REASONING UNDER UNCERTAINTY

Universidad Carlos III de Madrid

AI

Outline

Outline

Uncertainty in AI

Uncertainty

- In classic logic we express absolute certainty: true or false
  - *Ann works for Google* $\rightsquigarrow$ *works(Ann,Google)*
  - Production Systems (PS) use assertions considered as true

Uncertainty in AI

Uncertainty

- In classic logic we express absolute certainty: true or false
  - *Ann works for Google* $\rightsquigarrow$ *works(Ann,Google)*
  - Production Systems (PS) use assertions considered as true
- Uncertainty over some assertion
  - *weather tomorrow, traffic today, vehicle position*
- Simple solutions:

Uncertainty in AI

Uncertainty

- In classic logic we express absolute certainty: true or false
    - *Ann works for Google* $\rightsquigarrow$ *works(Ann,Google)*
    - Production Systems (PS) use assertions considered as true
- Uncertainty over some assertion
    - *weather tomorrow, traffic today, vehicle position*
- Simple solutions:
    - omit it
    - represent different results as a disjunction
        - assert *(good weather OR bad weather)*
        - we would need new algorithms
        - *(good weather OR bad weather)*
          if good weather then play tennis $\rightsquigarrow$???
- Problem: utility of the PS decreases

Uncertainty

Handling uncertainty

- Probability:
  - example: there is a 90% of chance that tomorrow rains
  - solution: probabilitistic reasoning, bayesian networks
- Vagueness:
  - example: X is tall/short, it is cold
  - solution: fuzzy logic

Reasoning under uncertainty

Short history

- Bayes theorem (1763)
- Fuzzy logic (Zadeh, 1965)
- Certainty factors in expert systems: MYCIN (1976), PROSPECTOR (1979)
- Dempster-Schafer (1976)
- Bayesian networks (Pearl, 1986)
- Sequences of decisions: MDP, HMM, POMDP

Outline

Outline

Probabilities

- Probability can be interpreted as a measure of:
  - proportion of times something is true
    - *20 students passed the exam out of 22*
    - a physical phenomenon
    - can be experimentally measured

Probabilities

- Probability can be interpreted as a measure of:
  - proportion of times something is true
    - *20 students passed the exam out of 22*
    - a physical phenomenon
    - can be experimentally measured
  - degree of belief over something
    - *I think Real Madrid will win the Liga with a 80%*
    - can vary over people
    - or intelligent system

Probabilities

- Probability can be interpreted as a measure of:
  - proportion of times something is true
    - *20 students passed the exam out of 22*
    - a physical phenomenon
    - can be experimentally measured
  - degree of belief over something
    - *I think Real Madrid will win the Liga with a 80%*
    - can vary over people
    - or intelligent system
- Probability calculus does not depend on the interpretation
  - Probabilities range in [0.0..1.0]
  - Probability=0: false
  - Probability=1: true

Random variables

Propositional logic

- we describe states as sets of boolean variables: p, q, r
- an interpretation is a truth assignment to those variables: $p = V, q = F, r = V$

Random variables

Propositional logic

- we describe states as sets of boolean variables: p, q, r
- an interpretation is a truth assignment to those variables: $p = V, q = F, r = V$

Probability theory

- we use a set of random variables that can take values on a given domain:
  - one die: $X \in \{1, 2, 3, 4, 5, 6\}$
  - two dice: $X \in \{1, 2, 3, 4, 5, 6\}$, $Y \in \{1, 2, 3, 4, 5, 6\}$

Random variables

Propositional logic

- we describe states as sets of boolean variables: p, q, r
- an interpretation is a truth assignment to those variables: $p = V, q = F, r = V$

Probability theory

- we use a set of random variables that can take values on a given domain:
  - one die: $X \in \{1, 2, 3, 4, 5, 6\}$
  - two dice: $X \in \{1, 2, 3, 4, 5, 6\}, Y \in \{1, 2, 3, 4, 5, 6\}$
- the associated value to a random variable is unknown
- we can assign a probability to each value
  - die: $P(X = 1) = \frac{1}{6}, \ldots, P(X = 6) = \frac{1}{6}$
- these probabilities define a probability distribution

Example

- Given a robot in a 100×100 grid with a given orientation
- Define its random variables and domains

Example

- Given a robot in a 100×100 grid with a given orientation
- Define its random variables and domains
  - Random variables

$$X \in \{0, \ldots, 99\}, Y \in \{0, \ldots, 99\}, \theta \in \{0, \ldots, 359\}$$

  - what would be the random variables of two robots?

