Home Work 2

We consider the Windy Gridworld with the following changes:

- We allow diagonal moves, so there are now 8 possible actions from each square (king's moves). The initial policy is probability 1/8 for each direction
$$\{'n','ne','e','se','s',,'sw','w','nw'\}$$
- There is now a 'death square' at position (1,10) which is a terminal state with reward -100
- The wind has the same effect as before so for example if you are in state $(6, 4)$ and you move 'ne', without wind you would go to $(5, 5)$ but with the wind you go to $(4, 5)$. If you are in $(6, 7)$ and move 'se' you go to $(5, 8)$ and moving 'e' you go to $(4, 8)$ etc.

1. Use both SARSA and Q-learning to find optimal policies when starting from (7,1),from (4,1) and from (1,7)

2. What happens if we do not allow diagonal moves (so only rook's moves)?

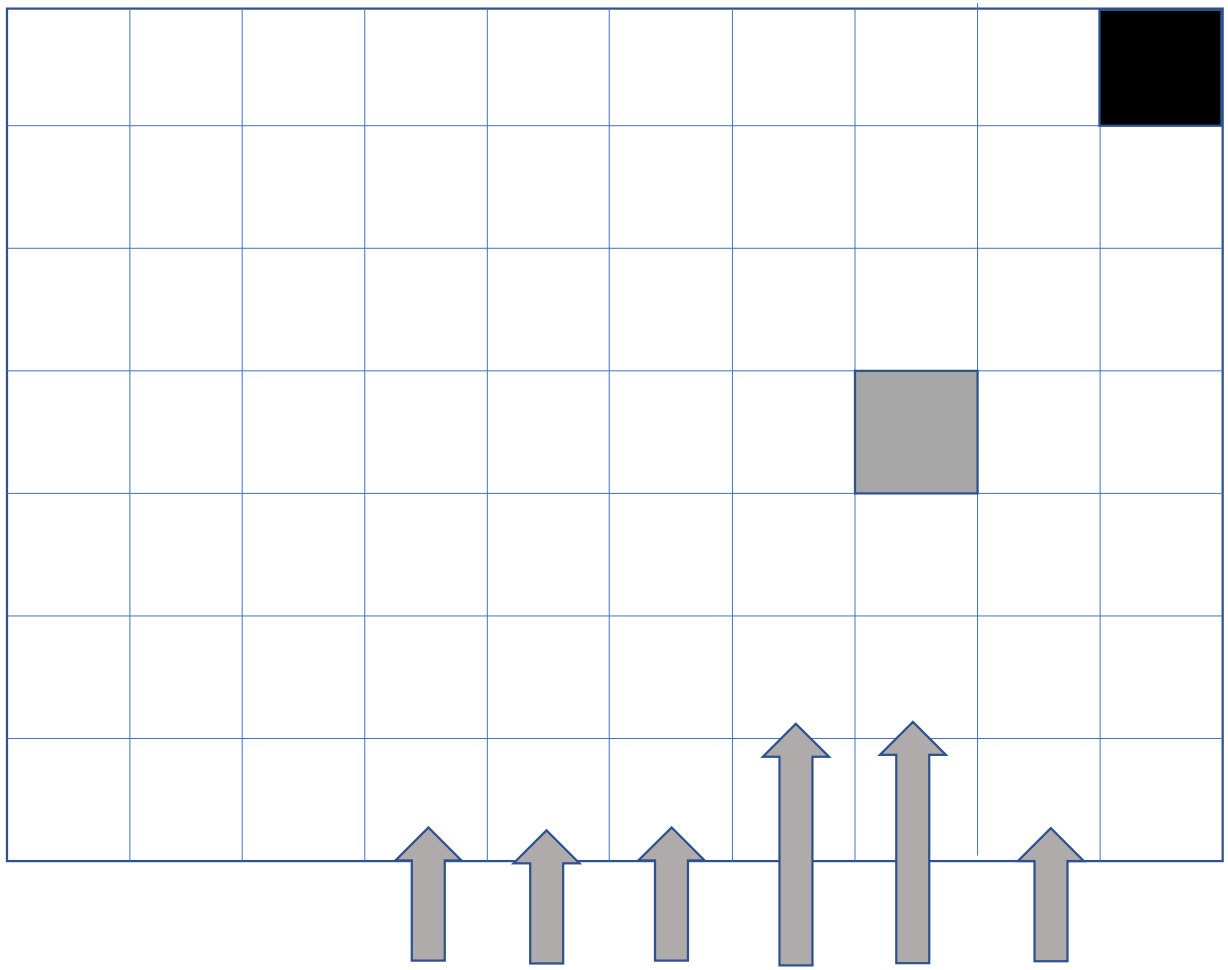
3. Ask an LLM to generate the code for this problem without specifying which algorithm to use and determine which algorithm it chooses

4. Ask it to generate code for SARSA and for Q-Learning and check that it gives the same results as your own code

5. Can you with just the model description get it to generate optimal trajectories without writing the code, for instance ask it to return the trajectory from $(1, 1)$ follwing an optimal policy

6. See what it does if you don't allow diagonal moves

7. Add an absorbing state with reward 5 at $(1, 4)$ and do 1,2,3,4,5,6 for this grid



The blue square is an obsorbing state with reward 5