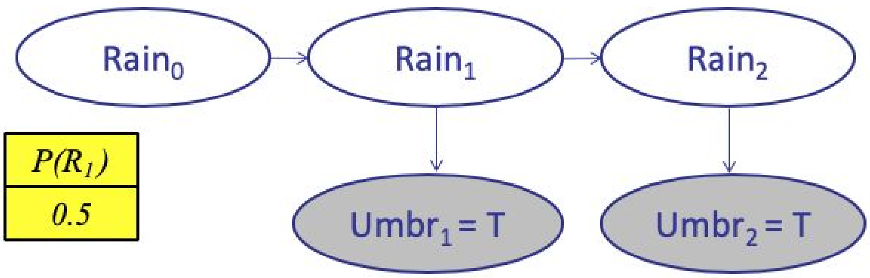
Question 1. Given an HMM with parameters specified below, calculate P(Rain$_2$ | Umbr$_1$, Umbr$_2$), i.e. that it's rainy on day 2 after observing umbrellas on day 1 and day 2.

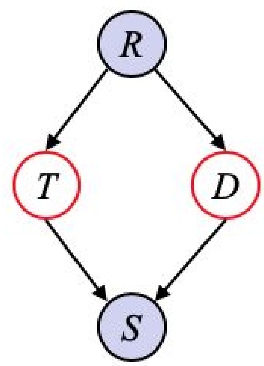


$P(R_1)$
0.5

R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.8
f	0.6

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.3

Question 2: Given the Bayes Net shown to the right, use D-separation to answer the following questions about conditional independence:



A) $T \perp\!\!\!\perp D$

B) $T \perp\!\!\!\perp D \mid R$

C) $T \perp\!\!\!\perp D \mid R, S$

Q4. [17 pts] Probability and Bayes Nets

(a) [2 pts] Suppose $A \perp\!\!\!\perp B$. Determine the missing entries (x, y) of the joint distribution $P(A, B)$, where A and B take values in $\{0, 1\}$.

$$P(A = 0, B = 0) = 0.1$$
$$P(A = 0, B = 1) = 0.3$$
$$P(A = 1, B = 0) = x$$
$$P(A = 1, B = 1) = y$$

$x =$ ____________, $y =$ ____________

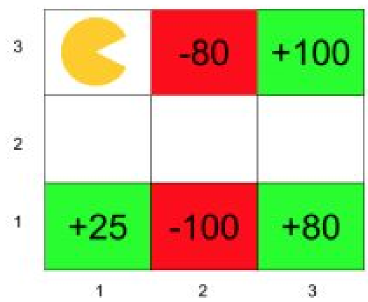
(b) [3 pts] Suppose $B \perp\!\!\!\perp C \mid A$. Determine the missing entries (x, y, z) of the joint distribution $P(A, B, C)$.

$$P(A = 0, B = 0, C = 0) = 0.01$$
$$P(A = 0, B = 0, C = 1) = 0.02$$
$$P(A = 0, B = 1, C = 0) = 0.03$$
$$P(A = 0, B = 1, C = 1) = x$$
$$P(A = 1, B = 0, C = 0) = 0.01$$
$$P(A = 1, B = 0, C = 1) = 0.1$$
$$P(A = 1, B = 1, C = 0) = y$$
$$P(A = 1, B = 1, C = 1) = z$$

$x =$ ____________, $y =$ ____________, $z =$ ____________

Q8. [8 pts] Q-Learning Strikes Back

Consider the grid-world given below and Pacman who is trying to learn the optimal policy. If an action results in landing into one of the shaded states the corresponding reward is awarded during that transition. All shaded states are terminal states, i.e., the MDP terminates once arrived in a shaded state. The other states have the *North, East, South, West* actions available, which deterministically move Pacman to the corresponding neighboring state (or have Pacman stay in place if the action tries to move out of the grad). Assume the discount factor $\gamma = 0.5$ and the Q-learning rate $\alpha = 0.5$ for all calculations. Pacman starts in state (1, 3).



(a) [2 pts] What is the value of the optimal value function V^* at the following states:

$V^*(3,2) =$ __________ $V^*(2,2) =$ __________ $V^*(1,3) =$ __________

(b) [3 pts] The agent starts from the top left corner and you are given the following episodes from runs of the agent through this grid-world. Each line in an Episode is a tuple containing (s, a, s', r).

Episode 1	Episode 2	Episode 3
(1,3), S, (1,2), 0	(1,3), S, (1,2), 0	(1,3), S, (1,2), 0
(1,2), E, (2,2), 0	(1,2), E, (2,2), 0	(1,2), E, (2,2), 0
(2,2), S, (2,1), -100	(2,2), E, (3,2), 0	(2,2), E, (3,2), 0
	(3,2), N, (3,3), +100	(3,2), S, (3,1), +80

Using Q-Learning updates, what are the following Q-values after the above three episodes:

$Q((3,2),N) =$ __________ $Q((1,2),S) =$ __________ $Q((2,2),E) =$ __________

(c) Consider a feature based representation of the Q-value function:

$$Q_f(s,a) = w_1 f_1(s) + w_2 f_2(s) + w_3 f_3(a)$$

$f_1(s)$: The x coordinate of the state $\qquad$ $f_2(s)$: The y coordinate of the state

$$f_3(N) = 1,\ f_3(S) = 2,\ f_3(E) = 3,\ f_3(W) = 4$$

(i) [2 pts] Given that all w_i are initially 0, what are their values after the first episode:

$w_1 =$ __________ $w_2 =$ __________ $w_3 =$ __________

(ii) [1 pt] Assume the weight vector w is equal to $(1, 1, 1)$. What is the action prescribed by the Q-function in state $(2, 2)$?

____