Events

- Probability is defined for a sample space Ω, domain of a random variable
  - X: unknown position
  - Ω: values $0, \ldots, 99$

Events

- Probability is defined for a sample space Ω, domain of a random variable
  - X: unknown position
  - Ω: values $0, \ldots, 99$
- An atomic event $e \in \Omega$ is a possible situation (world)
  - $e_0 : X = 0, e_1 : X = 1, \ldots e_{99} : X = 99$

Events

- Probability is defined for a sample space Ω, domain of a random variable
  - X: unknown position
  - Ω: values $0, \ldots, 99$
- An atomic event $e \in \Omega$ is a possible situation (world)
  - $e_0 : X = 0, e_1 : X = 1, \ldots e_{99} : X = 99$
- A probability distribution assigns a number $P(e)$ to each $e \in \Omega$, such that
  - $0 \leq P(e) \leq 1$ and $\sum_{e \in \Omega} P(e) = 1$
  - $P(e_0) = P(e_1) = \ldots = P(e_{99}) = \frac{1}{100}$

Events

- Probability is defined for a sample space Ω, domain of a random variable
  - X: unknown position
  - Ω: values $0, \ldots, 99$
- An atomic event $e \in \Omega$ is a possible situation (world)
  - $e_0 : X = 0, e_1 : X = 1, \ldots e_{99} : X = 99$
- A probability distribution assigns a number $P(e)$ to each $e \in \Omega$, such that
  - $0 \leq P(e) \leq 1 \quad$ and $\quad \sum_{e \in \Omega} P(e) = 1$
  - $P(e_0) = P(e_1) = \ldots = P(e_{99}) = \frac{1}{100}$
- An event A is a set of atomic events: $A \subseteq \Omega$
  - $P(A) = \sum_{e \in A} P(e)$
  - $P($ robot is on one of the first four columns$) =$ $P(e_0) + P(e_1) + P(e_2) + P(e_3) = \frac{4}{100} = \frac{1}{25}$

Events

- Probability is defined for a sample space Ω, domain of a random variable
  - X: unknown position
  - Ω: values $0, \ldots, 99$
- An atomic event $e \in \Omega$ is a possible situation (world)
  - $e_0 : X = 0, e_1 : X = 1, \ldots e_{99} : X = 99$
- A probability distribution assigns a number $P(e)$ to each $e \in \Omega$, such that
  - $0 \leq P(e) \leq 1$ and $\sum_{e \in \Omega} P(e) = 1$
  - $P(e_0) = P(e_1) = \ldots = P(e_{99}) = \frac{1}{100}$
- An event A is a set of atomic events: $A \subseteq \Omega$
  - $P(A) = \sum_{e \in A} P(e)$
  - $P($ robot is on one of the first four columns$) =$ $P(e_0) + P(e_1) + P(e_2) + P(e_3) = \frac{4}{100} = \frac{1}{25}$
  - what is the probability that the robot is in the first row?

“A priori” probability

- The probability distribution of a random variable is usually represented as a $\langle$ vector $\rangle$
- Example:
  - $P(X) = \langle P(X=0), \ldots, P(X=99) \rangle$
- The joint probability distribution is the distribution for several variables. E.g. $P(X, Y)$, $P(X, Y, \theta)$

$$
\begin{aligned}
P(X, Y) &= \langle P(X=0, Y=0), P(X=0, Y=1), \ldots, P(X=99, Y=99) \rangle \\
&= \langle P(X=0, Y=0), P(X=0, Y=1), \ldots, P(X=99, Y=99) \rangle \\
&= \langle \frac{1}{10000}, \frac{1}{10000}, \ldots, \frac{1}{10000} \rangle
\end{aligned}
$$

- This distribution is “a priori” or unconditional, since it does not depend on any condition

Example

- Given a robot in a 100×100 grid with a given orientation
- Define its random variables, domains, and probability distribution

Example

- Given a robot in a 100×100 grid with a given orientation
- Define its random variables, domains, and probability distribution
  - Random variables

$$X \in \{0, \ldots, 99\}, Y \in \{0, \ldots, 99\}, \theta \in \{0, \ldots, 359\}$$

