Q1. Mini Power Pellets

Note: For this problem, any answers that require division can be left written as a fraction.

PacLabs has just created a new type of mini power pellet that is small enough for Pacman to carry around with him when he's running around mazes. Unfortunately, these mini-pellets don't guarantee that Pacman will win all his fights with ghosts, and they look just like the regular dots Pacman carried around to snack on.

Pacman just ate a snack (P), which was either a mini-pellet ($+p$), or a regular dot ($-p$), and is about to get into a fight (W), which he can win ($+w$) or lose ($-w$). Both these variables are unknown, but fortunately, Pacman is a master of probability. He knows that his bag of snacks has 5 mini-pellets and 15 regular dots. He also knows that if he ate a mini-pellet, he has a 70% chance of winning, but if he ate a regular dot, he only has a 20% chance.

(a) What is $P(+w)$, the marginal probability that Pacman will win?

(b) Pacman won! Hooray! What is the conditional probability $P(+p \mid +w)$ that the food he ate was a mini-pellet, given that he won?

Pacman can make better probability estimates if he takes more information into account. First, Pacman's breath, B, can be bad ($+b$) or fresh ($-b$). Second, there are two types of ghost (M): mean ($+m$) and nice ($-m$). Pacman has encoded his knowledge about the situation in the following probability distribution $P(M, P, B, W) = P(M)P(P)P(W|M, P)P(B|P)$:

P(M)	
+m	0.50
-m	0.50

P(P)	
+p	0.25
-p	0.75

P(W|M, P)			
+m	+p	+w	0.60
+m	+p	-w	0.40
+m	-p	+w	0.10
+m	-p	-w	0.90
-m	+p	+w	0.80
-m	+p	-w	0.20
-m	-p	+w	0.30
-m	-p	-w	0.70

$$P(M, P, B, W) = P(M)P(P)P(W|M, P)P(B|P)$$

P(B|P)		
+p	+b	0.80
+p	-b	0.20
-p	+b	0.40
-p	-b	0.60

(c) What is the probability of the atomic event $(-m, +p, +w, -b)$, where Pacman eats a mini-pellet and has fresh breath before winning a fight against a nice ghost?

For the remainder of this question, use the half of the joint probability table that has been computed for you below:

$P(M, P, W, B)$				
$+m$	$+p$	$+w$	$+b$	0.0600
$+m$	$+p$	$+w$	$-b$	0.0150
$+m$	$+p$	$-w$	$+b$	0.0400
$+m$	$+p$	$-w$	$-b$	0.0100
$+m$	$-p$	$+w$	$+b$	0.0150
$+m$	$-p$	$+w$	$-b$	0.0225
$+m$	$-p$	$-w$	$+b$	0.1350
$+m$	$-p$	$-w$	$-b$	0.2025

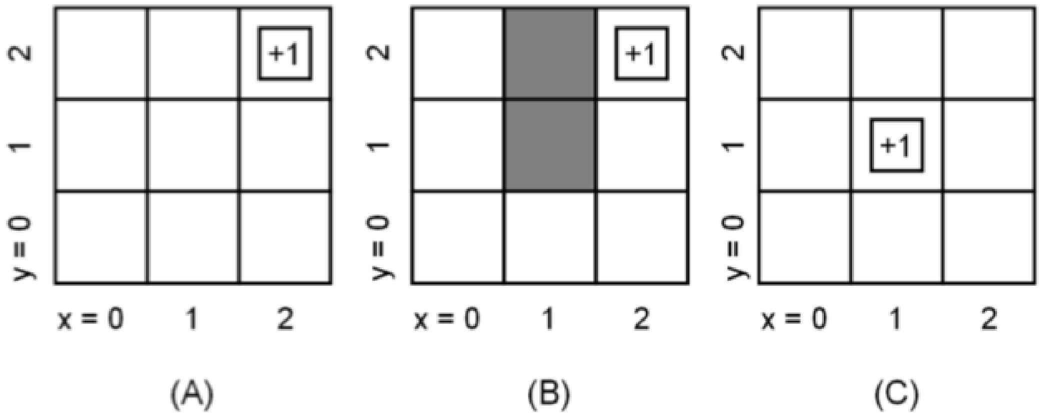
(d) What is the marginal probability, $P(+m, +b)$ that Pacman encounters a mean ghost and has bad breath?

(e) Pacman observes that he has bad breath and that the ghost he's facing is mean. What is the conditional probability, $P(+w \mid +m, +b)$, that he will win the fight, given his observations?

(f) Pacman's utility is +10 for winning a fight, -5 for losing a fight, and -1 for running away from a fight. Pacman wants to maximize his expected utility. Given that he has bad breath and is facing a mean ghost, should he stay and fight, or run away? Justify your answer numerically!

Q2. The Expressiveness of Feature-Space Representations

Consider the gridworlds below. The agent can take actions N, S, E, W, which always move it one square in the respective direction unless the agent is at the edge of the gridworld or moving into a wall. The squares marked with +1 are pre-terminal states, from which the agent can exit the gridworld. Rewards for all transitions are zero, except the exit transition, which has reward +1. Assume a discount of 0.5.



(a) Fill in the optimal values for grid (A), and specify the optimal policies for grids (B) and (C) by placing an arrow in each empty square.

Imagine we have a set of real-valued features $f_i(s)$ for each non-terminal state $s = (x, y)$, and we wish to approximate the optimal utility values $V^*(s)$ by $V(s) = \sum_i w_i \cdot f_i(s)$ (linear feature-based approximation).

(b) If our features are $f_1(x, y) = x$ and $f_2(x, y) = y$, give values of w_1 and w_2 for which a one-step look-ahead policy extracted from V will be optimal in grid A.

(c) Can we represent the actual optimal values V^* for grid (A) using these two features? Why or why not?

(d) For each of the features sets listed below, state which (if any) of the grid MDPs above can be solved, in the sense that we can express some (possibly non-optimal) values which produce optimal one-step look- ahead policies.

i. $f_1(x, y) = x$ and $f_2(x, y) = y$.

ii. For each (i, j), a feature $f_{i,j}(x, y) = 1$ if $(x, y) = (i, j)$, 0 otherwise.

iii. $f_1(x, y) = (x - 1)^2, f_2(x, y) = (y - 1)^2$ and $f_3(x, y) = 1$.