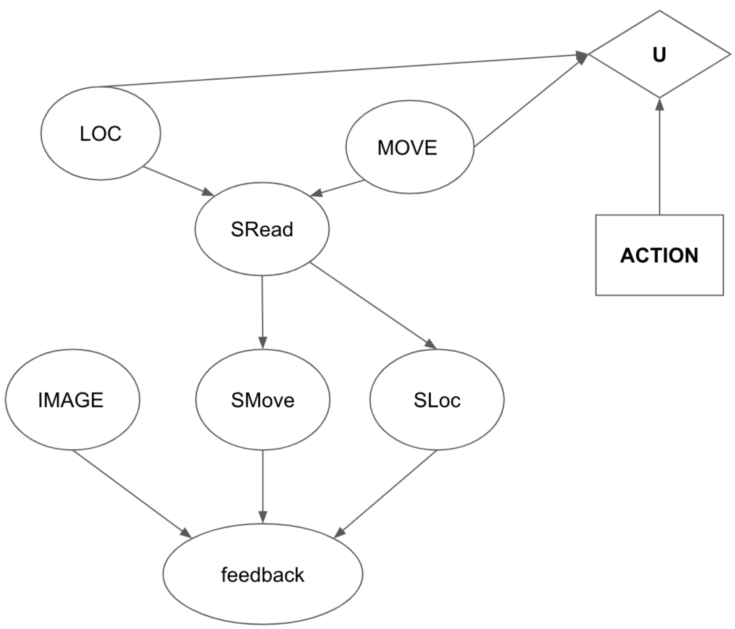
CS 188 Spring 2026 | Introduction to Artificial Intelligence | Exam Prep 9 Solutions

Q1. Vehicle Perception Indication

A vehicle is trying to identify the situation of the world around it using a set of sensors located around the vehicle.

Each sensor reading (SRead) is based off of an object's location (LOC) and an object's movement (MOVE). The sensor reading will then produce various values for its predicted location (SLoc) and predicted movement (SMove). The user will receive these readings, as well as the the image (IMAGE) as feedback.

(a) The vehicle takes an action, and we assign some utility to the action based on the object's location and movement. Possible actions are MOVE TOWARDS, MOVE AWAY, and STOP. Suppose the decision network faced by the vehicle is the following.



(i) Based on the diagram above, which of the following **could possibly be** true?

- [x] VPI (Image) = 0
- [ ] VPI (SRead) < 0
- [ ] VPI (SMove, SRead) > VPI (SRead)
- [x] VPI (Feedback) = 0
- ○ None of the above

VPI(Image) = 0 because there is not active path connecting Image and U

VPI cannot be negative, so option 2 is not selected.

VPI(SMove, SRead) = VPI(SMove | SRead) + VPI(SRead), therefore we can cancel VPI(SRead) from both side, and it becomes asking if VPI(SMove | SRead) > 0. And we can see that cannot be true, because shading in SRead, there is no active path connecting SMove and U.

There is an active path connecting Feedback and U, therefore VPI(Feedback) ≥ 0. It could still be 0 because active path only gives the possibility of > 0.

(ii) Based on the diagram above, which of the following **must necessarily be** true?

- [x] VPI (Image) = 0
- [ ] VPI (SRead) = 0

■ VPI (SMove, SRead) = VPI (SRead)

☐ VPI (Feedback) = 0

○ None of the above

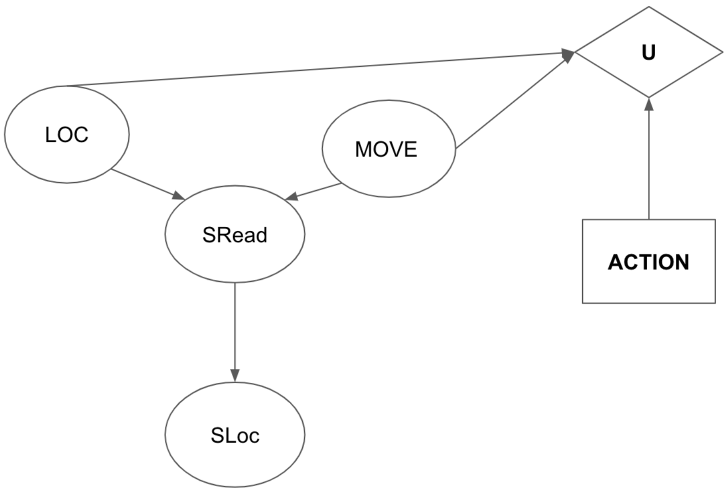
VPI(Image) = 0 because there is not active path connecting Image and U