  - Probability distribution (*unknown position*):

$$P(X = 0, Y = 0) = P(X = 0, Y = 1) = \ldots =$$
$$P(X = 0, Y = 99) = P(X = 1, Y = 0) = \ldots =$$
$$P(X = 99, Y = 99) = \frac{1}{100 \times 100}$$

Example

- Given a robot in a 100×100 grid with a given orientation
- Define its random variables, domains, and probability distribution
  - Random variables

$$X \in \{0, \dots, 99\}, Y \in \{0, \dots, 99\}, \theta \in \{0, \dots, 359\}$$

  - Probability distribution (*unknown position*):

$$P(X = 0, Y = 0) = P(X = 0, Y = 1) = \dots =$$
$$P(X = 0, Y = 99) = P(X = 1, Y = 0) = \dots =$$
$$P(X = 99, Y = 99) = \frac{1}{100 \times 100}$$

  - Probability distribution (*unknown position and orientation*):

Example

- Given a robot in a 100×100 grid with a given orientation
- Define its random variables, domains, and probability distribution
  - Random variables

$$X \in \{0, \ldots, 99\}, Y \in \{0, \ldots, 99\}, \theta \in \{0, \ldots, 359\}$$

  - Probability distribution (*unknown position*):

$$P(X = 0, Y = 0) = P(X = 0, Y = 1) = \ldots =$$
$$P(X = 0, Y = 99) = P(X = 1, Y = 0) = \ldots =$$
$$P(X = 99, Y = 99) = \frac{1}{100 \times 100}$$

  - Probability distribution (*unknown position and orientation*):

$$P(X = 0, Y = 0, \theta = 0) = \ldots$$
$$P(X = 100, Y = 100, \theta = 360) = \frac{1}{100 \times 100 \times 360}$$

Example

- Given a robot in a 100×100 grid with a given orientation
- Define its random variables, domains, and probability distribution
  - Random variables

$$X \in \{0, \dots, 99\}, Y \in \{0, \dots, 99\}, \theta \in \{0, \dots, 359\}$$

  - Probability distribution (*unknown position*):

$$P(X = 0, Y = 0) = P(X = 0, Y = 1) = \dots =$$
$$P(X = 0, Y = 99) = P(X = 1, Y = 0) = \dots =$$
$$P(X = 99, Y = 99) = \frac{1}{100 \times 100}$$

  - Probability distribution (*unknown position and orientation*):

$$P(X = 0, Y = 0, \theta = 0) = \dots$$
$$P(X = 100, Y = 100, \theta = 360) = \frac{1}{100 \times 100 \times 360}$$

- what would be the probability distribution of unknown position and orientation of two robots?

Conditional probability

$$P(A|B) = \frac{P(A, B)}{P(B)} \text{ if } P(B) \neq 0$$

- $P(A|B)$ can be interpreted as the updated probability of A, once B has been observed
- Examples
    - $P(X = 0)$?
    - $P(X = 0|X < 10)$?
    - $P(X = 0|Y = 0)$?

Conditional distribution

- Conditional distribution is a vector of vectors (array)
- Example:

$$
\begin{aligned}
P(X|Y) &= \langle P(X|Y=0), \ldots, P(X|Y=99) \rangle \\
&= \langle \langle \frac{1}{10000}, \ldots, \frac{1}{10000} \rangle, \ \ldots, \ \langle \frac{1}{10000}, \ldots, \frac{1}{10000} \rangle \rangle
\end{aligned}
$$

Law of total probability

- Given a set of pairwise disjoint events A_i such that their union is the whole sample space and another event B:

$$P(B) = \sum_{i=1}^{n} P(B, A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Law of total probability

- Given a set of pairwise disjoint events A_i such that their union is the whole sample space and another event B:

$$P(B) = \sum_{i=1}^{n} P(B, A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

- Thus, if we have a random variable A with possible disjoint values $a_1, \ldots, a_n$ and an event B:

$$P(B) = \sum_{i=1}^{n} P(B, A = a_i) = \sum_{i=1}^{n} P(B|A = a_i)P(A = a_i)$$

Law of total probability

- Given a set of pairwise disjoint events A_i such that their union is the whole sample space and another event B:

$$P(B) = \sum_{i=1}^{n} P(B, A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

