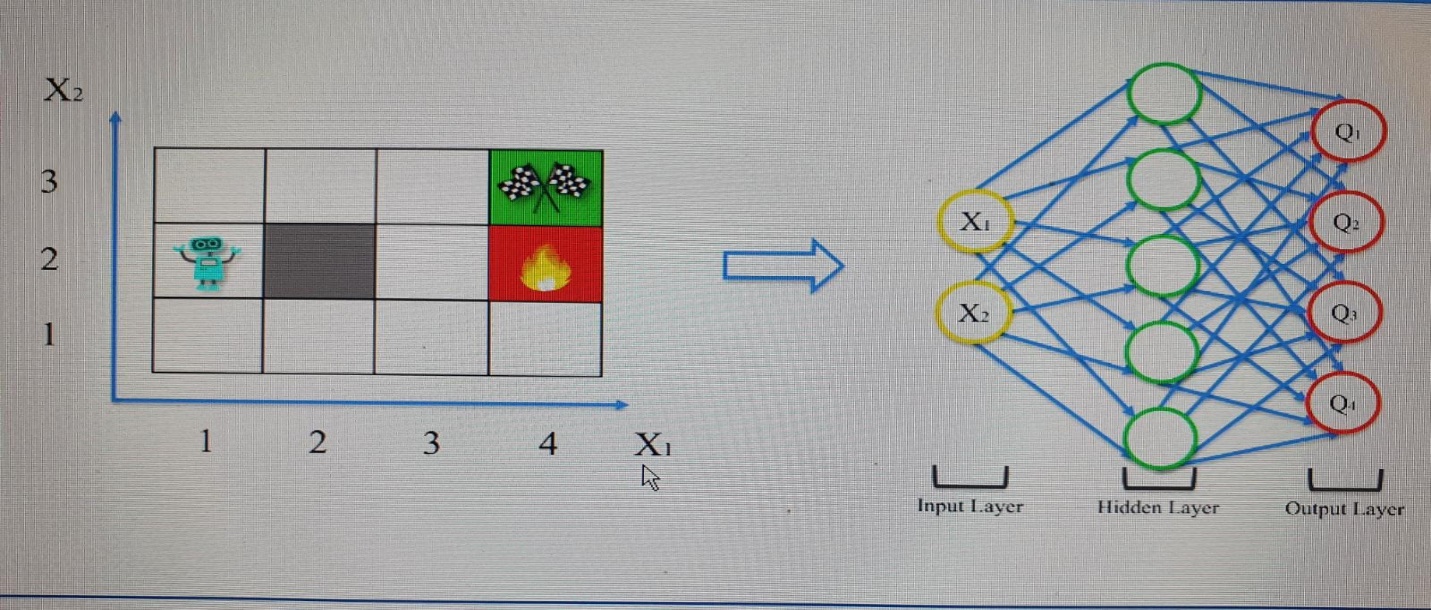
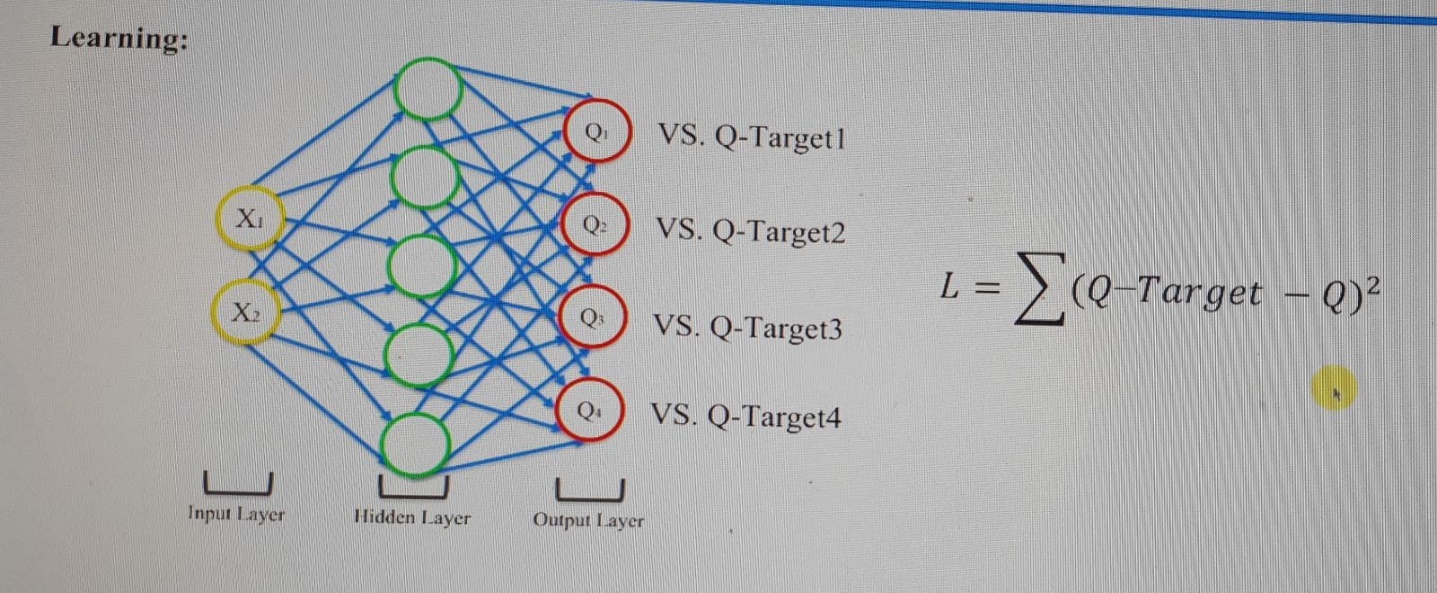