VPI(SRead) could be > 0 because SRead-MOVE-U is an active path between SRead and U

VPI(SMove, SRead) = VPI(SMove | SRead) + VPI(SRead), therefore we can cancel VPI(SRead) from both side, and it becomes asking if VPI(SMove | SRead) == 0. And we can see that must true, because shading in SRead, there is no active path connecting SMove and U.

VPI(Feedback) could be > 0 because feedback-SLoc-SRead-MOVE-U is an active path

Let's assume that your startup has less money, so we use a simpler sensor network. One possible sensor network can be represented as follows.



You have distributions of $P(\text{LOC})$, $P(\text{MOVE})$, $P(SRead|\text{LOC}, \text{MOVE})$, $P(SLoc|SRead)$ and utility values $U(a, l, m)$.

(b) Complete the equation for determining the expected utility for some ACTION a.

$$EU(a) = \Big(\underline{\quad\textbf{(i)}\quad}\ \underline{\quad\textbf{(ii)}\quad}\ \underline{\quad\textbf{(iii)}\quad}\Big)\ U(a, l, m)$$

(i) ● $\sum_l P(l)$ ○ $\sum_{sloc} P(sloc|l)$ ○ $\sum_l \sum_{sloc} P(sloc|l)$ ○ 1

(ii) ● $\sum_m P(m)$ ○ $\sum_m P(sloc|m)$ ○ $\sum_l \sum_m \sum_{sloc} P(sloc|l)P(sloc|m)$ ○ 1

(iii) ○ $* \sum_l \sum_m \sum_{sloc} P(sloc|l)P(sloc|m)$ ○ $+ \sum_l \sum_m \sum_{sloc} P(sloc|l)P(sloc|m)$
○ $+ \sum_l \sum_m \sum_{sloc} P(sloc|l, m)P(l)P(m)$ ● $* 1$

$EU(a) = \sum_l P(l) \sum_m P(m)U(a, l, m)$

We can eliminate SRead and SLoc via marginalization, so they don't need to be included the expression

(c) Your colleague Bob invented a new sensor to observe values of $SLoc$.

(i) Suppose that your company had no sensors till this point. Which of the following expression is equivalent to VPI($SLoc$)?

■ $\text{VPI}(SLoc) = (\sum_{sloc} P(sloc)\,\text{MEU}(SLoc = sloc)) - \max_a \text{EU}(a)$

■ $\text{VPI}(SLoc) = \text{MEU}(SLoc) - \text{MEU}(\emptyset)$

□ $\text{VPI}(SLoc) = \max_{sloc} \text{MEU}(SLoc = sloc) - \text{MEU}(\emptyset)$

○ None of the above

Option 2 is correct by definition, and option 1 is the expanded version of option 2

(ii) Gaagle, an established company, wants to sell your startup a device that gives you $SRead$. Given that you already have Bob's device (that gives you $SLoc$), what is the maximum amount of money you should pay for Gaagle's device? Suppose you value $1 at 1 utility.

□ $\text{VPI}(SRead)$

■ $\text{VPI}(SRead) - \text{VPI}(SLoc)$

□ $\text{VPI}(SRead, SLoc)$

■ $\text{VPI}(SRead, SLoc) - \text{VPI}(SLoc)$

○ None of the above

Choice 4 is correct by definition

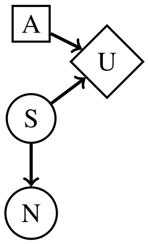
Choice 2 is true because VPI(SLoc | SRead) = 0, and thus

VPI(SRead) = VPI(SRead) + 0 = VPI(SRead) + VPI(SLoc | SRead) = VPI(SRead, SLoc), which makes choice 2 the same as choice 4

Q2. [Optional] Dressed to Impress

(a) Alice is invited to a party tonight which is said to be once-in-a-lifetime. However, this mysterious party doesn't publicize who is going and thus Alice has no idea whether the size S of the party will be be large ($S = +s$) or tiny ($S = -s$). The size can affect the noise N outside the party, and it will either be noisy ($N = +n$) or quiet ($N = -n$). Alice has three dresses: blue, red and yellow. Each dress will have a different utility for her depending on the size of the party. Let's help her decide which will be best!

