as typical performance. An agent that avoids immediate backtracking except at dead ends is a useful intermediate strategy.

Problem sheet — Lecture 1

1. Explain the difference between a deterministic reflex agent, a partly randomised agent and a fully randomised agent in a maze.

2. Design a maze-navigation rule that avoids immediately reversing direction except at dead ends. Explain when it may fail.

3. An agent needs 24, 24, 24, 24 steps in four runs; another needs 12, 30, 42, 20. Compare mean, minimum and maximum and discuss reliability.

4. How would you compare four agents across three mazes if some runs never reach the exit?

2 Lecture 2: State-Space Search and Heuristics

A search problem specifies an initial state, possible actions, a transition model, a goal test and a path-cost function. The search *frontier* stores states that may be expanded next. Breadth-first search prioritises depth; uniform-cost search prioritises accumulated cost $g(n)$; greedy best-first search prioritises estimated remaining cost $h(n)$; and A* uses

$$f(n) = g(n) + h(n).$$

A heuristic is *admissible* if it never overestimates the true remaining cost. With nonnegative costs and appropriate handling of repeated states, A* can find an optimal path under standard admissibility/consistency conditions. A cycle check prevents repeated traversal of the same state.

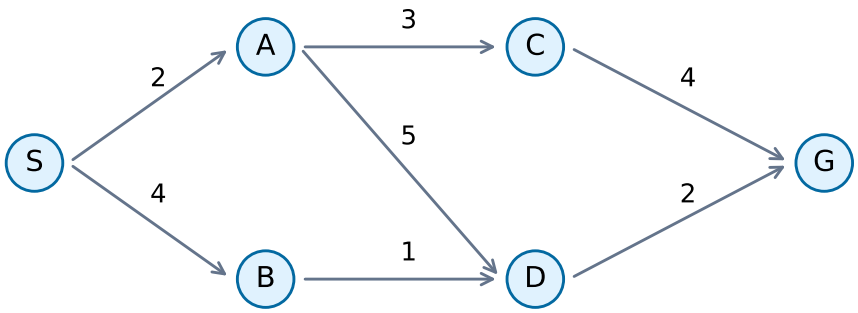


Figure: A small weighted graph for comparing search strategies; edge labels denote costs.

Problem sheet — Lecture 2

1. For the graph in the figure, find a minimum-cost path from S to G using uniform-cost search. Record expansion order.

2. Choose admissible heuristic values for the graph and show the A* priorities $g + h$.

3. Under which conditions does breadth-first search behave like uniform-cost search? Explain whether uniform-cost search is a special case of A*.

4. Formulate the three-missionaries and three-cannibals river-crossing problem as a state-space search, including legal states and actions.

3 Lecture 3: Propositional Logic and Entailment

Propositional logic represents statements using Boolean variables and connectives $\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$. A *model* assigns truth values to variables. A formula is satisfiable if at least one model makes it true; it is valid if every model makes it true. Semantic entailment $\alpha \models \beta$ means that every model of α is also a model of β.

$$A \Rightarrow B \equiv \neg A \vee B, \qquad \neg(A \wedge B) \equiv \neg A \vee \neg B.$$

Truth tables offer a direct method for checking entailment, equivalence and satisfiability. With n Boolean variables, there are 2^n possible assignments before constraints are imposed. Clausal form is convenient for automated reasoning and resolution.

Problem sheet — Lecture 3

1. Determine, with counterexamples or proofs, whether $(A \wedge B) \models (A \Leftrightarrow B)$ and $(A \Leftrightarrow B) \models (A \vee B)$.

2. For Boolean variables A, B, C, D, count the satisfying assignments of $B \vee C$ and of $\neg A \vee \neg B \vee \neg C \vee \neg D$.

3. Translate the unicorn statements into propositional logic: mythical implies immortal; not mythical implies mortal and mammal; immortal or mammal implies horned; horned implies magical. Determine which conclusions are entailed.

4. Use a truth table to test $(C \vee (\neg A \wedge \neg B)) \equiv ((A \Rightarrow C) \wedge (B \Rightarrow C))$.

4 Lecture 4: First-Order Logic and Knowledge Representation

Propositional logic cannot directly express statements about arbitrary objects. First-order logic introduces constants, variables, predicates, functions and quantifiers. The universal quantifier $\forall x$ means "for every x"; the existential quantifier $\exists x$ means "there exists an x".

$$\forall x\,[\mathrm{Student}(x) \Rightarrow \mathrm{Learns}(x)], \qquad \exists x\ \mathrm{Student}(x).$$

The order of quantifiers matters: $\forall x \exists y\, R(x,y)$ generally differs from $\exists y \forall x\, R(x,y)$. Inference combines premises with valid rules to derive new statements. Knowledge representation requires an explicit domain and carefully chosen predicates.

Problem sheet — Lecture 4

1. Express in first-order logic: every student attends at least one lecture; some students attend every lecture.

2. Explain the difference between $\forall x \exists y\, R(x,y)$ and $\exists y \forall x\, R(x,y)$ using a concrete example.

3. Create a small knowledge base describing people invited to a party, and formulate two logical queries.

4. Describe how a Boolean logic circuit for $(A \vee B) \wedge \neg C$ corresponds to a propositional formula.

5 Lecture 5: Probability, Bayes Rule and Independence

Uncertainty is represented by probability distributions. Conditional probability is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0.$$

Bayes' theorem reverses conditioning:

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}.$$

The denominator can be obtained using total probability. Events X and Y are independent if $P(X, Y) = P(X)P(Y)$; conditional independence given Z requires $P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$. Independence conditional on a known parameter need not imply marginal independence when that parameter is uncertain. Expected utility or payoff is a probability-weighted average, $\mathbb{E}[U] = \sum_i p_i U_i$.