- Thus, if we have a random variable A with possible disjoint values $a_1, \dots, a_n$ and an event B:

$$P(B) = \sum_{i=1}^{n} P(B, A = a_i) = \sum_{i=1}^{n} P(B|A = a_i)P(A = a_i)$$

- Example

$$\begin{aligned} P(X = 0) &= \sum_{i=0}^{99} P(X = 0, Y = i) \\ &= P(X = 0, Y = 0) + P(X = 0, Y = 1) + \dots + P(X = 0, Y = 99) \\ &= \frac{1}{10000} + \frac{1}{10000} + \dots + \frac{1}{10000} = \frac{100}{10000} = \frac{1}{100} \end{aligned}$$

Conditional probability

- Conditional probability

$$P(A|B) = \frac{P(A,B)}{P(B)} \text{ if } P(B) \neq 0$$

- Product rule

$$P(A,B) = P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$$

- Bayes rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \alpha P(B|A)P(A)$$

Conditional probability

- Conditional probability

$$P(A|B) = \frac{P(A,B)}{P(B)} \text{ if } P(B) \neq 0$$

- Product rule

$$P(A,B) = P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$$

- Bayes rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \alpha P(B|A)P(A)$$

- Sometimes obtaining $P(A|B)$ is easier than $P(B|A)$

$$P(\text{Cause}|\text{Effect}) = \frac{P(\text{Effect}|\text{Cause})P(\text{Cause})}{P(\text{Effect})}$$

  it is usually easier to ask an expert $P(\text{Effect}|\text{Cause})$ than $P(\text{Cause}|\text{Effect})$

Chain rule

- Product rule $P(X_1, X_2) = P(X_1)P(X_2|X_1)$
- Recursive application: chain rule

$$P(X_1, \ldots, X_n) = P(X_1, \ldots, X_{n-1})P(X_n|X_1, \ldots, X_{n-1})$$

$$\begin{aligned}
&= P(X_1, \ldots, X_{n-2})P(X_{n-1}|X_1, \ldots, X_{n-2})P(X_n|X_1, \ldots, X_{n-1}) \\
&= \ldots \\
&= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \ldots P(X_n|X_1, \ldots, X_{n-1}) \\
&= P(X_1) \times \prod_{i=2}^{n} P(X_i|X_1, \ldots, X_{i-1})
\end{aligned}$$

- Crucial for AI: we can compute joint probabilities only using conditional probabilities

Outline

Use of probabilities in AI

Typical tasks: decision making, classification, prediction, ...

- What is true?
  - use of classic logic: propositional satisfaction, circuit verification, production systems, shortest path, chess, ...

Use of probabilities in AI

Typical tasks: decision making, classification, prediction, ...

- What is true?
  - use of classic logic: propositional satisfaction, circuit verification, production systems, shortest path, chess, ...
- vs. what is more probable?
  - use of probabilities: bayesian networks, sequences prediction (speech recognition), classification (of language), weather forecast, videogames, ...

Use of probabilities in AI

Typical tasks: decision making, classification, prediction, ...

- What is true?
  - use of classic logic: propositional satisfaction, circuit verification, production systems, shortest path, chess, ...
- vs. what is more probable?
  - use of probabilities: bayesian networks, sequences prediction (speech recognition), classification (of language), weather forecast, videogames, ...
- What if the selected model is wrong?
  - In classic logic
    - incomplete model $\rightarrow$ ok
    - incorrect model $\rightarrow$ problem
  - In probabilities
    - it is usually more interesting to know the relation among probabilities than the exact numbers: $P(e) > P(e')$?
    - it could be more robust

Inference

- Main task:

  Compute probabilities of events e given some evidence o:
  $$P(e|o)$$

- Inference tasks:
  - compute posterior distribution given evidence
  - choose an action to achieve high reward given some evidence
  - decision making with optimal utility
  - classification
  - diagnosis

Inference

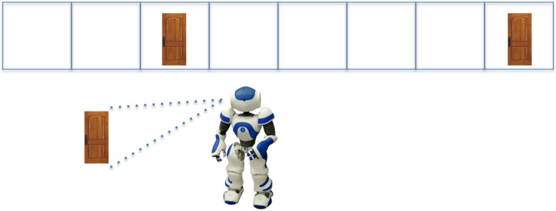
Compute posterior distribution given evidence $P(X|o)$

