

Epsilon in DRL



THE UNIVERSITY OF
SOUTHERN MISSISSIPPI®

DRL

- So far, a DRL model has multiple hyperparameters:
 - Epochs (episodes)
 - Dropout
 - Learning rate
 - Tied to optimization method
 - Epsilon
 - Linear decay
 - Exponential decay



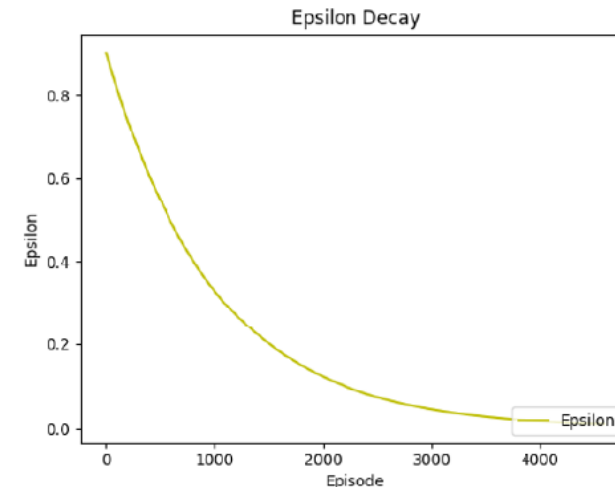
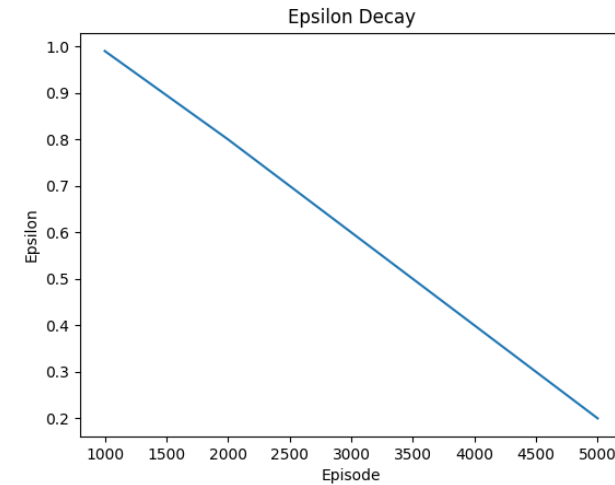
Gathering information

- The agent needs to gather information:
 - We control this by using:
 - Exploration – Exploitation methods
 - Upper confidence bound:
 - Looks for “uncertainty”
 - “Locks in” a Q-value and explores around it
 - Epsilon-greedy (ϵ -greedy)
 - Linear decay
 - Exponential decay
 - Other forms:
 - sinusoidal -> depends on the environment
 - Softmax = Not to be confused with Softmax activation method



ϵ -greedy

- With ϵ -greedy, the decay can be controlled:
 - Linear
 - Decay occurs in a linear form
 - Exploration – exploitation is “constant”
 - Variation:
 - Decay rate (decay step)
 - Exponential
 - Decay occurs in an exponential form
 - Exploration – exploitation is variable
 - Explore at the beginning, exploit at the end
 - Variation:
 - Decay rate (decay step, & explore, exploit rate)



Environments

- For example:
 - Slot machine:
 - One arm
 - Different combinations to get a reward
 - One action – get rewarded
 - One – arm bandit
 - Multiple – arm bandit
 - How to explore – exploit the environment:
 - One arm, or multiple arm
 - Linear decay
 - Exponential decay





THE UNIVERSITY OF
SOUTHERN MISSISSIPPI.

