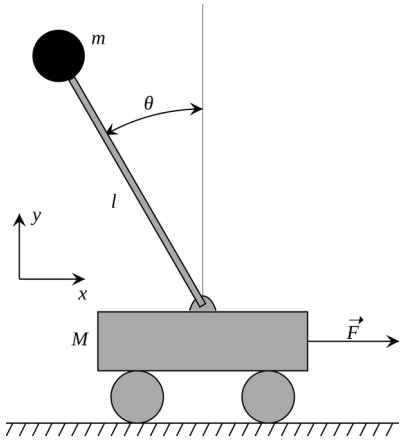
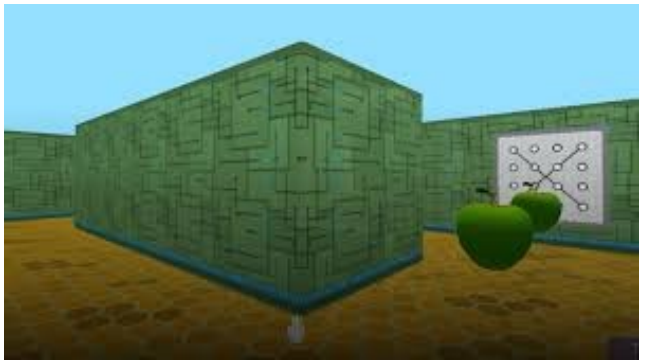
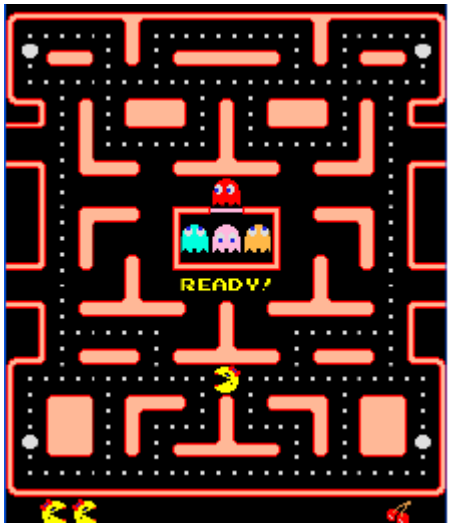
MECHTRON 2MD3
Data Structures and Algorithms for Mechatronics

Application of Trees:
Genetic Programming in Reinforcement Learning







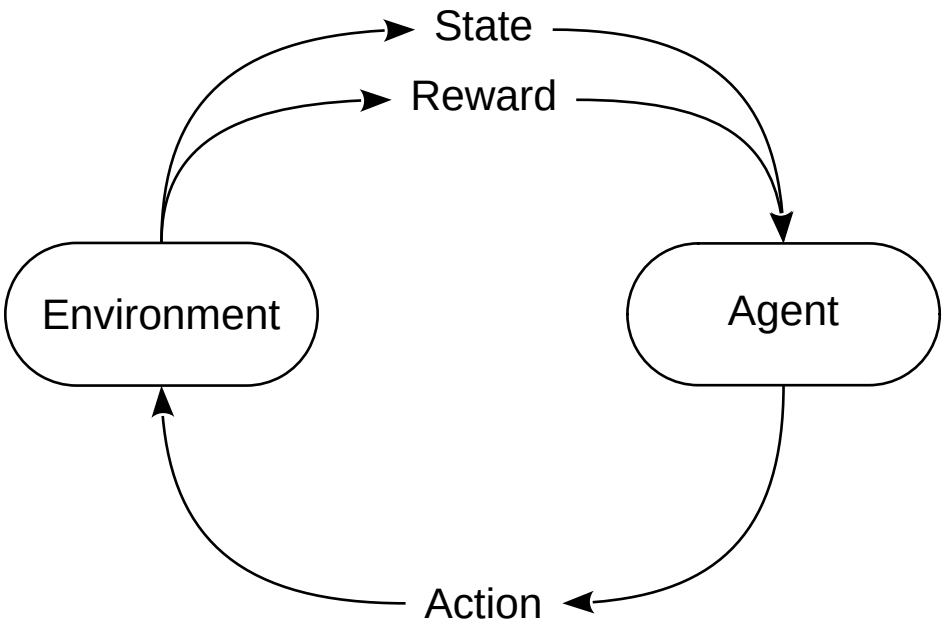
Stephen Kelly, PhD
Department of Computing and Software
kellys32@mcmaster.ca

Genetic Programming in Reinforcement Learning

Overview

- Reinforcement Learning (RL) characteristics
- Crash course in Genetic Programming with classic control RL example
- Assignment Overview

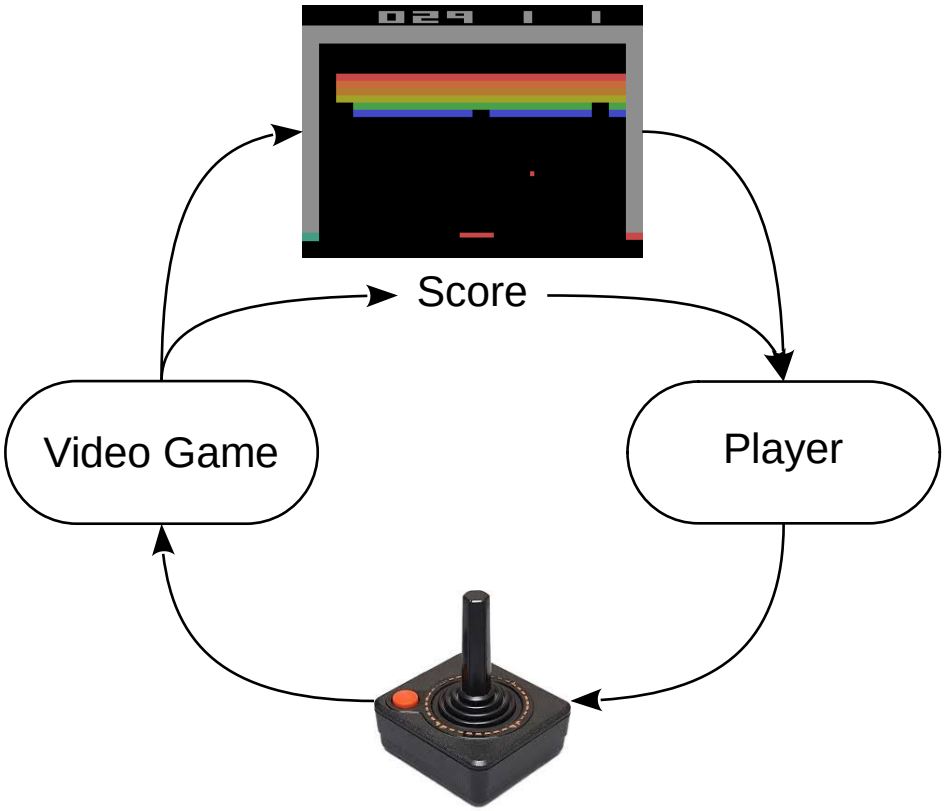
Reinforcement Learning



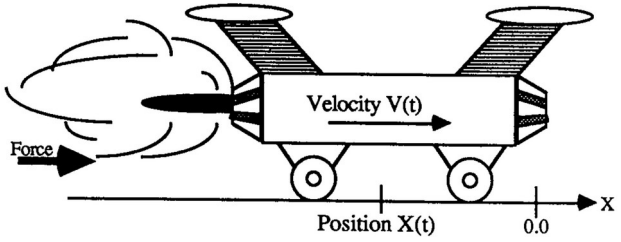
Agent's behavior adapted through trial-and-error interaction with environment