1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

$$P(X) = \langle \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8} \rangle$$

Inference

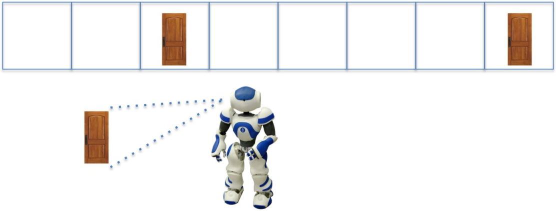
Compute posterior distribution given evidence $P(X|o)$



What is the probability distribution of the position of the robot (X) given that it observed a door ($o =$door), $P(X|o = \text{door})$?

Inference

Compute posterior distribution given evidence $P(X|o)$

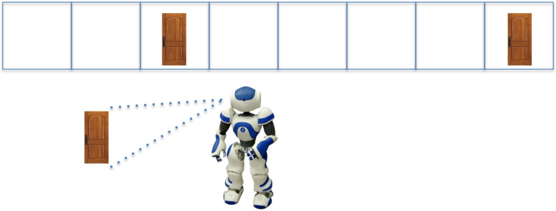


What is the probability distribution of the position of the robot (X) given that it observed a door ($o =$door), $P(X|o = \text{door})$?

$$P(X|o = \text{door}) = \langle 0, 0, \frac{1}{2}, 0, 0, 0, 0, 0, \frac{1}{2} \rangle$$

Inference

Compute posterior distribution given evidence $P(X|o)$



What is the probability distribution of the position of the robot (X) given that it observed a door ($o =$door), $P(X|o = \text{door})$?

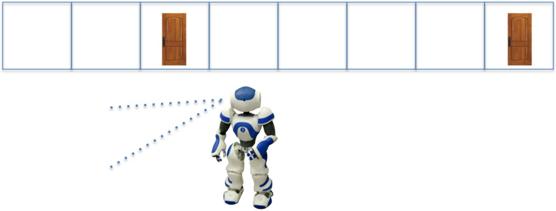
$$P(X|o = \text{door}) = \langle 0, 0, \frac{1}{2}, 0, 0, 0, 0, 0, \frac{1}{2} \rangle$$

$$P(X = 0|o = \text{door}) = \frac{P(\text{door}|X = 0)P(X = 0)}{P(\text{door})} = \frac{0 \times \frac{1}{8}}{\frac{2}{8}} = 0$$

$$P(X = 2|o = \text{door}) = \frac{P(\text{door}|X = 2)P(X = 2)}{P(\text{door})} = \frac{1 \times \frac{1}{8}}{\frac{2}{8}} = \frac{1}{2}$$

Inference

Compute posterior distribution given evidence $P(X|o)$

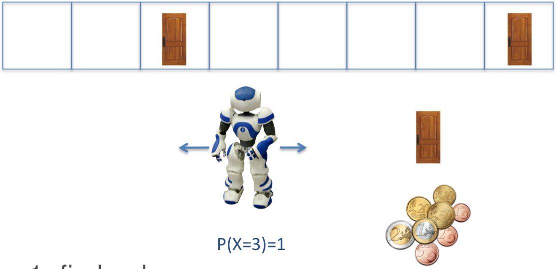


What is the probability distribution of the position of the robot (X) given that it did not observe a door ($o =$not-door), $P(X|o = \text{not-door})$?

Inference

Choose an action to achieve high reward ($R = 1$) given that variable X has a value

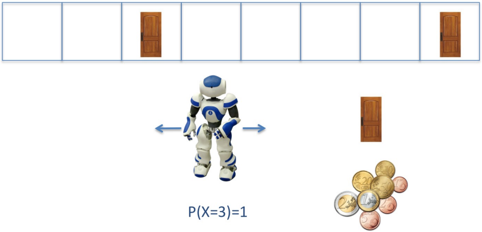
- if I can take Action=1 or Action=2, I choose Action=1 if $P(R = 1|\text{Action} = 1, X = v) > P(R = 1|\text{Action} = 2, X = v)$
- example:



- $R = 1$: find a door
- Actions: *Left* or *Right*

Inference

Choose an action to achieve high reward ($R = 1$) given that variable X has a value