We have the following decision network, where circles are chance nodes, squares are decision nodes, and diamonds are utility nodes:



S	$P(S)$
+s	0.5
-s	0.5

S	N	$P(N \mid S)$
+s	+n	0.7
+s	-n	0.3
-s	+n	0.1
-s	-n	0.9

S	A	U
+s	blue	80
-s	blue	60
+s	red	40
-s	red	100
+s	yellow	60
-s	yellow	40

(i) What is the expected utility of wearing each dress, with both S and N unobserved?

- EU(A=blue) = $80 * 0.5 + 60 * 0.5 = 70$
- EU(A=red) = $40 * 0.5 + 100 * 0.5 = 70$
- EU(A=yellow) = $60 * 0.5 + 40 * 0.5 = 50$

What is Alice's maximum expected utility?

- MEU({ }) = 70

(ii) Suppose Alice observes that the party is quiet, $N = -n$. Compute the following conditional probabilities with this observation:

- $P(+s|-n) = \frac{P(-n|+s)P(+s)}{P(-n|+s)P(+s)+P(-n|-s)P(-s)} = \frac{0.3 \cdot 0.5}{0.3 \cdot 0.5 + 0.9 \cdot 0.5} = 0.25$
- $P(-s|-n) = P(-s|-n) = 1 - P(+s|-n) = 0.75$

What is the expected utility of wearing each dress?

- EU(A=blue | $N = -n$) = $80 * 0.25 + 60 * 0.75 = \frac{145}{2} = 65$
- EU(A=red | $N = -n$) = $40 * 0.25 + 100 * 0.75 = \frac{155}{2} = 85$
- EU(A=yellow | $N = -n$) = $60 * 0.25 + 40 * 0.75 = 45$

What is Alice's maximum expected utility given that $N = -n$?

- MEU({N=-n}) = 85

(iii) Construct a formula for $VPI(N)$ for the given network. To decouple this problem from your work above, use any of the symbolic terms from the following list (rather than plugging in numeric values):

$P(+n|+s)$, $P(+n|-s)$, $P(-n|+s)$, $P(-n|-s)$, $P(+n)$, $P(-n)$, $P(+s)$, $P(-s)$, $MEU(\{\})$, $MEU(\{N = +n\})$, $MEU(\{N = -n\})$

- $VPI(N) = P(+n)MEU(\{N = +n\}) + P(-n)MEU(\{N = -n\}) - MEU(\{\})$

3 Decision Trees

You are a geek who hates sports. Trying to look cool at a party, you join a discussion that you believe to be about football and basketball. You gather information about the two main subjects of discussion, but still cannot figure out what sports they play.

Sport	**Position**	**Name**	**Height**	**Weight**	**Age**	**College**
?	Guard	Charlie Ward	6′02″	185	41	Florida State
?	Defensive End	Julius Peppers	6′07″	283	32	North Carolina

Fortunately, you have brought your CS 188 notes along, and will build some classifiers to determine which sport is being discussed. You come across a pamphlet from the Atlantic Coast Conference Basketball Hall of Fame, as well as an Oakland Raiders team roster, and create the following table:

Sport	**Position**	**Name**	**Height**	**Weight**	**Age**	**College**
Basketball	Guard	Michael Jordan	6′06″	195	49	North Carolina
Basketball	Guard	Vince Carter	6′06″	215	35	North Carolina
Basketball	Guard	Muggsy Bogues	5′03″	135	47	Wake Forest
Basketball	Center	Tim Duncan	6′11″	260	35	Oklahoma
Football	Center	Vince Carter	6′02″	295	29	Oklahoma
Football	Kicker	Tim Duncan	6′00″	215	33	Oklahoma
Football	Kicker	Sebastian Janikowski	6′02″	250	33	Florida State
Football	Guard	Langston Walker	6′08″	345	33	California

2.1 Entropy

Before we get started, let's review the concept of entropy.

(b) Give the definition of entropy for an arbitrary probability distribution $P(X)$.