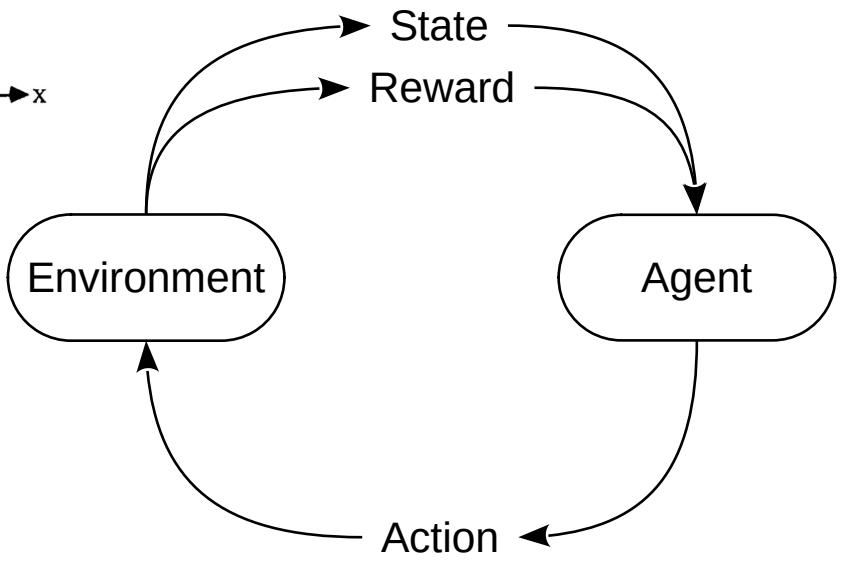
Goal: Agent maps states (situations) to actions such that reward is maximized

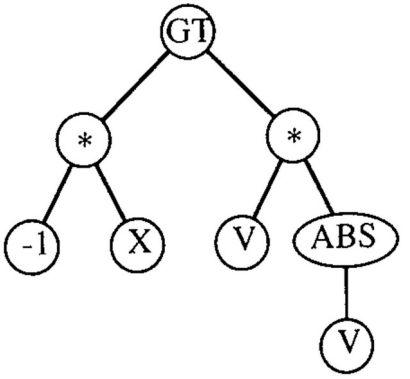
Reinforcement Learning



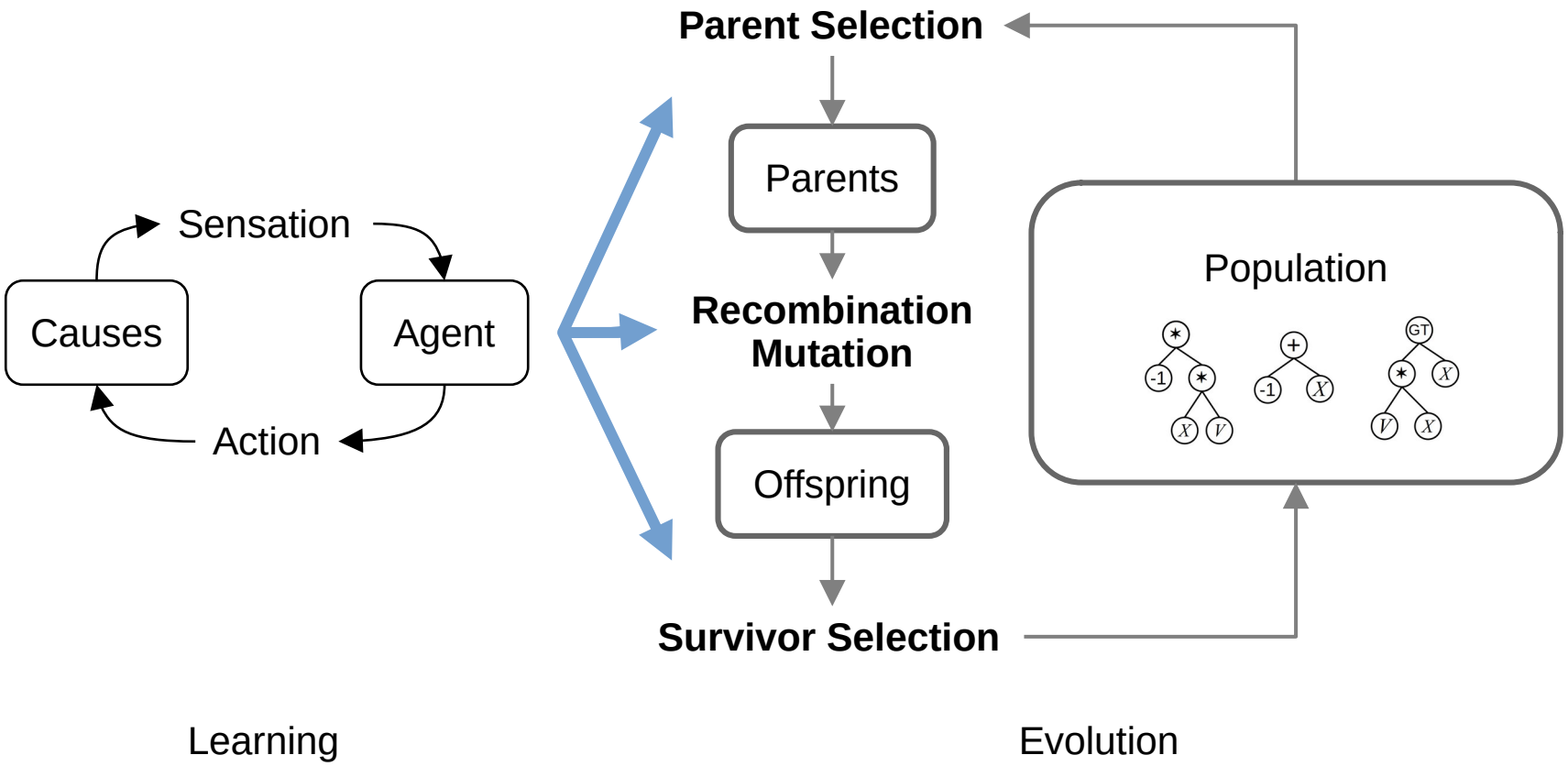
Genetic Programming in Reinforcement Learning