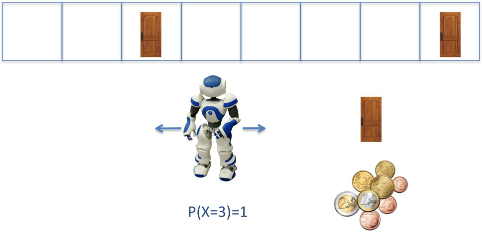
$$P(R = 1|\text{Action=Left}, X = 3) = 1$$

$$P(R = 1|\text{Action=Right}, X = 3) = 0$$

Then, select Left

Inference

Choose an action to achieve high reward ($R = 1$) given that variable X has a value



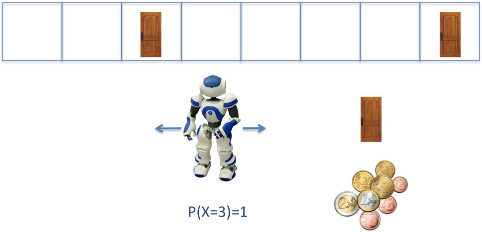
$$P(R = 1 | \text{Action=Left}, X = 3) = 1$$

$$P(R = 1 | \text{Action=Right}, X = 3) = 0$$

Then, select Left
Similar task: Choose an action to achieve high reward ($R = 1$) given some evidence o

Inference

Choose an action to achieve high reward ($R = 1$) given that variable X has a value



$$P(R = 1|\text{Action=Left}, X = 3) = 1$$

$$P(R = 1|\text{Action=Right}, X = 3) = 0$$

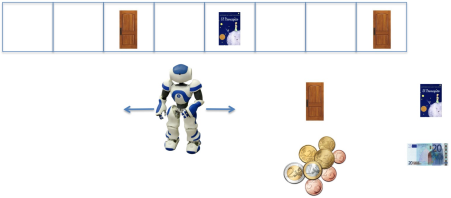
Then, select Left
Similar task: Choose an action to achieve high reward ($R = 1$) given some evidence o
What would happen if $X = 1$? And if $X = 7$?

Inference

Decisions with optimal utility

- Choose action a_j with best expected utility

$$
\begin{aligned}
&\text{ExpectedUtility}(\text{Action} = a_j) \\
&= \sum_i P(\text{Output} = i \mid \text{Action} = a_j, \text{evidence}) \times \text{ExpectedUtility}(\text{Output} = i)
\end{aligned}
$$

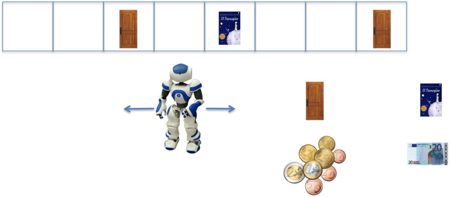


Inference

Decisions with optimal utility

- Choose action a_j with best expected utility

$$
\begin{aligned}
&\text{ExpectedUtility}(\text{Action} = a_j)\\
&= \sum_i P(\text{Output} = i \mid \text{Action} = a_j, \text{evidence}) \times \text{ExpectedUtility}(\text{Output} = i)
\end{aligned}
$$



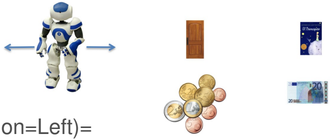
- Output: see door or book
- Action: Left or Right
- Evidence: robot position
- Utility: see door=3.88, see book=20

Inference

Decisions with optimal utility

- Choose action a_j with best expected utility

$$\text{ExpectedUtility}(\text{Action} = a_j)$$
$$= \sum_i P(\text{Output} = i | \text{Action} = a_j, \text{evidence}) \times \text{ExpectedUtility}(\text{Output} = i)$$



ExpectedUtility(Action=Left)=
$\sum_i P(\text{Output} = i|\text{Action=Left}, X) \times \text{ExpectedUtility}(\text{Output} = i) =$
$P(\text{Output=door}|\text{Action=Left}, X = 3) \times \text{ExpectedUtility}(\text{Output=door}) +$
$P(\text{Output=book}|\text{Action=Left}, X = 3) \times \text{ExpectedUtility}(\text{Output=book}) =$
$1 \times 3.88 + 0 \times 20 = 3.88$