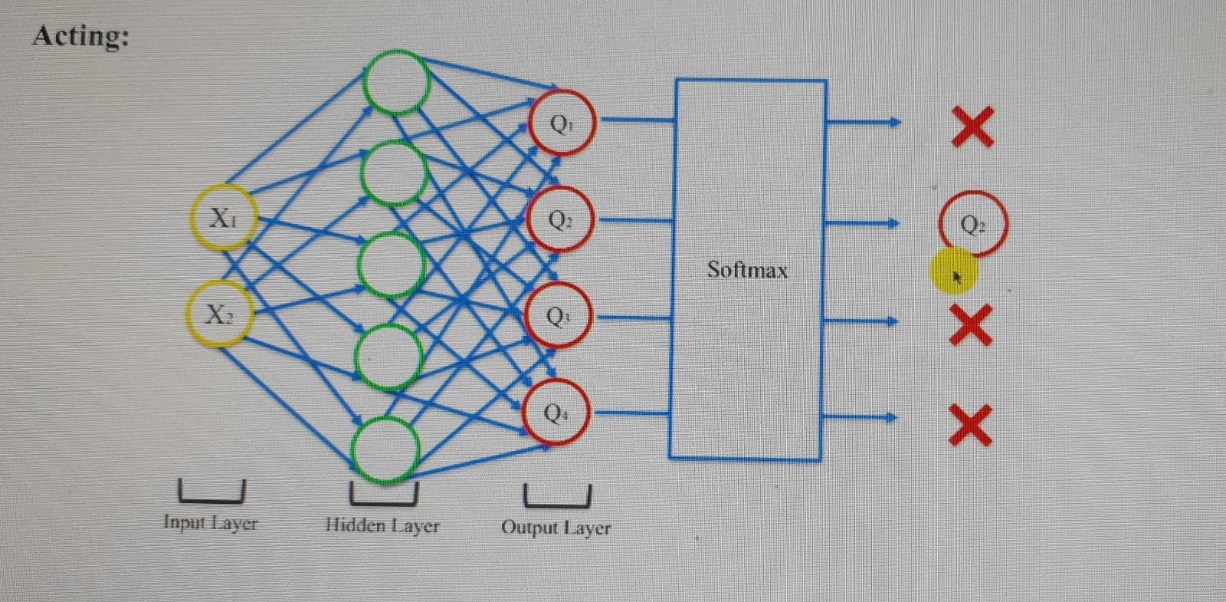
**Deep Q-Learning intuition**

