

## Same test bandit set

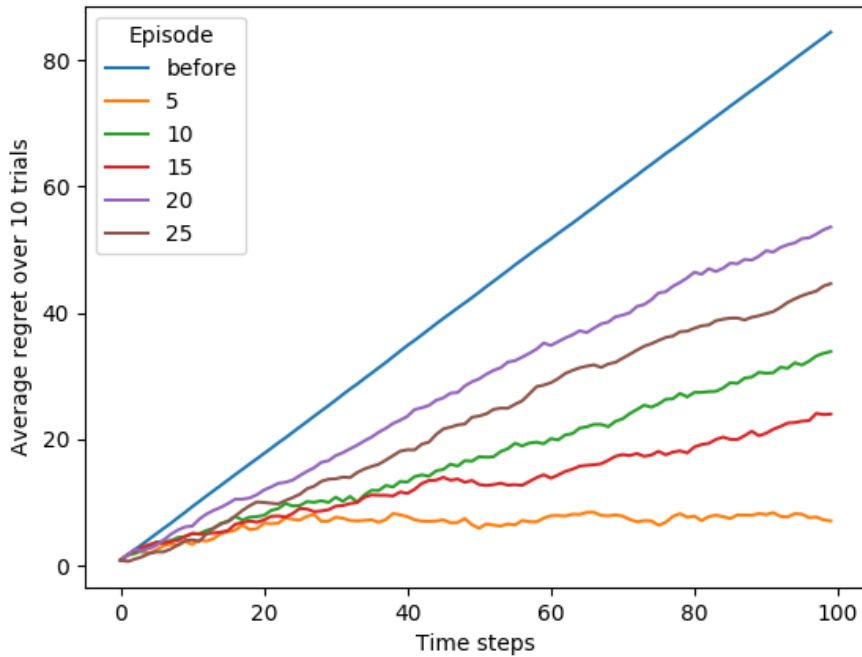
Test bandits

1: 0.364      1.315

2: -0.480     0.072

DQN

Average regret v/s time steps for 2 bandits. (epsilon = 0.2, beta = 0.9)



DQN 2 bandits

**mean = random.uniform(-10, 10)**

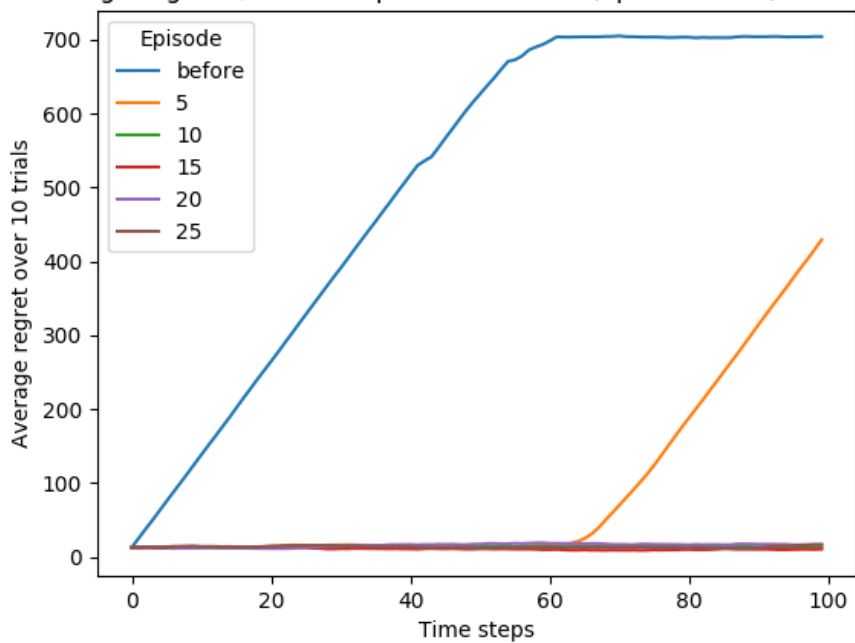
**sigma = random.uniform(0, 2)**

Test bandit:

-9.304 0.865

3.332 1.128

Average regret v/s time steps for 2 bandits. (epsilon = 0.2, beta = 0.9)



DQN

### 5 bandits

Test bandit:

0.433 1.089

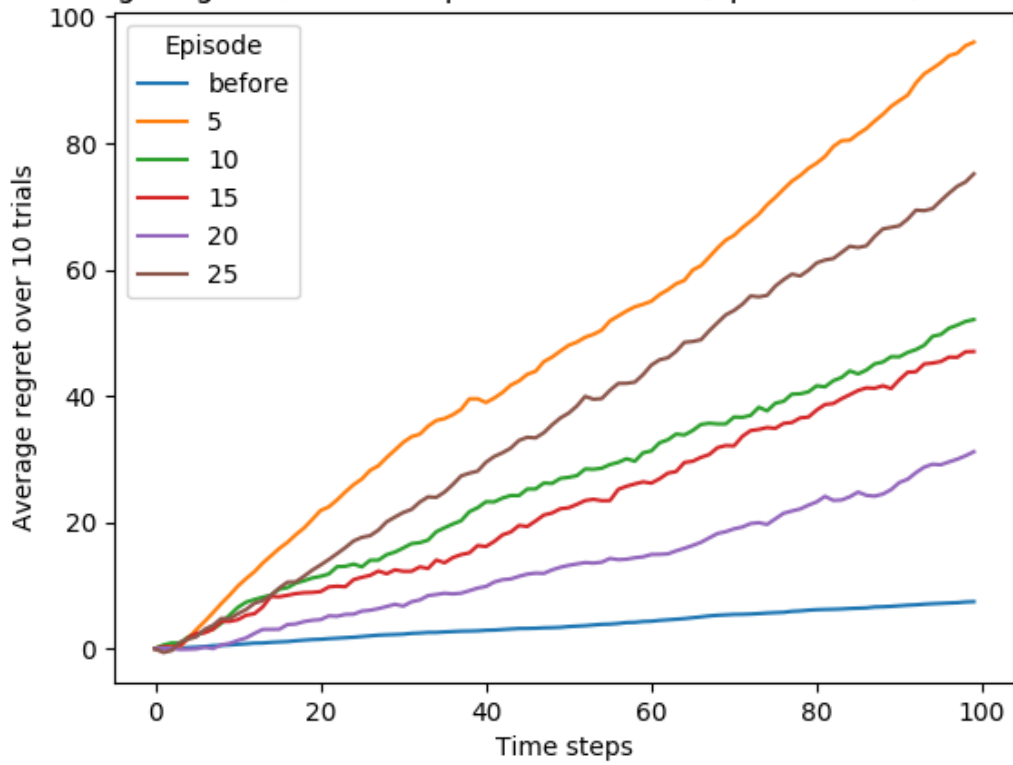
0.572 1.422

0.494 0.115

-0.706 0.830

-0.810 0.273

Average regret v/s time steps for 5 bandits. (epsilon = 0.2, beta = 0.9)



DQN

**8 bandits**

**mean = random.uniform(-10, 10)**

**sigma = random.uniform(0, 2)**

Test bandit:

9.945 1.630

-0.704 0.234

-6.633 1.415

8.687 1.003

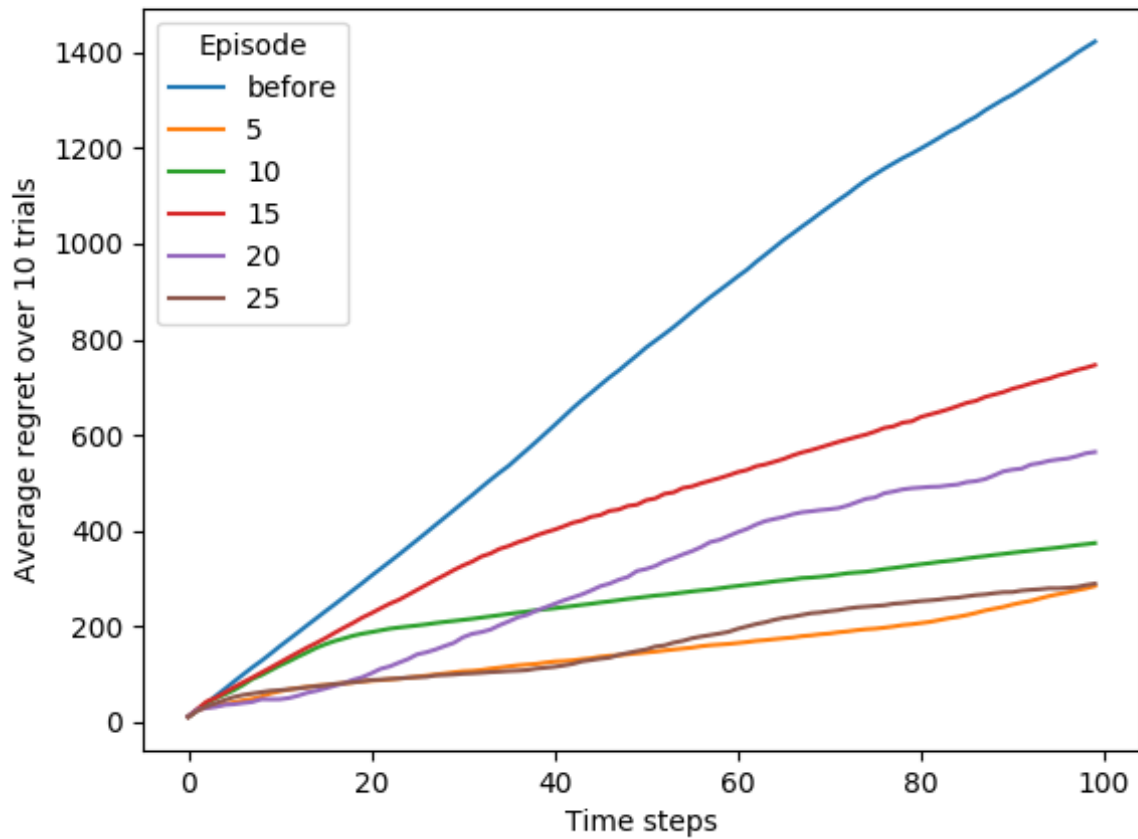
9.654 0.640

-7.794 0.998

-0.761 1.678

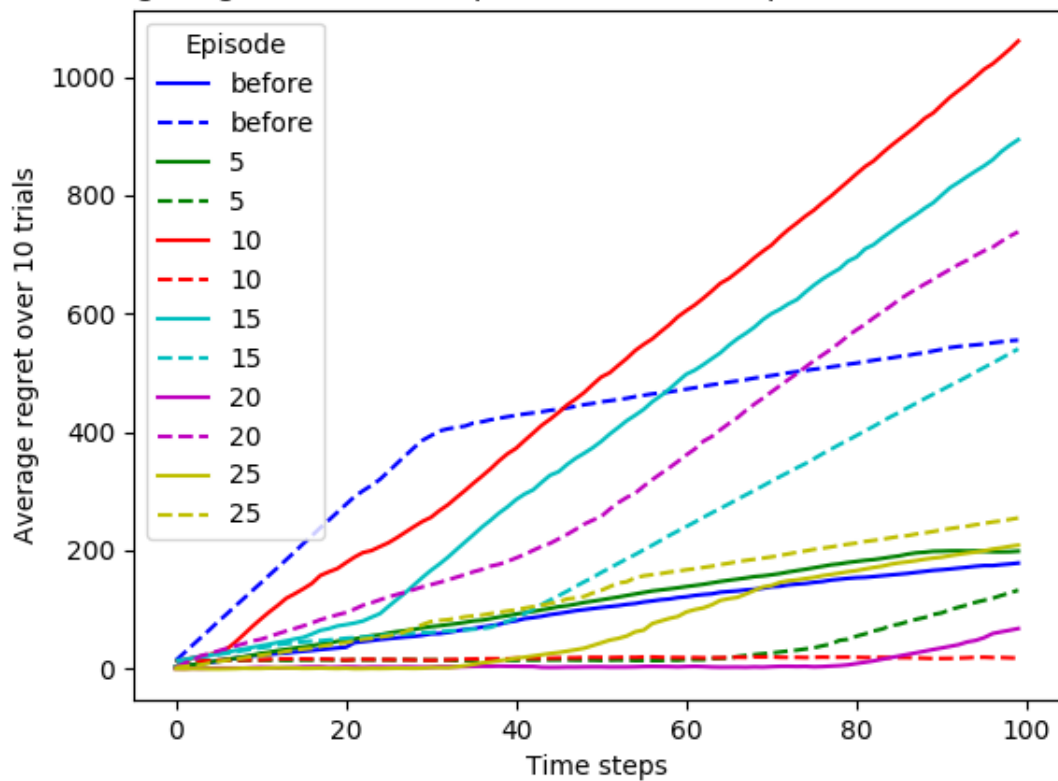
3.348 1.832

Average regret v/s time steps for 8 bandits. (epsilon = 0.2, beta = 0.9)



## DQN v/s Double Q

Average regret v/s time steps for 8 bandits. (epsilon = 0.2, beta = 0.9)



## 8 bandits

Test bandits :

3.273 0.259  
-1.100 1.235  
3.976 1.451  
-8.003 0.666  
3.791 1.242  
-7.160 0.807  
6.038 1.177  
3.850 1.114