**INITIALLY**

# defines the reward matrix  
r = np.array([[-1, -1, -1, -1, 0, -1],  
 [-1, -1, -1, 0, -1, 100],  
 [-1, -1, -1, 0, -1, -1],  
 [-1, 0, 0, -1, 0, -1],  
 [ 0, -1, -1, 0, -1, 100],  
 [-1, 0, -1, -1, 0, 100]]).astype("float64")  
q = np.zeros\_like(r)

# hyperparameters

gamma = 0.8 #vary this between 0 and 1 i.e., (0,1)

alpha = 0.9 #vary this between 0 and 1 inclusive, i.e., [0,1]

n\_episodes = 1E3 #try different values, e.g., 10, 500, 10000 and so forth

epsilon = 0.05 #you can experiment with this as well

**INITIAL OUTPUT**

Q Table

[[0.00000000e+000 0.00000000e+000 0.00000000e+000 0.00000000e+000

1.00000000e+000 0.00000000e+000]

[0.00000000e+000 0.00000000e+000 0.00000000e+000 7.41976889e-297

0.00000000e+000 1.00000000e+000]

[0.00000000e+000 0.00000000e+000 0.00000000e+000 1.00000000e+000

0.00000000e+000 0.00000000e+000]

[0.00000000e+000 7.56550285e-001 3.65809180e-003 0.00000000e+000

2.39791623e-001 0.00000000e+000]

[0.00000000e+000 0.00000000e+000 0.00000000e+000 2.45056560e-320

0.00000000e+000 1.00000000e+000]

[0.00000000e+000 0.00000000e+000 0.00000000e+000 0.00000000e+000

4.21778013e-048 1.00000000e+000]]

Optimal Path for starting state 0

0 -> 4 -> 5

Optimal Path for starting state 1

1 -> 5

Optimal Path for starting state 2

2 -> 3 -> 1 -> 5

Optimal Path for starting state 3

3 -> 1 -> 5

Optimal Path for starting state 4

4 -> 5

Optimal Path for starting state 5

5

**AFTER CHANGING HYPERPARAMETERS**

# hyperparameters  
gamma = 0.3 #vary this between 0 and 1 i.e., (0,1)  
alpha = 0.9 #vary this between 0 and 1 inclusive, i.e., [0,1]  
n\_episodes = 400 #try different values, e.g., 10, 500, 10000 and so forth  
epsilon = 0.05 #you can experiment with this as well

Q Table

[[0.00000000e+000 0.00000000e+000 0.00000000e+000 0.00000000e+000

1.00000000e+000 0.00000000e+000]

[0.00000000e+000 0.00000000e+000 0.00000000e+000 1.05142911e-005

0.00000000e+000 9.99989486e-001]

[0.00000000e+000 0.00000000e+000 0.00000000e+000 1.00000000e+000

0.00000000e+000 0.00000000e+000]

[0.00000000e+000 3.18707031e-001 3.30146330e-001 0.00000000e+000

3.51146639e-001 0.00000000e+000]

[2.56426235e-140 0.00000000e+000 0.00000000e+000 1.81457827e-060

0.00000000e+000 1.00000000e+000]

[0.00000000e+000 0.00000000e+000 0.00000000e+000 0.00000000e+000

0.00000000e+000 1.00000000e+000]]

Optimal Path for starting state 0

0 -> 4 -> 5

Optimal Path for starting state 1

1 -> 5

Optimal Path for starting state 2

2 -> 3 -> 4 -> 5

Optimal Path for starting state 3

3 -> 4 -> 5

Optimal Path for starting state 4

4 -> 5

Optimal Path for starting state 5

5