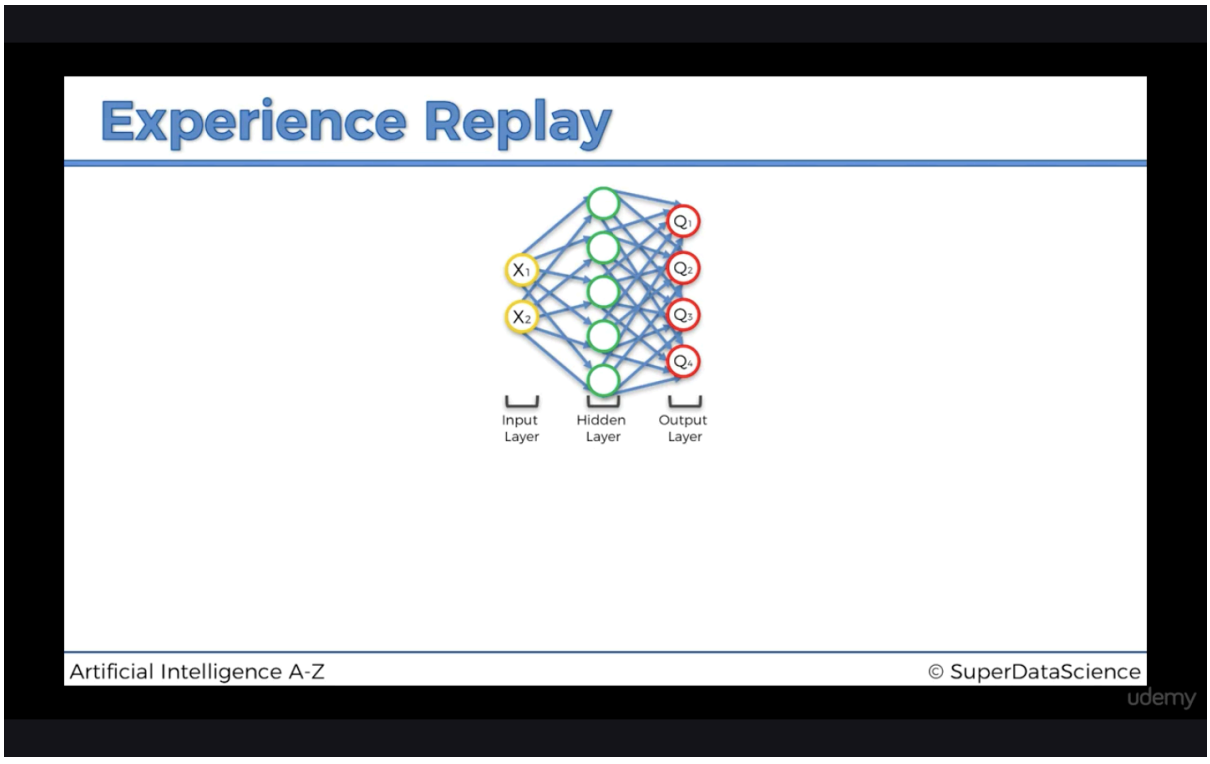
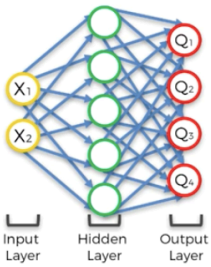
Experience replay



Experience Replay



Learning:

VS. Q-Target1

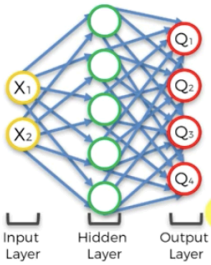
VS. Q-Target2

VS. Q-Target3

VS. Q-Target4

$$L = \sum (Q\text{-}Target - Q)^2$$

Experience Replay



Learning:

