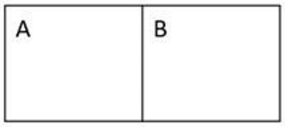
Part 3 Extra Examples

November 24, 2022 11:22 PM

Vacuum Scenario
Imagine a vacuum cleaner agent operating in a room with two squares which can be either “dirty” or “clean”. The environment reaches a terminal state when all the squares in the room are clean.



Suppose at each state, the agent can take one of the following actions:
Move to an adjacent square
Clean the current square

Note that with some small random chance, taking the “Clean” action will do nothing (i.e. not change the state).

The reward associated with each state is:
r(s) = -1 * number of dirty squares

Let’s use the following abbreviation for states:
A00 = Vacuum in A, A clean, B clean
A01 = Vacuum in A, A clean, B dirty
A10 = Vacuum in A, A dirty, B clean
A11 = Vacuum in A, A dirty, B dirty
B00 = Vacuum in B, A clean, B clean
B01 = Vacuum in B, A clean, B dirty
B10 = Vacuum in B, A dirty, B clean
B11 = Vacuum in B, A dirty, B dirty

Suppose we simulate two trials through the state-action space, and they are given below:
(A11, Clean) -> (A01, Move Right) -> (B01, Clean) -> (B01, Clean) -> (B00, None)
(B11, Clean) -> (B10, Clean) -> (B10, Move Left) -> (A10, Clean) -> (A00, None)

Example 1.
Consider the Vacuum Scenario above. Suppose the above trial was generated by a fixed policy. Estimate the value function V(s) associated with each state for the fixed policy described above using direct estimation. Use discount factor gamma = 0.5.

Example 2.
Consider the Vacuum Scenario above. Suppose the above trial was generated by a fixed policy. Estimate the value function V(s) associated with each state for the fixed policy described above using adaptive dynamic programming. Use discount factor gamma = 0.5. Initially estimate for V(s) = -1 for states with one or more dirty square; V(s) = 0 otherwise.

Example 3.
Consider the Vacuum Scenario above. Suppose the above trial was generated by a fixed policy. Estimate

the value function V(s) associated with each state for the fixed policy described above using temporal difference learning. Use discount factor gamma = 0.5; use learning rate alpha = 0.5. Initially estimate for V(s) = -1 for states with one or more dirty square; V(s) = 0 otherwise.

Example 4.
Consider the Vacuum Scenario above. Estimate the Q-value function Q(s, a) associated with each state-action pair using temporal difference Q-learning. described above using temporal difference learning. Use discount factor gamma = 0.5; use learning rate alpha = 0.5. Initially estimate for Q(s, a) = -1 for states with one or more dirty square; Q(s, a) = 0 otherwise.

Example 5.
Consider Scenario 1 above. Suggest two different features (or basis functions) that could be used for performing reinforcement learning with a large state space approximation in this environment.

Example 6.
Consider the knapsack problem.

We have n items, where the i-th item has weight w_i and value v_i. Given a knapsack that can carry up to weight W, which items should we put into the knapsack to maximize value?

Suppose we want to solve this using the genetic algorithm. Suggest a possible representation of solutions. Suggest the following operations:
Fitness function, genetic operator, mutation

Consider an instance of this problem with the following items:
W = 10. w1 = 4, v1 = 3; w2 = 3, v2 = 4; w3 = 6, v3 = 7; w4 = 3, v4 = 3.
Show one example iteration of the genetic algorithm for this problem.

Example 7.
Consider a 2-gram character model. Build our model using the following dataset:
“she sees cheese”

Compute the probability of the sequence of characters:
“hee”

Example 8.
Build a vocabulary and compute the occurrence vectors, count vectors, term frequency vectors, and tf-idf vectors for the following three documents.

“i scream you scream”
“we all scream for ice cream”
“ice cold ice cream”