Evolutionary Reinforcement Learning



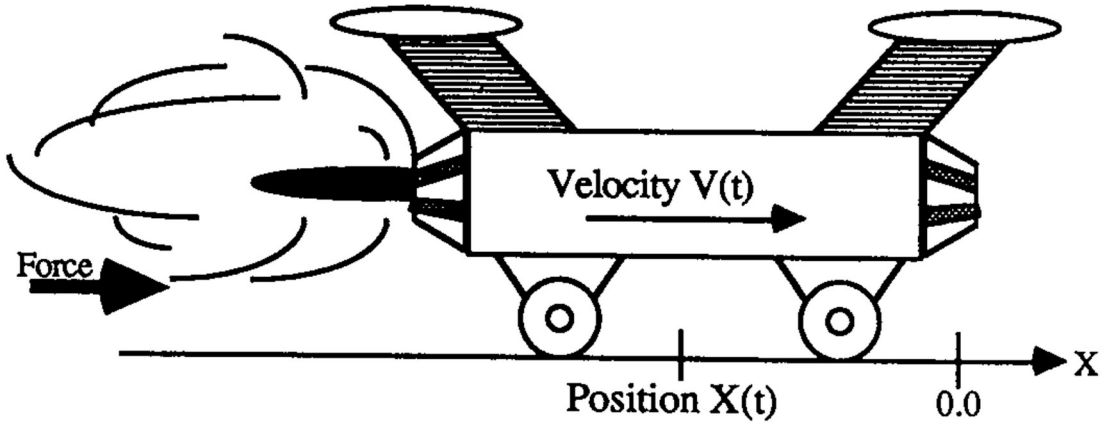
<u>Adaptation at two timescales</u>

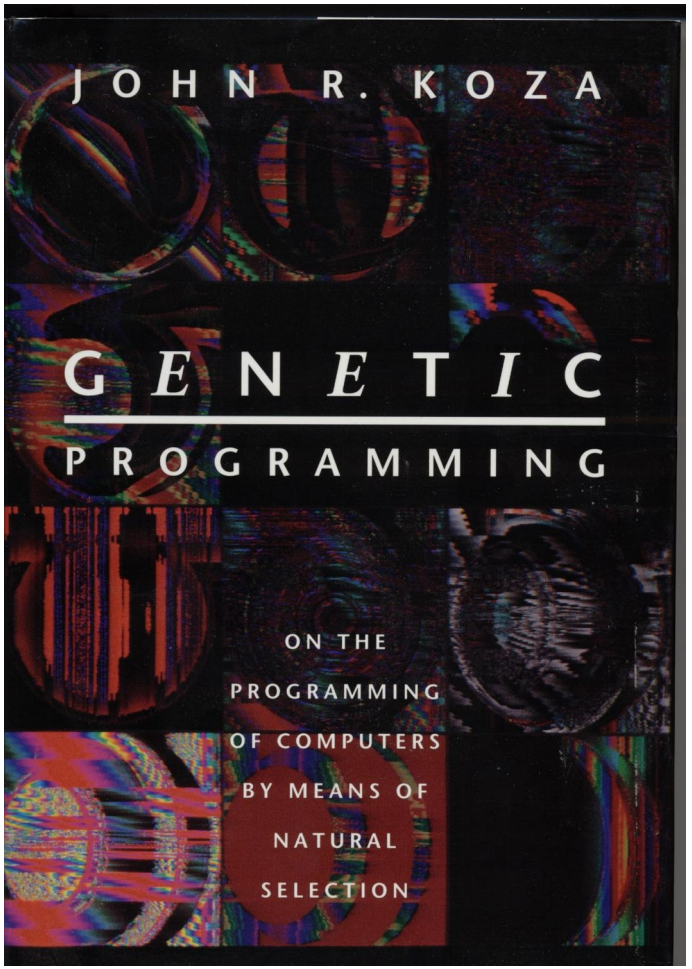
Learning: Agent learns to predict the effect of its actions

Evolution: Agent's experience influences evolutionary operators (selection and variation)

Optimal Control

Cart Centering (isotropic rocket) Problem



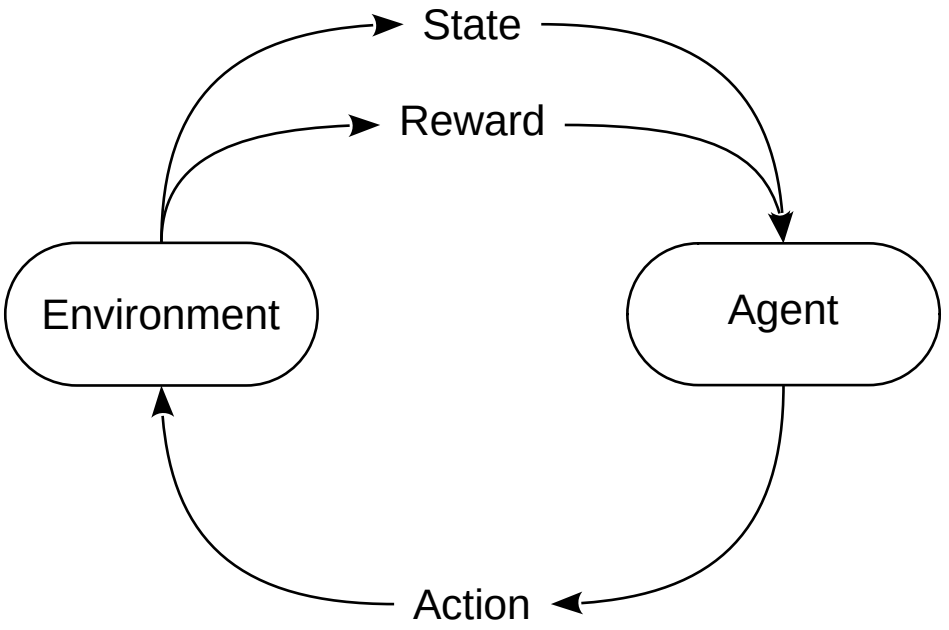


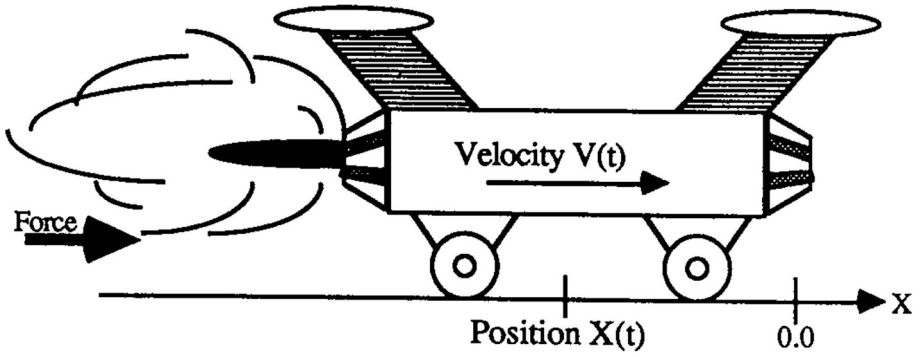
J. R. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection. Cambridge, MA: MIT Press, 1992

