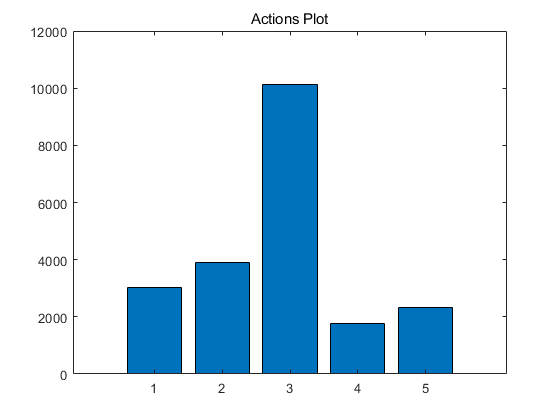
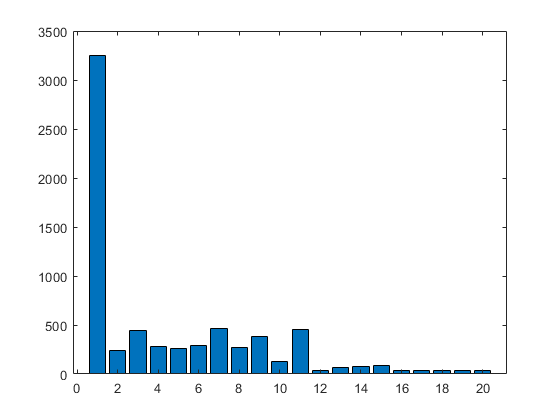
랜덤 실행 결과



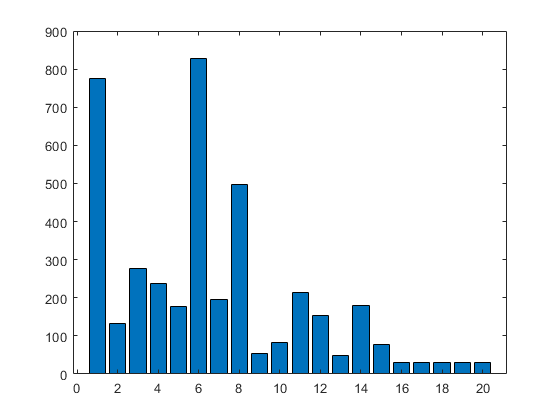
q-learning 하이퍼 파라미터 조정에 따른 훈련 결과

alpha 0.8

gamma0.99

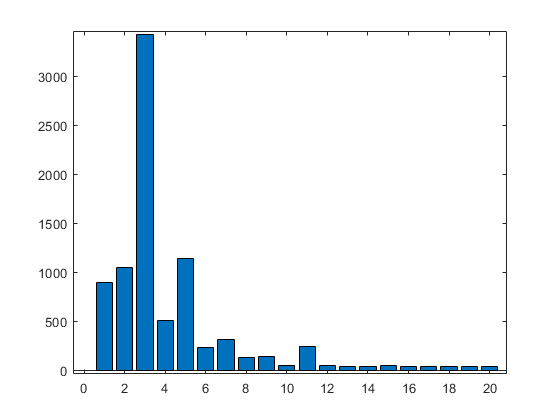
alpha 0.8

gamma0.7



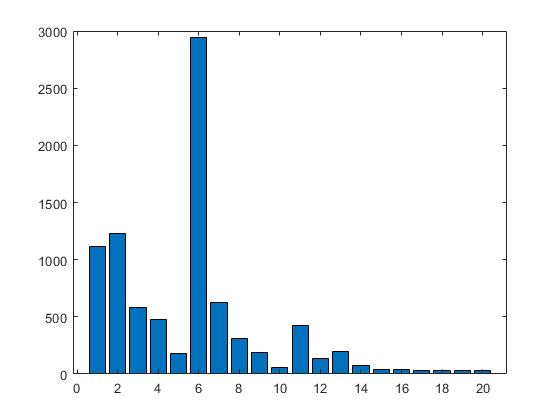
alpha 0.8

gamma 0.5



alpha 0.8

gamma0.3



미래보상에 대한 할인율 파라미터인 감마값을 조정하면서 훈련 결과를 관찰

작은 지도였기 때문에 수렴 시간에는 큰 차이가 없음

감마값이 높은 경우에 초기에 미로의 많은 부분을 탐색하고 점차적으로 시간이 내려오는 반면에 감마값이 낮은 경우 새로운 경로를 탐색하며 시간이 길어지는 경우가 좀 더 흩어져 있음