Chapter 9
Innleiðing í tilgjørdum viti

Notes

Contents

1 Uncertainty

Uncertainty arises when an agent lacks complete information about the world. It reflects the agent's **degree of belief** in various propositions.

- Uninformed agents must make decisions with incomplete information.
- Probability is used to quantify uncertainty.
- **Example:** Betting on a dice roll involves assigning probabilities to each possible face of the die.

2 Probability

Probability represents a **subjective measure of belief**. It helps agents reason under uncertainty.

Random Variables

A random variable can take different values due to randomness.

- Dice: $\text{domain}(X) = \{1, 2, 3, 4, 5, 6\}$
- Weather: $\text{domain}(\text{Weather}) = \{\text{sun, cloud, rain, wind, snow}\}$
- Traffic: $\{\text{none, light, heavy}\}$
- Flight status: $\{\text{on time, delayed, cancelled}\}$

Possible Worlds: A complete assignment of values to all variables. Denoted by ω, with $P(\omega)$ representing the probability of that world.

3 Axioms of Probability

1. Non-negativity: $0 \leq P(a)$
2. Normalization: $P(\text{true}) = 1$
3. Additivity (Mutual exclusivity): $P(a \vee b) = P(a) + P(b)$ if $a \wedge b = \text{false}$

4 Types of Probability

Unconditional (Prior) Probability

Belief in a proposition before seeing any evidence. **Example:** $P(\text{Rain}) = 0.3$

Conditional Probability

Belief in a proposition given some evidence. Notation: $P(a \mid b)$ **Example:** $P(\text{Traffic} = \text{heavy} \mid \text{Rain})$

Conditioning and Evidence

Evidence e eliminates possible worlds where e is false. We update our beliefs using:

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

- **Prior**: Belief before evidence
- **Likelihood**: Probability of evidence given a hypothesis
- **Posterior**: Updated belief

5 Probability Distribution

Definition: A function assigning probabilities to each value in a random variable's domain.

Flight Status	Probability
On Time	0.6
Delayed	0.3
Cancelled	0.1

6 Independence

Two events are independent if knowledge of one does not affect the other.

$$P(a \wedge b) = P(a) \cdot P(b)$$

Example: Rolling two dice.

7 Bayes' Rule

$$P(a \mid b) = \frac{P(b \mid a) \cdot P(a)}{P(b)}$$

Example: Medical Diagnosis

- $P(\text{Disease}) = 0.01$
- $P(\text{Positive Test} \mid \text{Disease}) = 0.99$
- $P(\text{Positive Test}) = 0.05$

$$P(\text{Disease} \mid \text{Positive}) = \frac{0.99 \cdot 0.01}{0.05} = 0.198$$

8 Causal vs. Evidential Reasoning

Causal Reasoning

Inferring effects from causes. **Example:** $P(\text{Wet} \mid \text{Rain})$

Evidential Reasoning

Inferring causes from effects. **Example:** $P(\text{Rain} \mid \text{Wet})$

9 Rules of Probability

- Negation: $P(\neg a) = 1 - P(a)$
- Inclusion-Exclusion: $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$
- Marginalisation: $P(a) = P(a \wedge b) + P(a \wedge \neg b)$
- Law of Total Probability: $P(a) = P(a \mid b)P(b) + P(a \mid \neg b)P(\neg b)$

10 Conditional Independence

Variables A and B are conditionally independent given C if:

$$P(A \mid B, C) = P(A \mid C)$$

Example: Given a disease C, symptoms A and B may be conditionally independent.

11 Bayesian Networks

A Bayesian network is a **Directed Acyclic Graph (DAG)** representing dependencies between variables.

- Nodes: Random variables
- Arrows: Causal/statistical dependencies
- Each node has a conditional probability table (CPT)

Structure:

For each node X:

$$P(X \mid \text{Parents}(X))$$

Example Network:

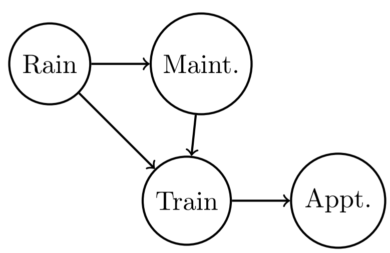
$$\text{Rain} \rightarrow \text{Maintenance} \rightarrow \text{Train} \rightarrow \text{Appointment}$$

Parents:

- Train: Rain, Maintenance
- Appointment: Train

CPTs:

- $P(\text{Rain})$
- $P(\text{Maintenance} \mid \text{Rain})$
- $P(\text{Train} \mid \text{Rain}, \text{Maintenance})$
- $P(\text{Appointment} \mid \text{Train})$



12 Joint Probability using Chain Rule

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{Parents}(X_i))$$

Example with Previous Network:

$P(\text{Rain}, \text{Maintenance}, \text{Train}, \text{Appointment}) = P(\text{Rain}) \cdot P(\text{Maintenance} \mid \text{Rain}) \cdot P(\text{Train} \mid \text{Rain}, \text{Maintenance}) \cdot P$

13 Inference in Bayesian Networks

Inference involves using known knowledge (evidence) to deduce new knowledge (unknown or hidden variables).