Optimal Control

Cart Centering (isotropic rocket) Problem

Cart can move left or right on frictionless 1-dimensional track



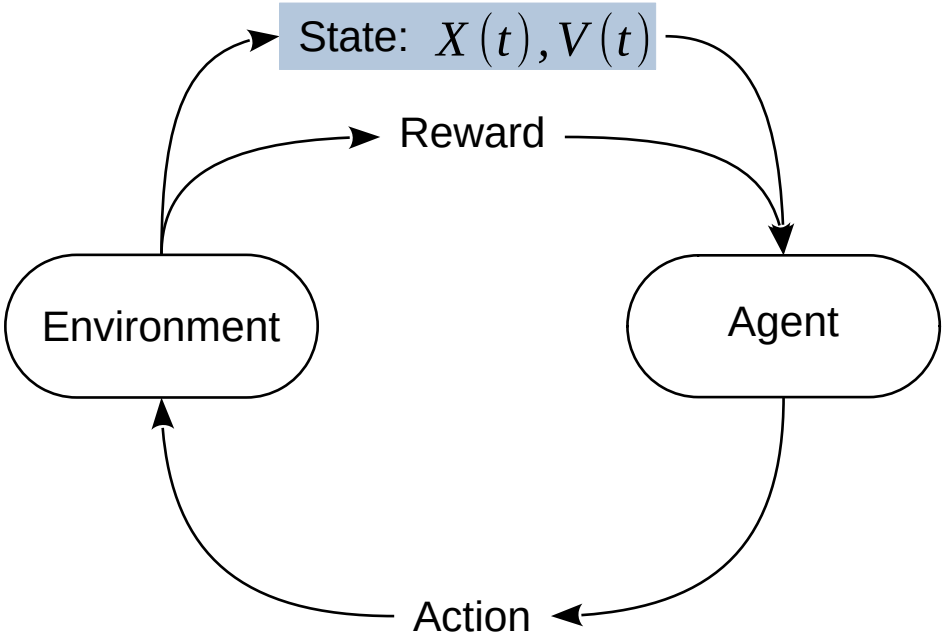


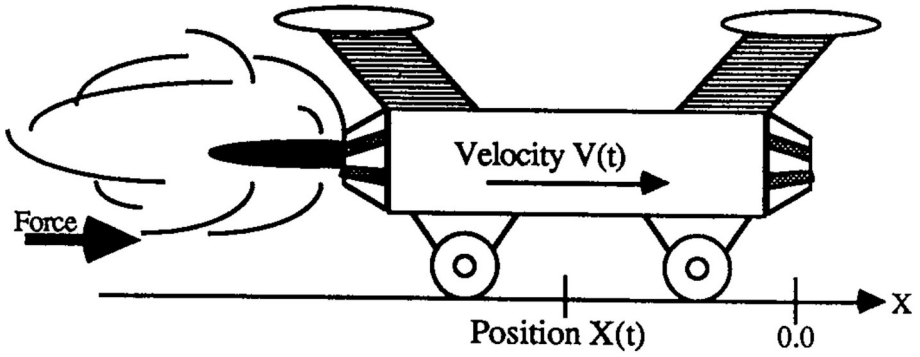
Optimal Control

Cart Centering (isotropic rocket) Problem

Cart can move left or right on frictionless 1-dimensional track

- **State:** X position and velocity of cart



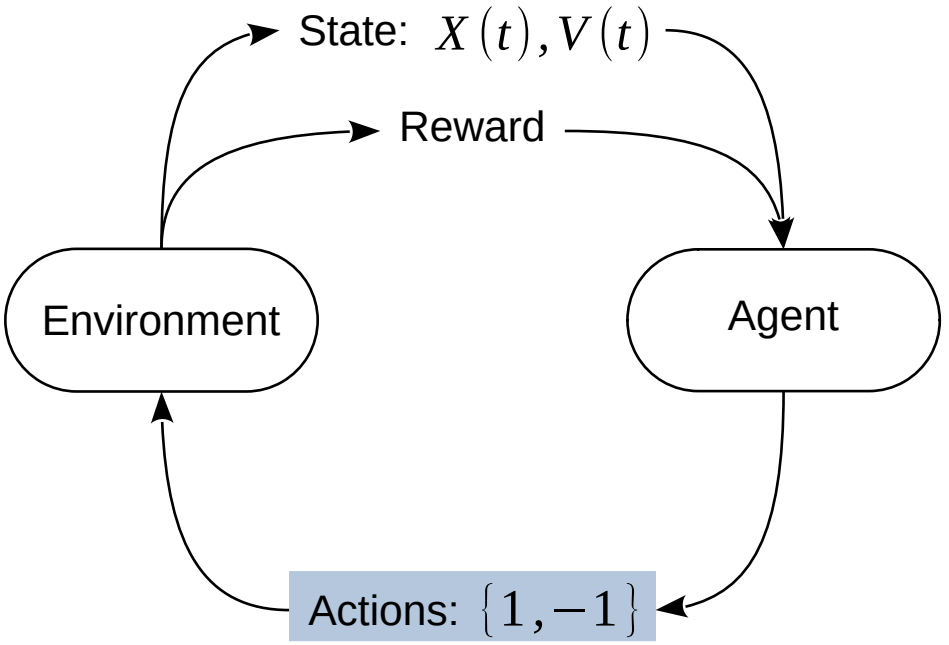


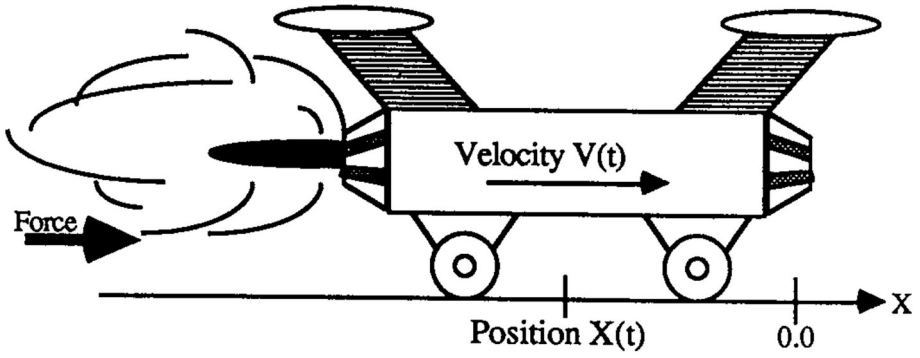
Optimal Control

Cart Centering (isotropic rocket) Problem

Cart can move left or right on frictionless 1-dimensional track

- **State:** X position and velocity of cart
- **Action:** Force to accelerate cart left or right



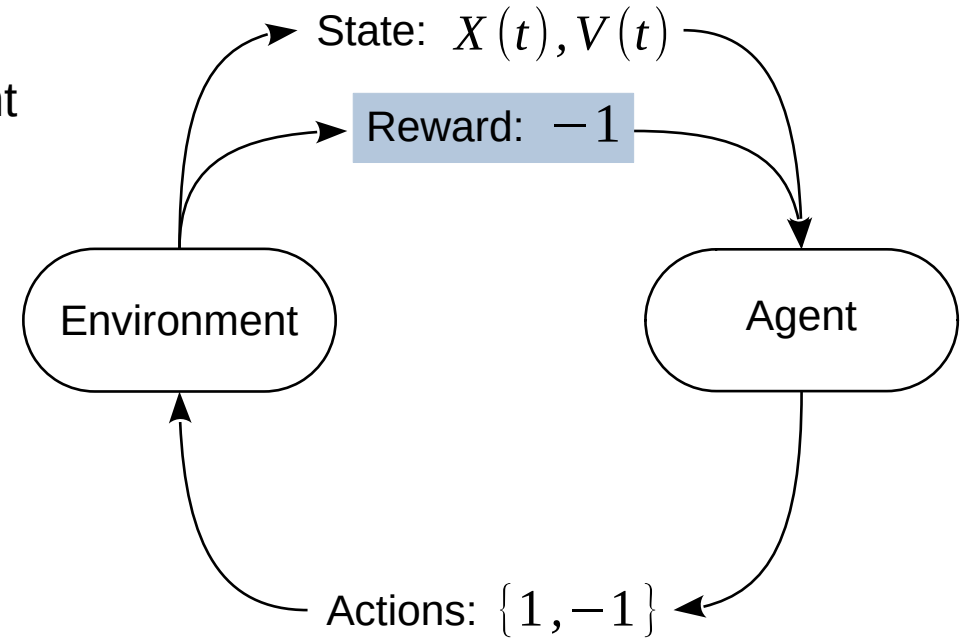


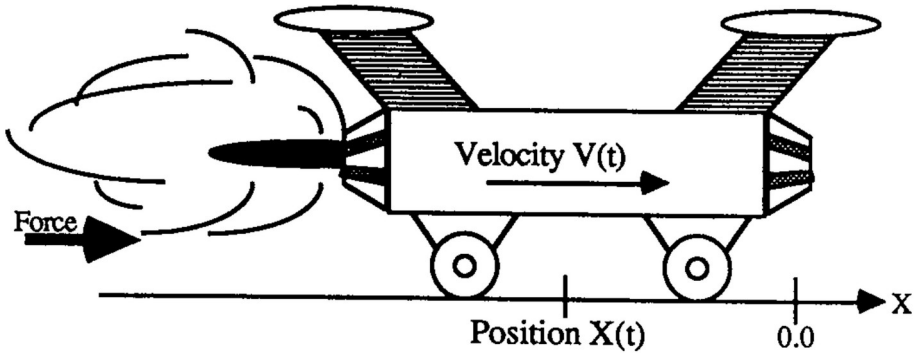
Optimal Control

Cart Centering (isotropic rocket) Problem

Cart can move left or right on frictionless 1-dimensional track

- **State:** X position and velocity of cart
- **Action:** Force to accelerate cart left or right
- **Reward:** -1



