COMP3411/9414 Artificial Intelligence
Session 1, 2019

Tutorial Solutions - Week 9

This page was last updated: 04/30/2019 06:07:07

Activity 10.1 Conditional Probability

Only 4% of the population are color blind, but 7% of men are color blind. What percentage of color blind people are men?

> If we assume 50% of the population are men, then the fraction of "color blind men" is 0.5 * 0.07 = 0.035
> This means the fraction of "color blind women" is 0.04 - 0.035 = 0.005
> Therefore, the fraction of color blind people who are men is 0.035 / 0.04 = 87.5%.

Activity 10.2 Enumerating Probabilities

(Exercise 13.8 from Russell & Norvig)

	toothache		**¬toothache**	
	catch	**¬catch**	**catch**	**¬catch**
cavity	.108	.012	.072	.008
¬cavity	.016	.064	.144	.576

1. Given the full joint distribution shown in Figure 13.3 (also on page 17 of the Uncertainty lecture slides), calculate the following:

   a. P(toothache ∧ ¬ catch) = 0.012 + 0.064 = 0.076

   b. P(catch) = 0.108 + 0.016 + 0.072 + 0.144 = 0.34

   c. P(cavity | catch) = P(cavity ∧ catch) / P(catch)
   = (0.108 + 0.072)/(0.108 + 0.072 + 0.016 + 0.144) = 0.18 / 0.34 = 0.53

d. $P(\text{cavity} \mid \text{toothache} \lor \text{catch}) =$
$P(\text{cavity} \land (\text{toothache} \lor \text{catch})) / P(\text{toothache} \lor \text{catch})$
$= (0.108 + 0.012 + 0.072)/(0.108 + 0.012 + 0.072 + 0.016 + 0.064 + 0.144)$
$= 0.192 / 0.416 = 0.46$

2. Verify the conditional independence claimed in the lecture slides by showing that $P(\text{catch} \mid \text{toothache}, \text{cavity}) = P(\text{catch} \mid \text{cavity})$

   $P(\text{catch} \mid \text{toothache} \land \text{cavity}) =$
   $P(\text{catch} \land \text{toothach} \land \text{cavity}) / P(\text{toothache} \land \text{cavity})$
   $= 0.108 /(0.108 + 0.012) = 0.108 / 0.12 = 0.9$

   $P(\text{catch} \mid \text{cavity}) = P(\text{catch} \land \text{cavity}) / P(\text{cavity})$
   $= (0.108 + 0.072)/(0.108 + 0.012 + 0.072 + 0.008) = 0.18 / 0.2 = 0.9$

Activity 10.3 Wumpus World with Three Pits

Consider the same Wumpus World situation shown above, but this time making a different prior assumption about the placement of the Pits - namely, that exactly three Pits have been placed randomly among the sixteen squares at the beginning of the game. Using this new prior, calculate:

1. the probability of a Pit in (1,3)
2. the probability of a Pit in (2,2)

We start with these equations from the Lecture Notes:

$$P(Pit_{1,3} \mid known) = \Sigma_{fringe \in F}\, P(Pit_{1,3} \mid fringe)\, \Sigma_{other}\, P(known \mid fringe,other)P(fringe,other) \,/\, P(known)$$

$$P(known) = \Sigma_{fringe \in F}\, \Sigma_{other}\, P(known \mid fringe,other)\, P(fringe,other)$$

In this case, $P(known \mid fringe,other) = 1$ if and only if $fringe \in F$ and *fringe, other* between them contain exactly three Pits. For these configurations, $P(fringe,other)$ becomes a constant that can be cancelled from the numerator and denominator. If there are two pits in the fringe, the other pit can be placed in 10 different squares. If there is only one pit in the fringe, the other two pits can be placed in 45 different ways. Therefore,