Steps to Construct a Bayesian Network

1. **What is observed?** (Evidence)
2. **What do you want to find out?** (Query)
3. **What other variables simplify the model?** (Hidden or latent variables)

Hidden Variables

Hidden variables are unobserved variables that affect the probabilities of observed or queried variables.

Example:

$$\alpha \cdot P(\text{Appointment}, \text{Light}, \text{No})$$

Here, α is a normalization constant that ensures the resulting distribution sums to 1.

14 Marginalisation and Normalisation

- **Marginalisation:**
$$P(a) = P(a, b) + P(a, \neg b)$$
- **Conditional Inference:**
$$P(X \mid e) = \alpha \cdot P(X, e) = \alpha \cdot \sum_y P(X, e, y)$$
where y ranges over hidden variables.
- α is the normalisation constant.

Challenges

- Inference by enumeration scales poorly with many variables. - The complexity of exact inference depends heavily on the network's structure.

15 Exact and Approximate Inference Methods

Exact Inference Methods

- **Inference by Enumeration**
- **Variable Elimination**

Approximate Inference Methods

- **Stochastic Simulation**
- **Variational Methods**
- **Sampling Techniques**

16 Sampling Techniques

Sampling is used to approximate distributions by drawing random values from them.

Cumulative Probability Distribution

- Example: Weather distribution

$$
\begin{aligned}
P(X = v_1) &= 0.3 \quad \text{(Sunny)} \\
P(X = v_2) &= 0.4 \quad \text{(Rainy)} \\
P(X = v_3) &= 0.1 \quad \text{(Cloudy)} \\
P(X = v_4) &= 0.2 \quad \text{(Foggy)}
\end{aligned}
$$

Steps:

1. Order the values of the domain X
2. Generate the cumulative distribution
3. Draw a random number $y \sim \text{Uniform}(0, 1)$
4. Select the value of X based on where y falls in the cumulative distribution

Rejection Sampling

- Generate full samples.
- Reject samples that do not match the evidence.
- Inefficient if evidence is rare.

Likelihood Weighting

- Fix the evidence variables.
- Sample other variables conditioned on evidence.
- Weight each sample by its likelihood.

Importance Sampling

- Draw samples from an easier distribution.
- Weight the samples to correct the bias.

17 Uncertainty Over Time: Markov Models

Markov Assumption

The current state depends only on a finite number of previous states.

Markov Chain

A sequence of random variables $X_1, X_2, \ldots, X_n$ satisfying:

$$P(X_{t+1} \mid X_t, X_{t-1}, \ldots, X_1) = P(X_{t+1} \mid X_t)$$

Key Assumption: The future is independent of the past given the present.

Transition Model

$$P(X_{t+1} \mid X_t)$$

$$X_1 \longrightarrow X_2 \longrightarrow X_3 \longrightarrow \cdots$$

18 Hidden Markov Models (HMMs)

Definition

A Markov model with hidden (unobservable) states.

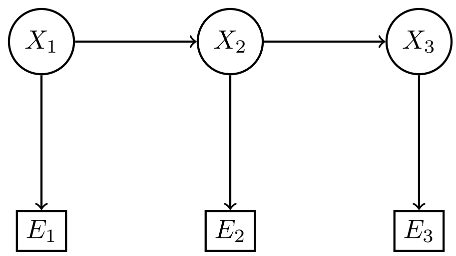
- **Transition Probability:** $P(X_{t+1} \mid X_t)$
- **Emission (Sensor) Probability:** $P(E_t \mid X_t)$

Sensor Markov Assumption

Current observation depends only on the current state.

Tasks in HMMs

- **Filtering:** Estimate current state given past observations
- **Prediction:** Estimate future states
- **Smoothing:** Estimate past states
- **Most Likely Explanation:** Find the most likely sequence of hidden states



19 Language Models

Simple Models: Bag of Words and Unigrams

- Words are treated as independent.
- $P(w_i)$: Probability distribution over words.
- Frequency vector indicates word usage.

Bigram Models

$$P(w_i \mid w_{i-1})$$

- The domain of each variable is the vocabulary.
- Models word sequences using adjacent word pairs.

Trigram Models

$$P(w_i \mid w_{i-1}, w_{i-2})$$

N-gram Models

- Extend bigrams/trigrams to any n-length sequence.
- Capture local dependencies in language.

Limitations of N-gram Models

- **Sparsity:** Many possible sequences are never seen in training data.
- **Data Size:** Large corpora are needed to get reliable estimates.
- **Lack of Semantic Understanding:** Words are treated as symbols, ignoring meaning and context.
- **Lack of Global Information:** Limited context prevents understanding long-distance dependencies.