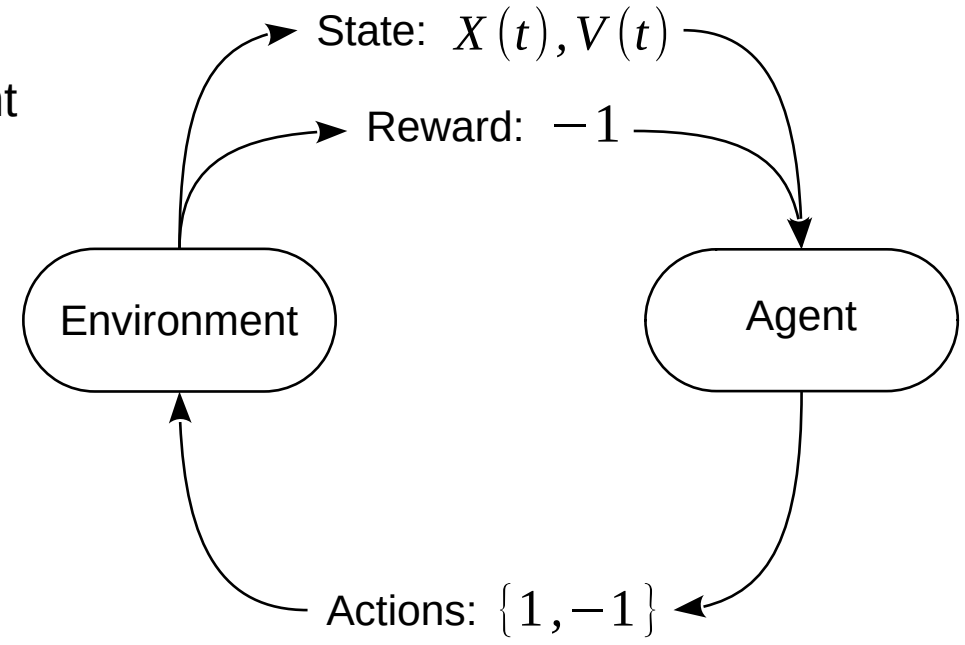


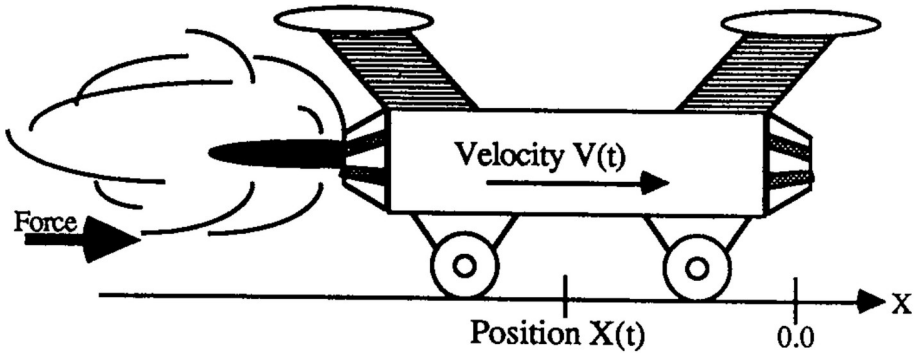
Optimal Control

Cart Centering (isotropic rocket) Problem

Cart can move left or right on frictionless 1-dimensional track

- **State:** X position and velocity of cart
- **Action:** Force to accelerate cart left or right
- **Reward:** -1
- **Initial State:** Random in (-0.75, 0.75)



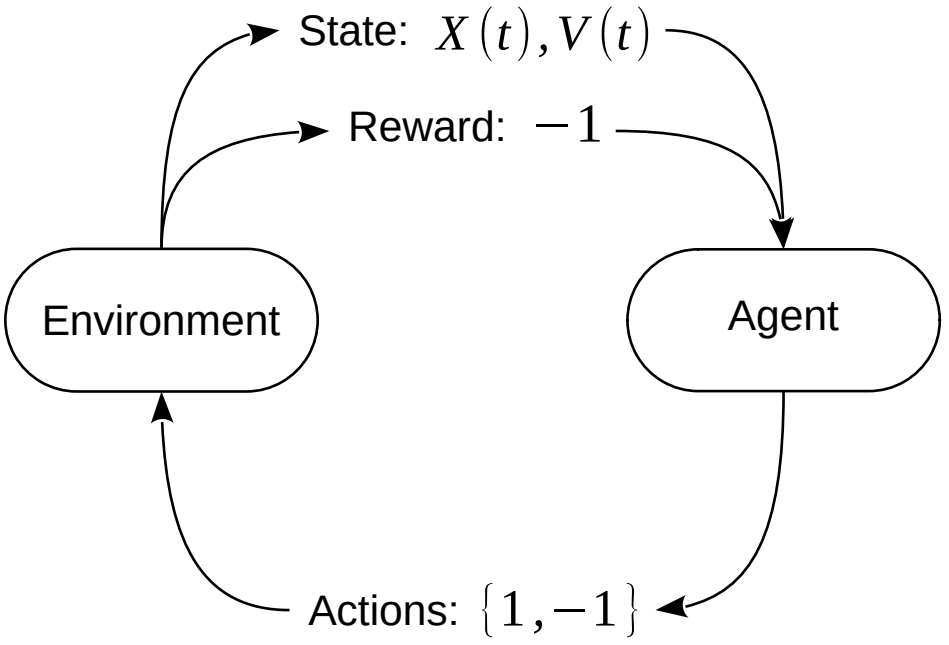


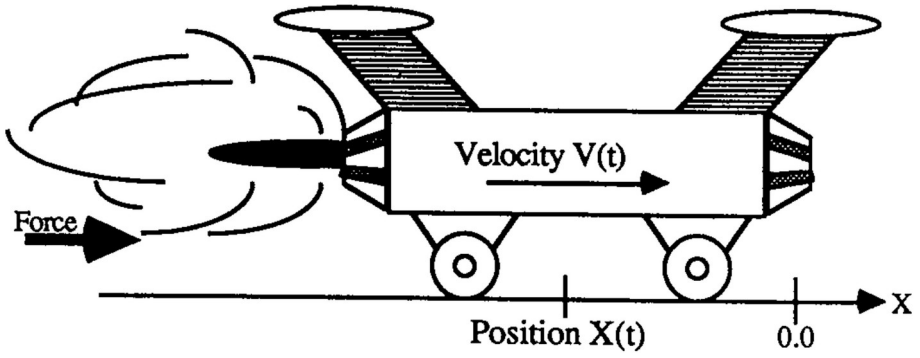
Optimal Control

Cart Centering (isotropic rocket) Problem

Cart can move left or right on frictionless 1-dimensional track

- **State:** X position and velocity of cart
- **Action:** Force to accelerate cart left or right
- **Reward:** -1
- **Initial State:** Random in (-0.75, 0.75)
- **Objective:** Center cart in minimal time