$P(Pit_{1,3} \mid known) = (1+10+10)/(1+10+10+10+45) = 21/76 \approx 0.276$

$P(Pit_{2,2} \mid known) = (1+10+10+45)/(1+10+10+10+45) = 66/76 \approx 0.868$

Activity 9.4 Q-Learning

Consider a world with two states $S = \{S_1, S_2\}$ and two actions $A = \{a_1, a_2\}$, where the transitions δ and reward r for each state and action are as follows:

$\delta(S_1, a_1) = S_1,\ r(S_1, a_1) = 0$
$\delta(S_1, a_2) = S_2,\ r(S_1, a_2) = -1$
$\delta(S_2, a_1) = S_2,\ r(S_2, a_1) = +1$
$\delta(S_2, a_2) = S_1,\ r(S_2, a_2) = +5$

1. Draw a picture of this world, using circles for the states and arrows for the transitions.

2. Assuming a discount factor of $\gamma = 0.9$, determine:
   a. the optimal policy $\pi^* : S \rightarrow A$

      $\pi^*(S_1) = a_2$
      $\pi^*(S_2) = a_2$

   b. the value function $V^* : S \rightarrow R$

      $V(S_1) = -1 + \gamma V(S_2)$
      $V(S_2) = 5 + \gamma V(S_1)$

      So $V(S_1) = -1 + 5\gamma + \gamma^2 V(S_1)$
      i.e. $V(S_1) = (-1 + 5\gamma) / (1 - \gamma^2) = 3.5 / 0.19 = 18.42$

$V(S_2) = 5 + \gamma V(S_1) = 5 + 0.9 * 3.5 / 0.19 = 21.58$

c. the "Q" function $Q : S \times A \rightarrow R$

$Q(S_1, a_1) = \quad \gamma V(S_1) = 16.58$

$Q(S_1, a_2) = \quad V(S_1) = 18.42$

$Q(S_2, a_1) = 1 + \gamma V(S_2) = 20.42$

$Q(S_2, a_2) = \quad V(S_2) = 21.58$

3. Write the Q values in a matrix:

Q	**a_1**	**a_2**
S_1	16.58	18.42
S_2	20.42	21.58

4. Trace through the first few steps of the Q-learning algorithm, with all Q values initially set to zero. Explain why it is necessary to force exploration through probabilistic choice of actions, in order to ensure convergence to the true Q values.

current state	**chosen action**	**new Q value**
S_1	a_1	$0 + \gamma*0 = 0$
S_1	a_2	$-1 + \gamma*0 = -1$
S_2	a_1	$1 + \gamma*0 = +1$

At this point, the table looks like this:

Q	**a_1**	**a_2**
S_1	0	-1
S_2	1	0

If we do not force exploration, the agent will always prefer action a_1 in state S_2, so it will never explore action a_2. This means that $Q(S_2, a_2)$ will remain zero forever, instead of converging to the true value of 21.58 . If we force exploration, the next few steps might look like this:

current state	chosen action	new Q value
S_2	a_2	$5 + \gamma*0 = 5$
S_1	a_1	$0 + \gamma*0 = 0$
S_1	a_2	$-1 + \gamma*5 = 3.5$
S_2	a_1	$1 + \gamma*5 = 5.5$
S_2	a_2	$5 + \gamma*3.5 = 8.15$

Now we have this table:

Q	$\mathbf{a_1}$	$\mathbf{a_2}$
$\mathbf{S_1}$	0	3.5
$\mathbf{S_2}$	5.5	8.15

From this point on, the agent will prefer action a_2 both in state S_1 and in state S_2. Further steps refine the Q value estimates, and, in the limit, they will converge to their true values.

current state	chosen action	new Q value
S_1	a_1	$0 + \gamma*3.5 = 3.15$
S_1	a_2	$-1 + \gamma*8.15 = 6.335$
S_2	a_1	$1 + \gamma*8.15 = 8.335$
S_2	a_2	$5 + \gamma*6.34 = 10.70$

etc...