Problem sheet — Lecture 5

1. A medical test has sensitivity and specificity 99%, and disease prevalence is 1 in 100,000. Find $P(\text{disease} \mid \text{positive})$ using Bayes' theorem.

2. A coin has an unknown head probability θ. Discuss whether successive tosses are independent conditional on θ and after integrating over uncertainty in θ.

3. A slot machine has three independent reels with four equally likely symbols. Three bars pay 21 coins, three bells 16, three lemons 5, three cherries 3, first two cherries and non-cherry third 2, and first cherry with non-cherry second 1. Calculate expected payout and probability of nonzero payout.

4. For three Boolean variables X, Y, Z, compare the degrees of freedom of a full joint distribution and a model assuming $X \perp Y \mid Z$.

6 Lecture 6: Bayesian Networks and Conditional Tables

A Bayesian network is a directed acyclic graph (DAG) whose nodes are random variables. An arrow denotes a direct probabilistic dependence encoded by a conditional probability table (CPT). Its joint distribution factorises as

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Pa}(X_i)).$$

A simple example chooses one of three biased coins, then records three tosses. The tosses are independent conditional on coin identity, but generally dependent marginally.

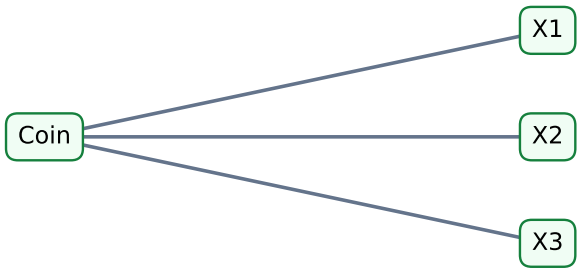


Figure: A latent coin identity influences three observed tosses. Network structure encodes conditional independences and can greatly reduce the number of probabilities required.

Problem sheet — Lecture 6

1. Three equally likely coins have head probabilities 0.30, 0.60 and 0.75. Draw a Bayesian network for three tosses and give the relevant CPTs.

2. For the same model, observe two heads and one tail. Write the posterior probability for each coin and identify the most likely coin.

3. Construct a Bayesian network for reactor temperature, sensor failure, measured temperature, alarm failure and alarm signal. Explain which dependencies are essential.

4. Two astronomers count the same stars with independent instrument failures. Propose a Bayesian-network structure and justify conditional independence assumptions.

7 Lecture 7: Inference and Approximation in Bayesian Networks

Probabilistic inference computes a query distribution from observations, such as $P(Q \mid e)$. Hidden variables can be eliminated by summation:

$$P(Q, e) = \sum_{\mathbf{z}} P(Q, e, \mathbf{z}).$$

Exact inference may become expensive as dependencies increase. An approximate network omits selected edges and thereby imposes additional conditional independences. Its quality can be assessed by comparing the resulting joint probabilities with a reference joint distribution. For example, one simple discrepancy measure is

$$D_1(P, Q) = \sum_{\mathbf{x}} |P(\mathbf{x}) - Q(\mathbf{x})|.$$

Rounding CPT entries can slightly affect normalization; retain adequate precision during calculations.

Problem sheet — Lecture 7

1. Given three binary variables (study time, homework preparation, exam result), draw a fully expressive Bayesian network and explain how its CPTs reproduce any joint distribution.

2. Compare $P(L)P(H)P(K \mid L, H)$ and $P(H)P(L \mid H)P(K \mid L)$. State the independence assumptions and how you would measure approximation error.

3. Explain the distinction between a polytree and a multiply connected Bayesian network. Why can multiple paths complicate inference?

4. Derive a formula for $P(X_1 \mid X_3)$ in the coin network after summing over coin identity and the unobserved tosses.

8 Lecture 8: Decision Networks and Rational Choice

A decision network combines chance variables, decision nodes and a utility function. Decisions should maximise expected utility, not necessarily the probability of a favourable outcome. For decision d and uncertain state s,

$$\mathrm{EU}(d) = \sum_{s} P(s \mid d) U(d, s), \qquad d^{*} = \arg\max_{d} \mathrm{EU}(d).$$

A student's choice to buy a textbook affects understanding, exam performance and monetary cost. When a book may be used in the exam, its influence on exam success need not be mediated only by understanding.

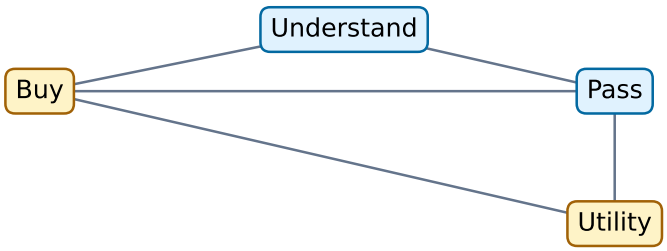


Figure: A simplified influence structure. A decision can influence multiple chance variables and utility.

Problem sheet — Lecture 8

1. A book costs 100 monetary units; passing an exam yields utility 2000. Given $P(M \mid B) = 0.9$, $P(M \mid \neg B) = 0.7$, and $P(P \mid B, M) = 0.9$, $P(P \mid B, \neg M) = 0.5$, $P(P \mid \neg B, M) = 0.8$, $P(P \mid \neg B, \neg M) = 0.3$, compute the expected utility of each decision.

2. Draw the decision network for the textbook example, marking chance, decision and utility nodes.

3. Explain how a change in book price could reverse the preferred decision, even when all probabilities remain unchanged.

4. Give an example where maximising the probability of success differs from maximising expected utility.