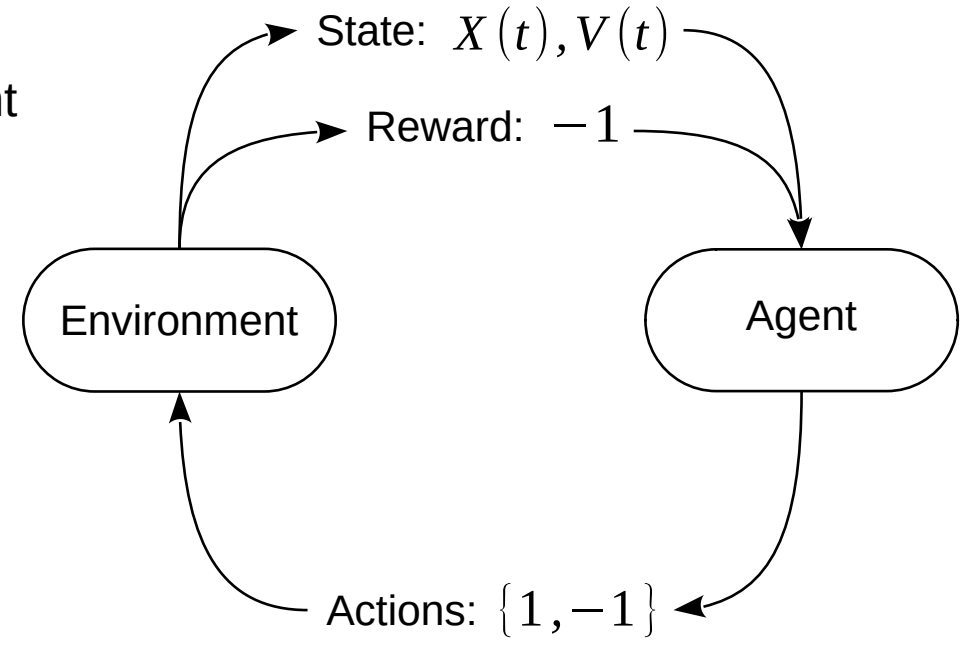


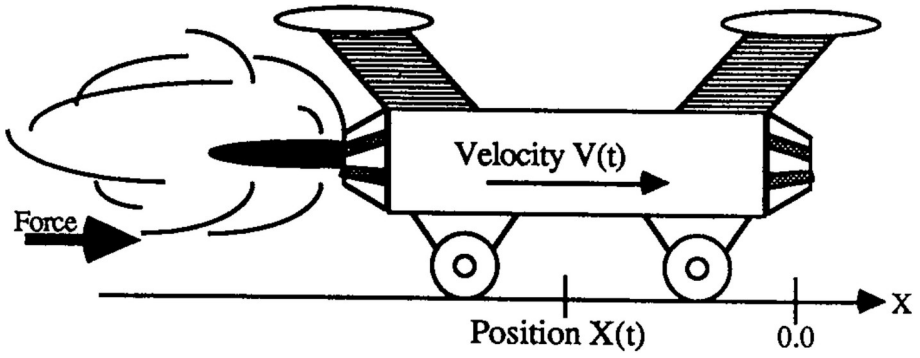
Optimal Control

Cart Centering (isotropic rocket) Problem

Cart can move left or right on frictionless 1-dimensional track

- **State:** X position and velocity of cart
- **Action:** Force to accelerate cart left or right
- **Reward:** -1
- **Initial State:** Random in (-0.75, 0.75)
- **Objective:** Center cart in minimal time
- **Target state:** $|(X(t))| \leq 0.01, |(V(t))| \leq 0.01$