Inference

Decisions with optimal utility

- Choose action a_j with best expected utility

$$\begin{aligned}
&\text{ExpectedUtility}(\text{Action} = a_j) \\
&= \sum_i P(\text{Output} = i | \text{Action} = a_j, \text{evidence}) \times \text{ExpectedUtility}(\text{Output} = i)
\end{aligned}$$

ExpectedUtility(Action=Left)=3.88
ExpectedUtility(Action=Right)=
$\sum_i P(\text{Output} = i | \text{Action=Right}, X) \times \text{ExpectedUtility}(\text{Output} = i) =$
$P(\text{Output=door} | \text{Action=Right}, X = 3) \times \text{ExpectedUtility}(\text{Output=door}) +$
$P(\text{Output=book} | \text{Action=Right}, X = 3) \times \text{ExpectedUtility}(\text{Output=book}) =$
$0 \times 3.88 + 1 \times 20 = 20$

Inference

- Classification
  - given some observations o to what class they belong?
  - compare $P(\text{Class} = 1|o)$ vs. $P(\text{Class} = 2|o)$
  - examples:
    - given some client data (average money on bank, location of house, monthly earnings),
      determine whether s/he will correctly return a mortgage (Class=1) or not (Class=2)
    - given some image (number of pixels of a given luminosity, number of lines),
      determine whether it belongs to a cat (Class=cat), a dog (Class=dog), or something different (Class=other)
  - computation (Naïve Bayes)
    - $\text{Class}=\arg\max_{c\in\text{Classes}} P(\text{Class}=c|o)$
    - $P(\text{Class} = c|o) = \frac{P(o|\text{Class}=c) \times P(\text{Class}=c)}{P(o)}$

Inference

- Classification
  - given some observations o to what class they belong?
  - compare $P(\text{Class} = 1|o)$ vs. $P(\text{Class} = 2|o)$
  - examples:
    - given some client data (average money on bank, location of house, monthly earnings),
      determine whether s/he will correctly return a mortgage (Class=1) or not (Class=2)
    - given some image (number of pixels of a given luminosity, number of lines),
      determine whether it belongs to a cat (Class=cat), a dog (Class=dog), or something different (Class=other)
  - computation (Naïve Bayes)
    - $\text{Class}{=}\arg\max_{c \in \text{Classes}} P(\text{Class}{=}c|o)$
    - $P(\text{Class} = c|o) = \frac{P(o|\text{Class}{=}c) \times P(\text{Class}{=}c)}{P(o)}$
- Diagnosis
  - probability of an illness $I = 1$ given results of analysis o, $P(I = 1|o)$

Independence

- A and B are independent iff $P(A, B) = P(A)P(B)$
- In that case, it also holds: $P(A|B) = P(A)$

Independence

- A and B are independent iff $P(A, B) = P(A)P(B)$
- In that case, it also holds: $P(A|B) = P(A)$
- $P(RobotX, Orientation) = P(RobotX)P(Orientation)$
- Reduction in the size of the probability distribution:

Independence

- A and B are independent iff $P(A, B) = P(A)P(B)$
- In that case, it also holds: $P(A|B) = P(A)$
- $P(RobotX, Orientation) = P(RobotX)P(Orientation)$
- Reduction in the size of the probability distribution:

$$100 \times 360 = 36000 \rightsquigarrow 100 + 360 = 460$$

- Smaller description implies
  - more efficient algorithms
  - less data (probabilities) to be specified

Summary

- Discrete probabilities can be understood as assigning weights to all possible assignments of a set of variables
- Probability is a rigorous formalism to handle uncertainty
- Joint probabilities specify the probability of each atomic event
- Questions can be answered summing atomic events
- For some interesting problems we can reduce the number of parameters to represent the joint probability
- We need tools for expressive representations that also allow us efficient reasoning
- Independence will be crucial

Credits

Material based on:

- Previous material of UC3M
- Book and teaching notes of *Artificial Intelligence: A Modern Approach*. Russell&Novig. 2nd edition
- Teaching material by Héctor Geffner

Outline

Outline

Outline

References

Lotfi A. Zadeh.
Outline of a new approach to the analysis of complex systems and decision processes.
Transactions on Systems, Man and Cybernetics, 1(1), 1973.