Reinforcement Learning

Sauman Das

Overview

- Basic Intuition
- Atari Breakout Game
- Define Terms (i.e. state, action, reward)
- Value Learning
- Policy Learning
- Simulation Environments

What was provided in supervised learning?

Supervised Learning

- Dataset
  - (x, y) pairs
  - Take a long time to create
- Goal: Minimize the loss function

Reinforcement Learning

- No more dataset!
  - Much more applicable in the real-world
- What are we trying to learn?
  - Determine which **action** to take at a particular **state**
- Optimizing based off of **reward function**

Automated Video Game Decisions - Atari Breakout

- Atari is a basic game with a few possible moves
- While watching the game, determine…
  - Possible actions
  - Ways to measure reward

Reinforcement Learning

- No more dataset!
  - Much more applicable in the real-world
- What are we trying to learn?
  - Determine which **action** to take at a particular **state**
- Optimizing based off of **reward function**

Atari Actions, States, and Rewards

- Actions
  -
  -
  -
- State:
- Reward:

Actions, States, and Rewards

Reward Function

$$R_t = \sum_{i=t}^{\infty} \gamma^i r_i$$

What do we need in order to calculate the total reward?

Q Function

- Represents the expected ***long-term*** reward given a certain state and action.

$$Q(s_t, a_t) = \mathbb{E}[R_t | s_t, a_t]$$

Goal of RL: Determine a Policy

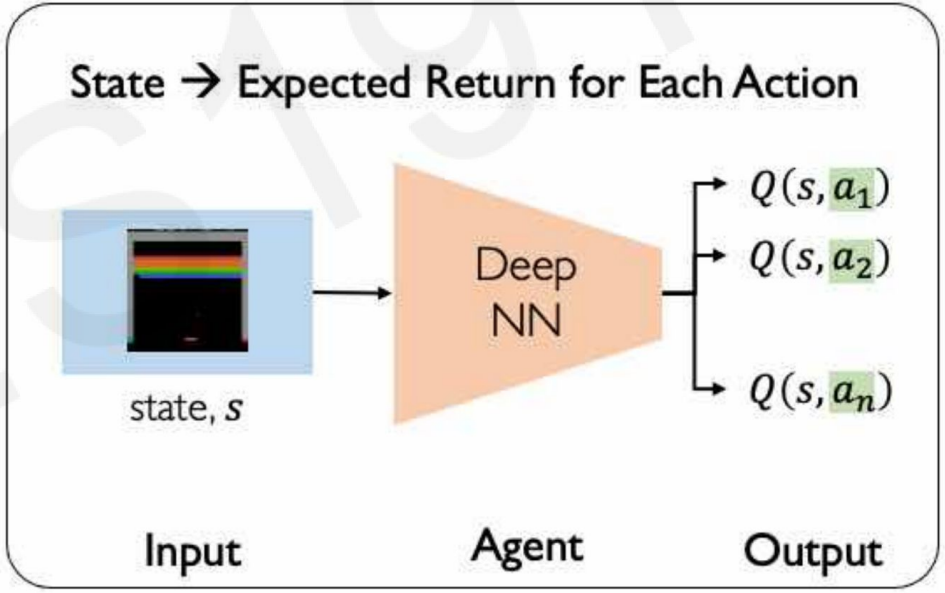
- Policy: The optimal action at a certain state
- Represent by the *symbol* π

$$\pi^*(s) = \underset{a}{\operatorname{argmax}}\, Q(s, a)$$

Value Learning

Finding Q for Each Action

- Pass in Current State through Neural Network
- Output: Q value for each action
- **How do we determine optimal action?**



Training a Deep Q Network

- Loss Function
- Target is known as the Bellman Optimality Equation → approximates reward if all the best actions were taken

$$\mathcal{L} = \mathbb{E}\left[\left\|\overbrace{\left(r + \gamma \max_{a'} Q(s', a')\right)}^{\text{target}} - \overbrace{Q(s, a)}^{\text{predicted}}\right\|^2\right]$$

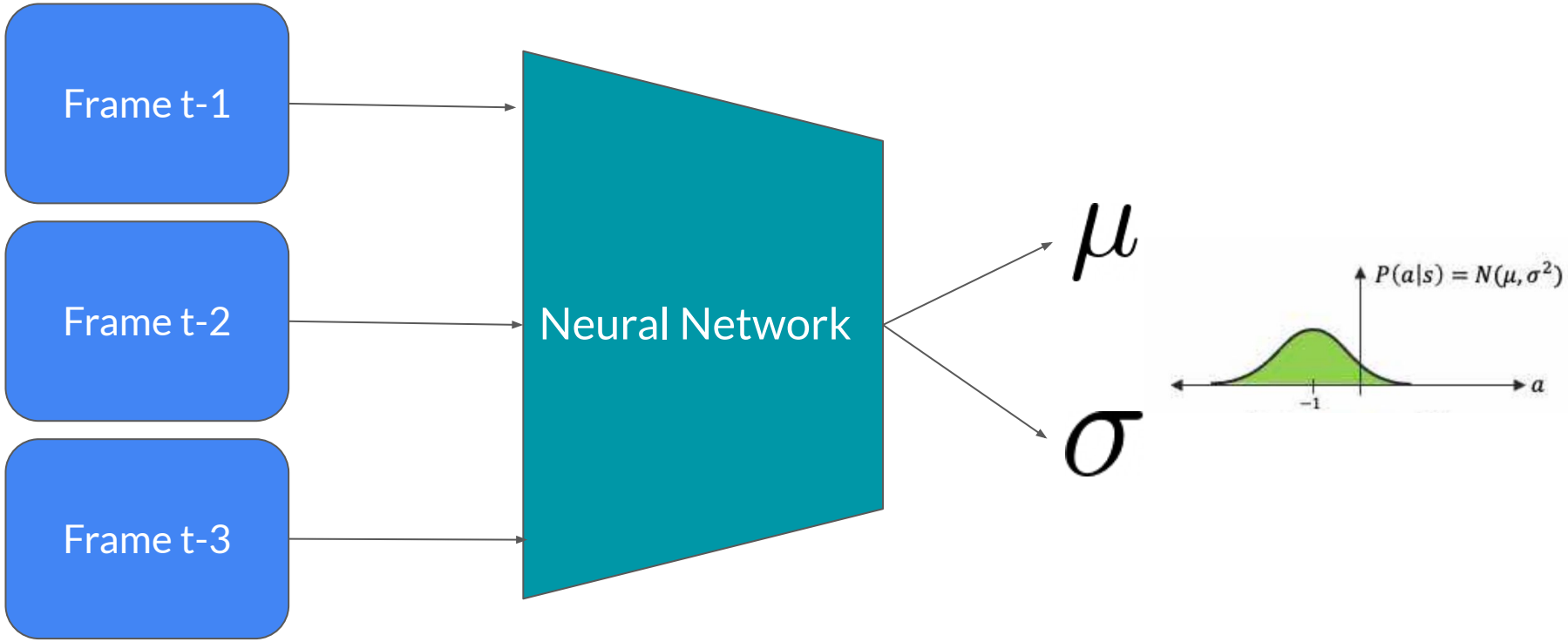
2 Hours Training

4 Hours Training

Another Game: Rider

What are the actions in Rider?

Policy Network - Predicting Continuous Outputs



Environments

- Virtual Environments or Simulations are really important for RL Tasks
- Training a self-driving car in real life is infeasible

Sources

- introtodeeplearning.com
  - Highly recommend this course
  - Very Short Course taught by MIT grad students