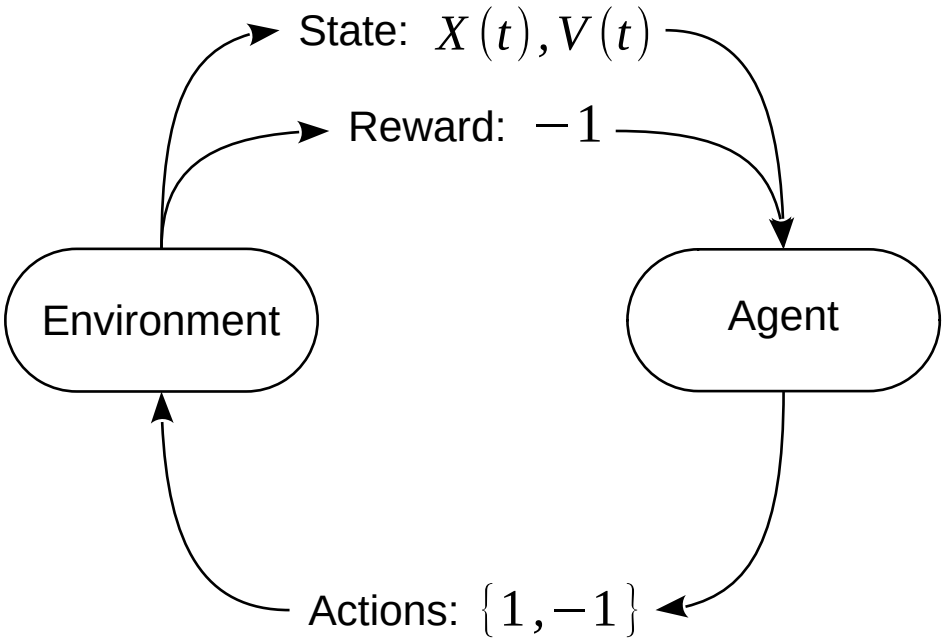


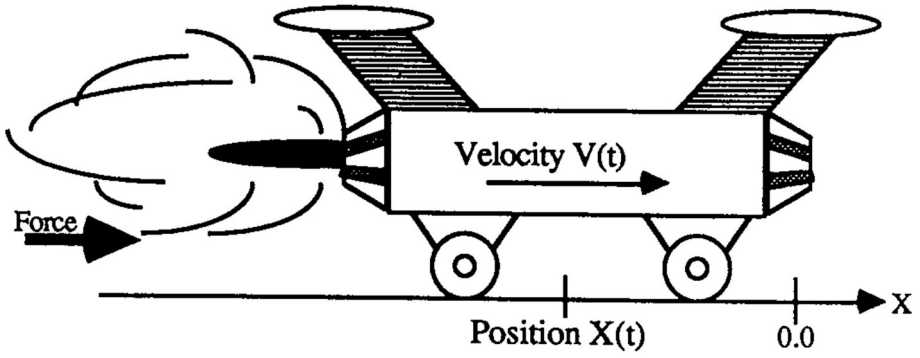
Optimal Control

Cart Centering (isotropic rocket) Problem

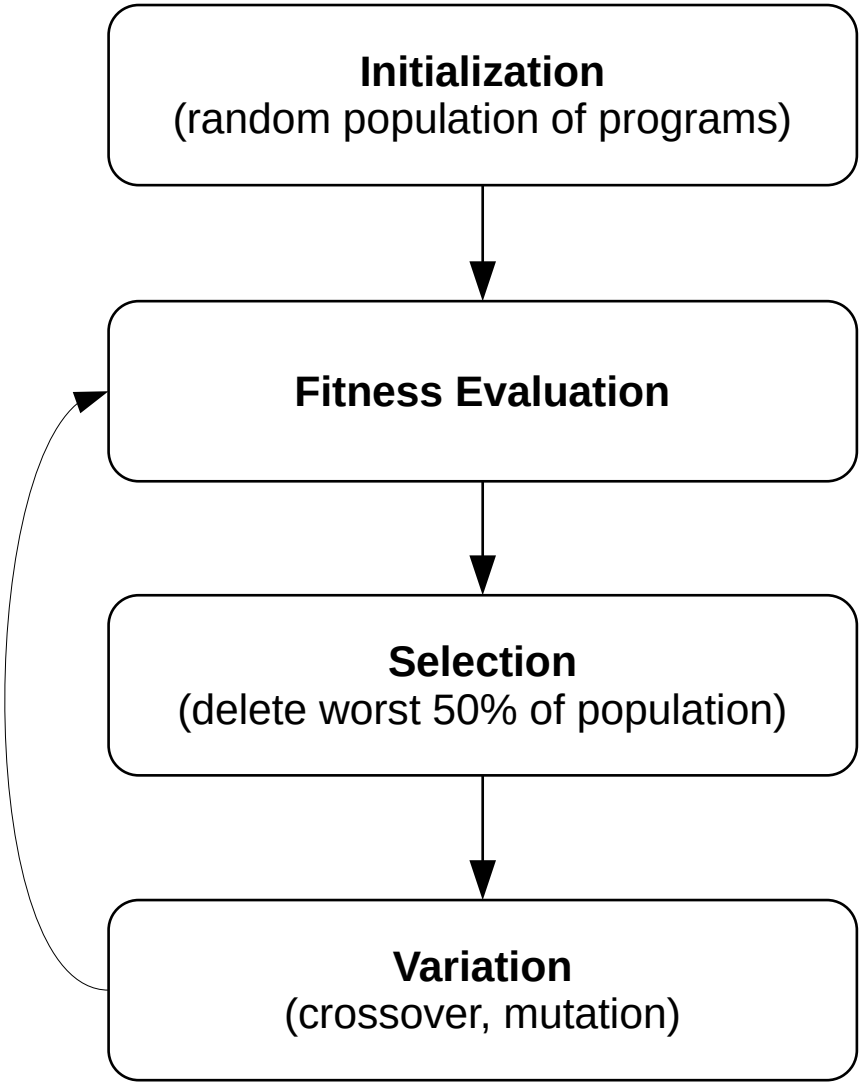
Cart can move left or right on frictionless 1-dimensional track

- **State:** X position and velocity of cart
- **Action:** Force to accelerate cart left or right
- **Reward:** -1
- **Initial State:** Random in (-0.75, 0.75)
- **Objective:** Center cart in minimal time
- **Target state:** $|(X(t))| \leq 0.01, |(V(t))| \leq 0.01$
- **Time limit:** 500 steps