$H(X) = \sum_x P(x) \log_2(1/P(x))$.
You can see this as the expected information content of the distribution.

(c) Draw a graph of entropy $H(X)$ vs. $P(X = 1)$ for a binary random variable X.

$P(X = 1)H(X)$

(d) What is the entropy of the distribution of Sport in the training data? What about Position?

To calculate the entropy for a random variable, we estimate the probability distribution and use the formula from the part above.
$P(S = \text{football}) = 1/2$, $P(S = \text{basketball}) = 1/2$
$P(P = \text{guard}) = 1/2$, $P(P = \text{kicker}) = 1/4$, $P(P = \text{center}) = 1/4$
$H(S) = \frac{\log_2(2)}{2} + \frac{\log_2(2)}{2} = 1$
$H(P) = \frac{\log_2(2)}{2} + \frac{\log_2(4)}{4} + \frac{\log_2(4)}{4} = 3/2$

2.2 Decision Trees

Central to decision trees is the concept of "splitting" on a variable.

(e) To review the concept of "information gain", calculate it for a split on the Sport variable.

Since the variable that we want to predict is Sport, we want to be calculating the entropy with respect to the variable Sport.
(a) i. Distribution before: 8 examples with (1/2, 1/2). (here the first number in the tuple is P(basketball), and the second

number is P(football)).

ii. Entropy before: $(\frac{\log(2)}{2} + \frac{\log(2)}{2}) = 1$

(b) i. Distribution after: 4 examples with (1, 0), 4 examples with (0, 1)

ii. Entropy after: $\frac{4}{8}(\frac{\log(1)}{1}) + \frac{4}{8}(\frac{\log(1)}{1}) = 0$

So, the information gain is $(1 - 0) = 1$, which is the greatest possible.

(f) Of course, in our situation this would not make sense, as Sport is the very variable we lack at test time. Now calculate the information gain for the decision "stumps" (one-split trees) created by first splitting on Position, Name, and College. Do any of these perfectly classify the training data? Does it make sense to use Name as a variable? Why or why not?

Note that here we will be splitting on different variables but still need to look at the entropy of the distribution of the variable we need to predict which is sport. So, the before case remains same as before.

(a) **Position**: Distribution after: 4 examples with (3/4, 1/4), 2 examples with (1/2, 1/2), 2 examples with (0, 1).

Entropy after: $\frac{4}{8}(\frac{\log(4/3)}{4/3} + \frac{\log(4)}{4}) + \frac{2}{8}(\frac{\log(2)}{2} + \frac{\log(2)}{2}) + \frac{2}{8}(\frac{\log(1)}{1}) = 0.66$

(b) **Name**: Distribution after: 1 example with (1, 0), 2 examples with (1/2, 1/2), 1 example with (0, 1), 2 examples with (1/2, 1/2), 1 example with (0,1), 1 example with (0,1).

Entropy after: 0.5

(c) **College**: Distribution after: 2 examples with (1, 0), 1 example with (1, 0), 3 examples with (1/3, 2/3), 1 example with (0, 1), 1 example with (0,1).

Entropy after: 0.34

Note that none of these variables completely classifies the data.

Regarding using the Name as a feature to use in classifying data: since we expect people's names to be unique, using them as a feature in learning is akin to using the unique ID of each data point. That is to say, it's quite a bad idea you will overfit to the training data.

(g) Decision trees can represent any function of discrete attribute variables. How can we best cast continuous variables (Height, Weight, and Age) into discrete variables?

Use an inequality relation, Attribute $> a$, where a is a split point chosen to give the highest information gain. E.g., an initial split on Age > 34 will perfectly classify the training data.

(h) Draw a few decision trees that each correctly classify the training data, and show how their predictions vary on the test set. What algorithm are you following?

We use the algorithm as given in the slides, and for each split use the variable that gives us the maximum information gain. In this given problem, as we observed above, the variable Age correctly classifies all of the training data, so that is the first variable that gets picked up, and the algorithm stops at that.

This decision tree would predict test example 1 (Age 41) to be Basketball and test example 2 (Age 32) to be Football.

(i) You may have noticed that the testing data has a value for Position that is missing in training data. What could we do in this case?

When we come to a split on a variable whose value for the test subject is missing in the tree, we could just choose the most likely branch of the split (the branch that leads to the node with the greatest number of items).