****

****

We want Loss (L) should be close to zero.

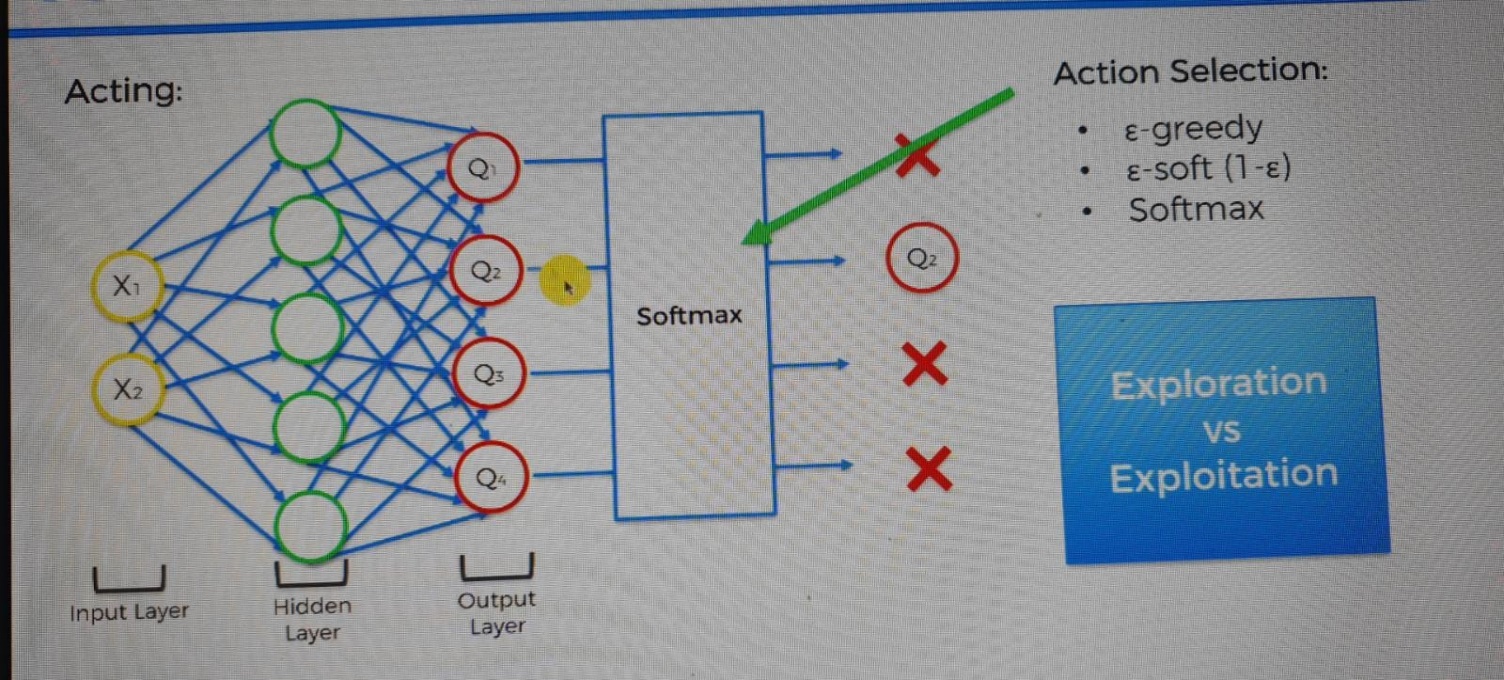
**Acting**

****

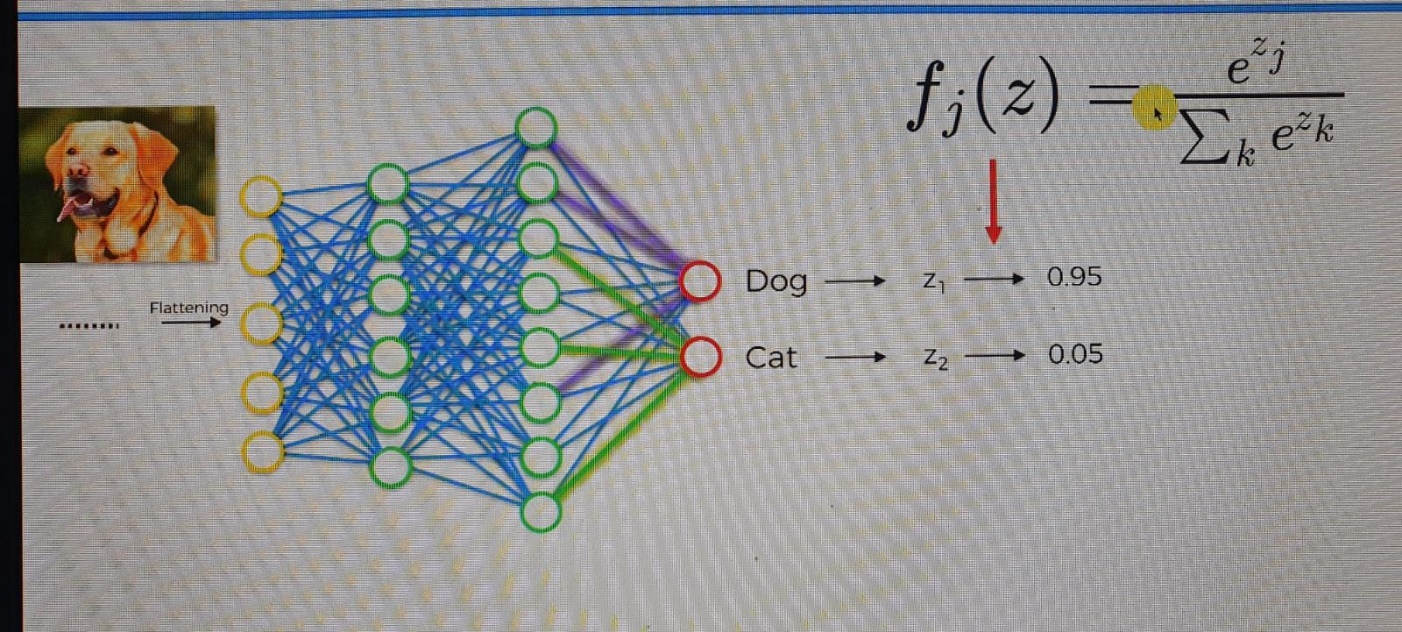
We pass the to softmax to choose the best Q (the best action).

**Experience Replay**

**Action Selection Policies**



Agent should not be stuck in local maxima. It must keep learning continuously

1. E-Greedy – chooses Q with the best value (90% with the best, 10% with other Qs)
2. E-soft(1-E) – selecting it oppositely to greedy (10% with best)
3. Softmax - 

https://www.amazon.com/Crash-Course-hands-introduction-reinforcement-ebook/dp/B07Z9DC228

https://docs.google.com/document/d/1tuqZVDI7hLM5jHx8se2AB26G0qIXZ3WgKPw293D7IcI/edit