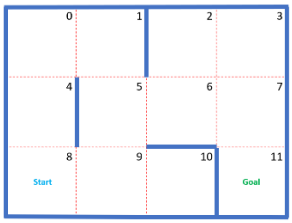
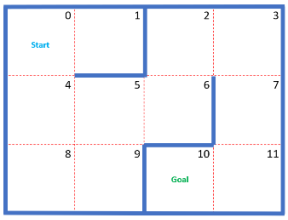
Deep Learning World / Predictive Analytics World
Exercises - Simple Q-Learning

1. Write a program that uses Q-learning to solve this 12-state maze. Write your program so that a shorter path to the goal state is preferred to a longer path. Use a gamma discount factor value of 0.5 but do not use a learning rate.



2. Modify your program to solve this maze.



3. Which statement best describes the equation?

$$Q(s_t, a_t) = [\,(1-\alpha) * Q(s_t, a_t)\,] + [\,\alpha * (r_t + \gamma * maxQ(s_{t+1}, a)\ \forall\ a\,)\,]$$

a.) The Bellman equation to update Q with learning rate alpha and discount factor gamma.
b.) The Ackermann function to estimate the best discount factor gamma.
c.) The SARSA (state-action-reward-state-action) equation for on-policy problems.
d.) The DDPG (deep deterministic policy gradient) equation for a continuous action space.

4. Which of the following statements regarding Q-learning gamma value is most accurate?

a.) In a fully deterministic environment, gamma should be set to -1.0 or less.
b.) The larger gamma is, the more the algorithm strives for long-term reward.
c.) When the value function is approximated by a neural network, gamma should be set to 1.0.
d.) None of the above are true.

5. Which of the following statements regarding Q-learning learning rate is most accurate?

a.) The learning rate controls the explore-exploit tradeoff, where larger values of the learning rate emphasize explore.
b.) The learning rate controls the explore-exploit tradeoff, where larger values of the learning rate emphasize exploit.
c.) In a fully deterministic environment, the learning rate should be set to e = 2.71828.
d.) In a stochastic environment, the learning rate should be set to pi = 3.14159.