VS. Q-Target1

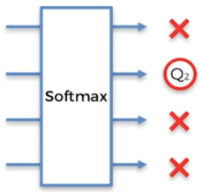
VS. Q-Target2

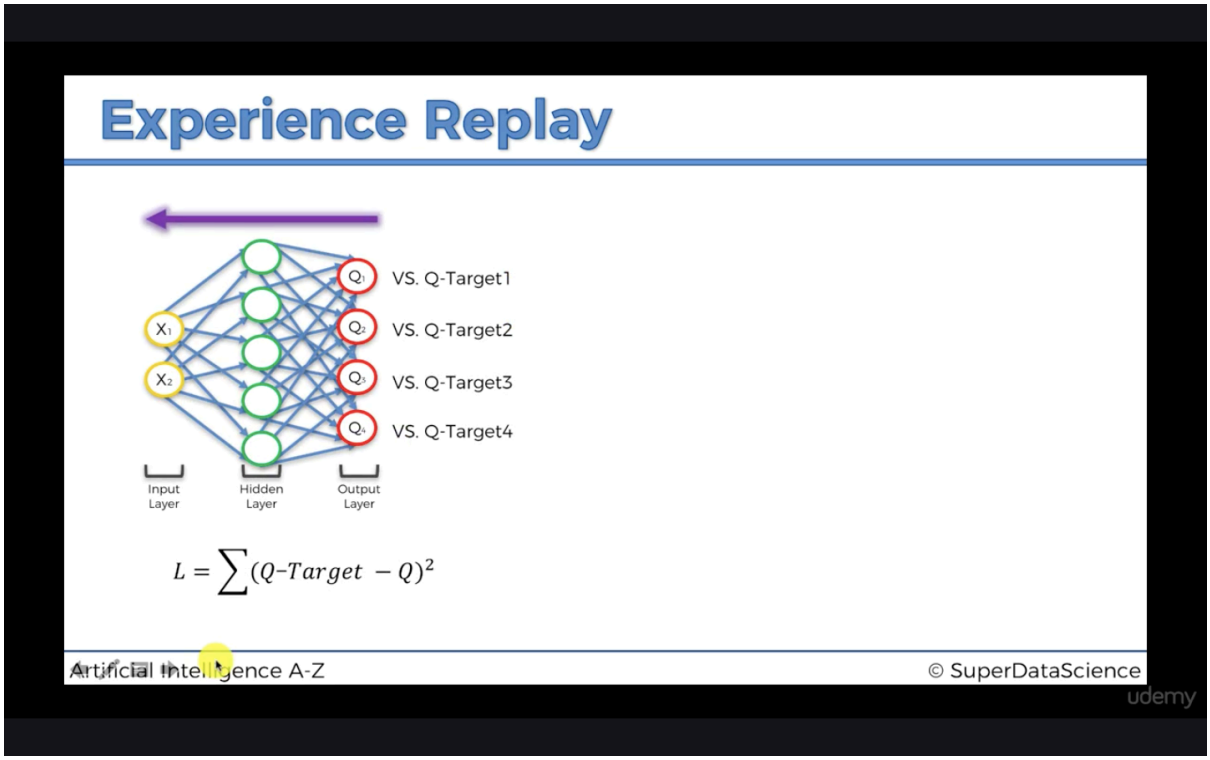
VS. Q-Target3

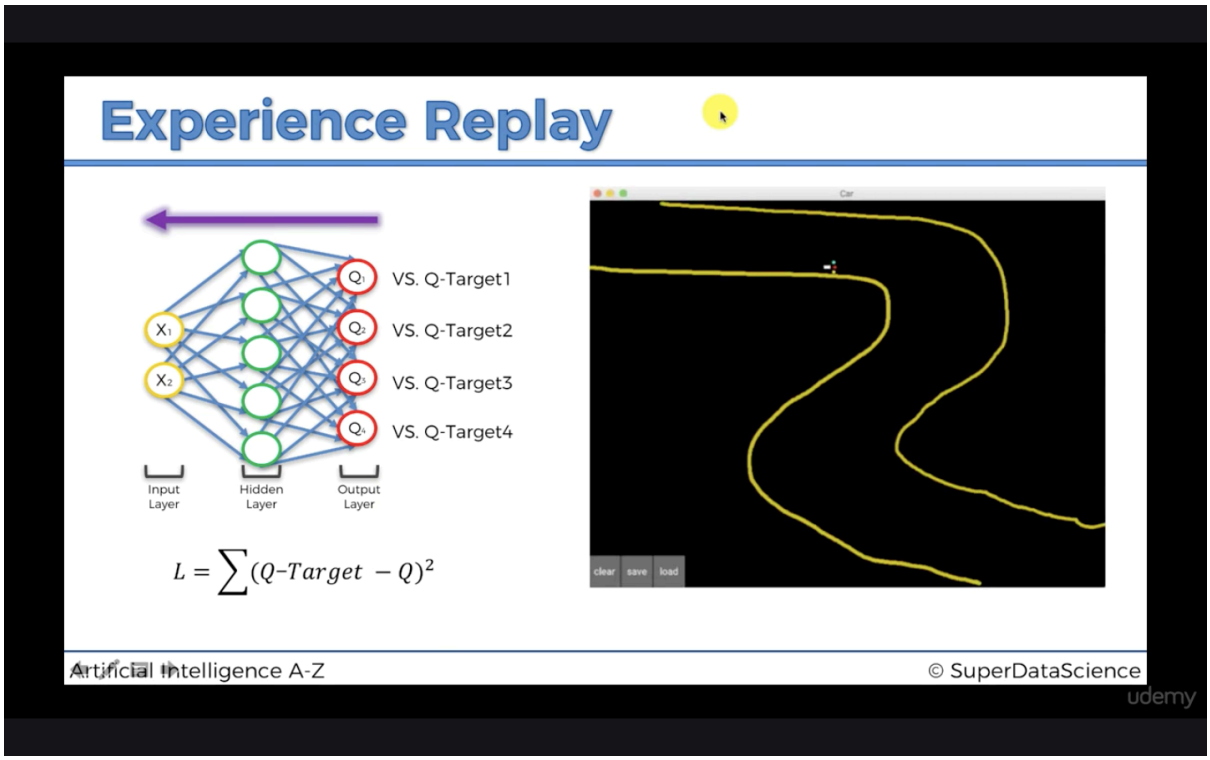
VS. Q-Target4

$$L = \sum (Q\text{-}Target - Q)^2$$

Acting:







The important thing to note that in this will not trained based on every single action in a state (because if we do the car will learn a single move that doesn't changes over and over. At the start, it learns how to drive on a straight line along the yellow line. With this technique, the car going to have a false perception). In here this will learn from taking a uniformly distributed sample from that batch of experiences that it has and then learns from them. It means that for example if we had 50 states at the beginning of straight line, they won't go to the network right away. They are actually saved into a memory of the agent and when it reaches to a threshold then the agent decides to learn it. If we learn from

every single action in a state, there is going to be a high bias that comes from the sequential nature of experience.