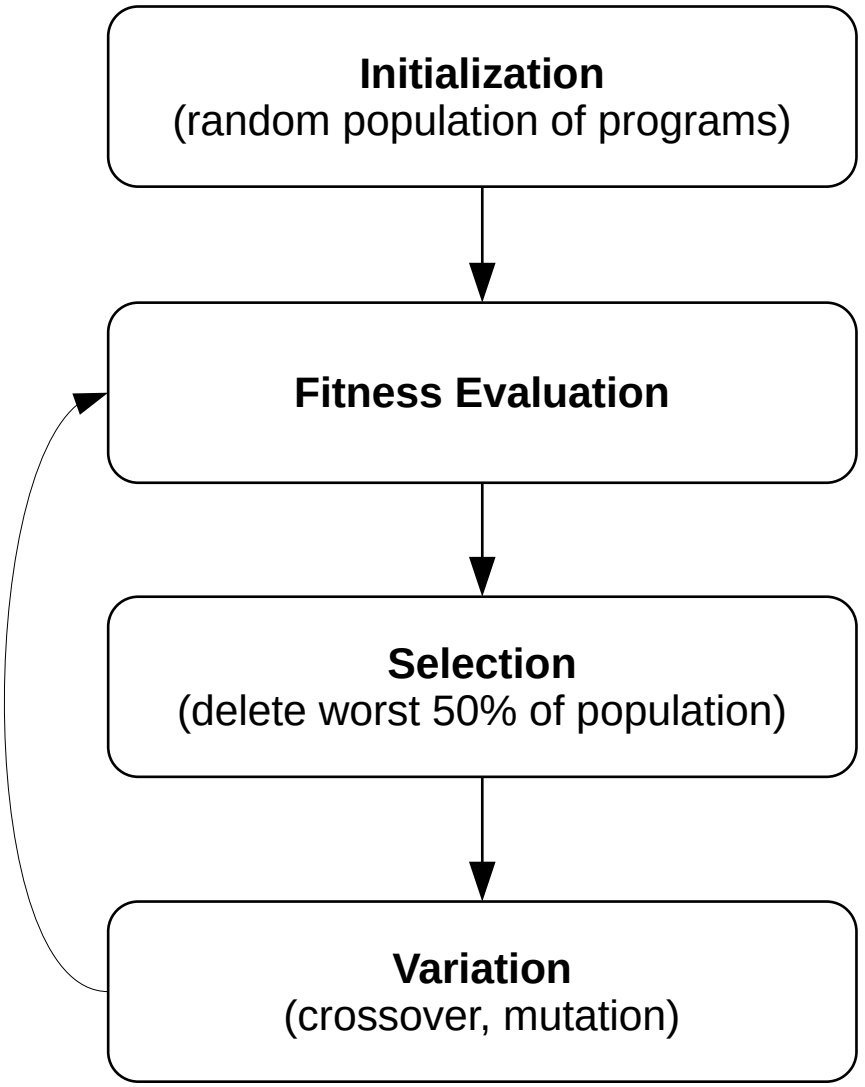


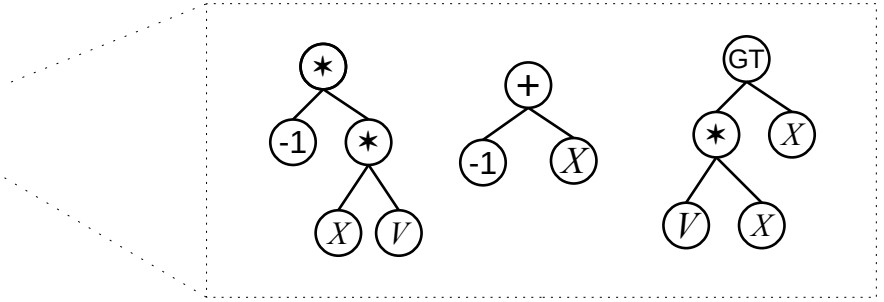


Genetic Programming



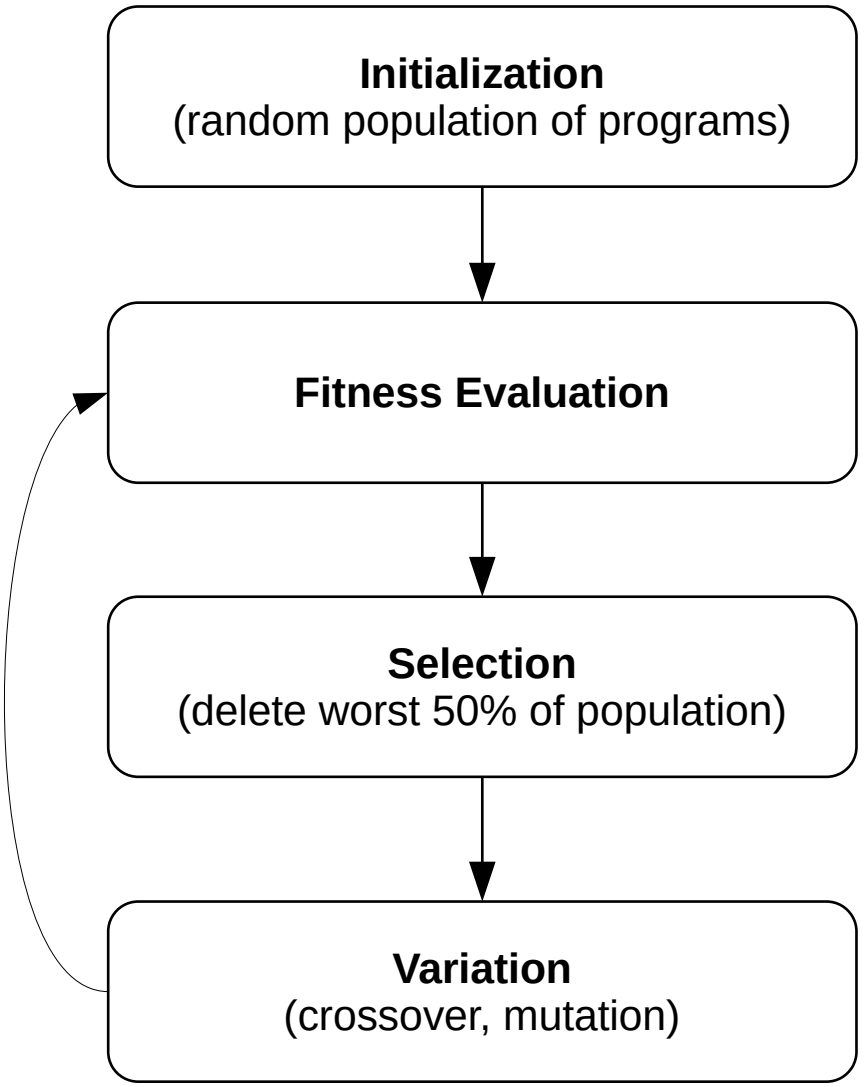
Genetic Programming





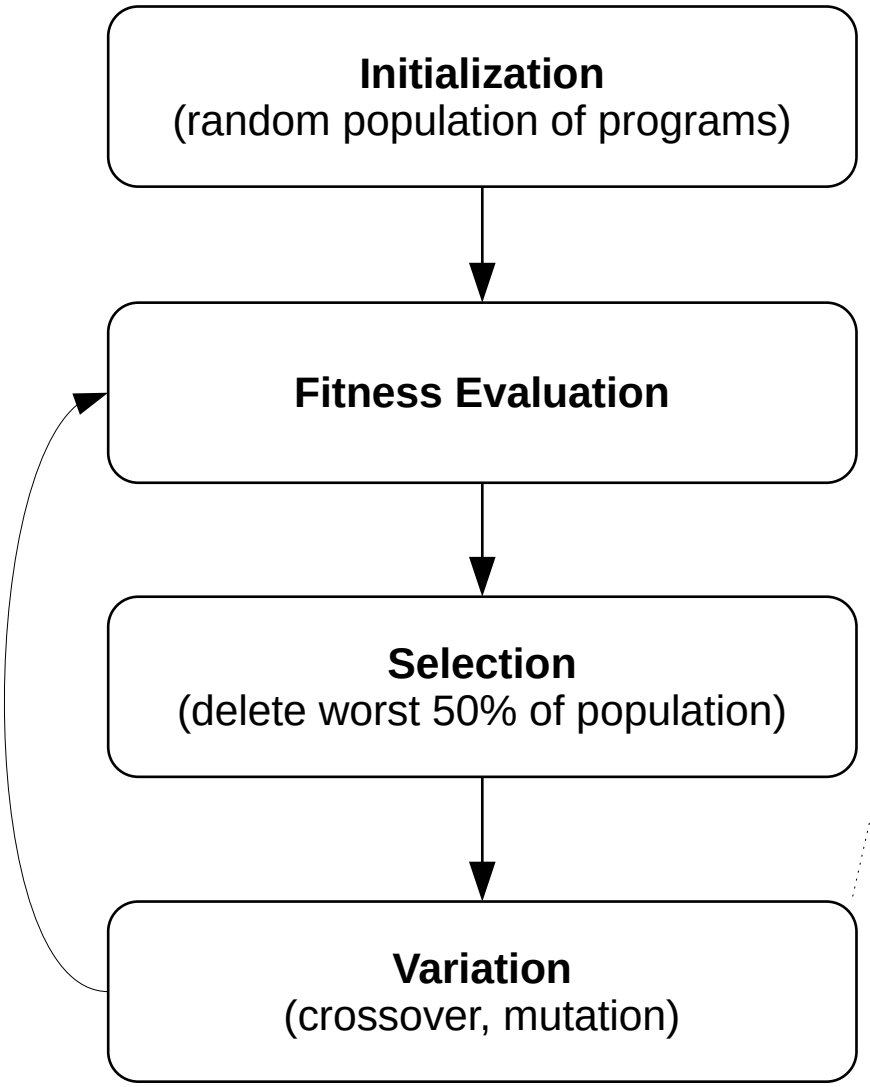
- wrapper maps positive output to 1, negative output to -1

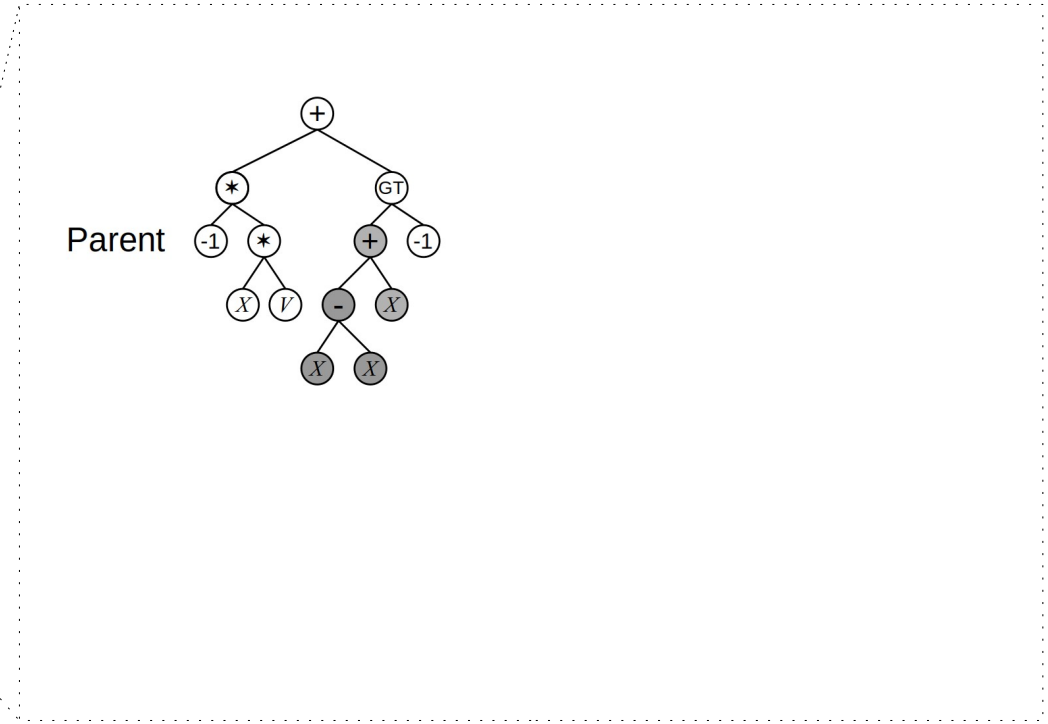
Genetic Programming



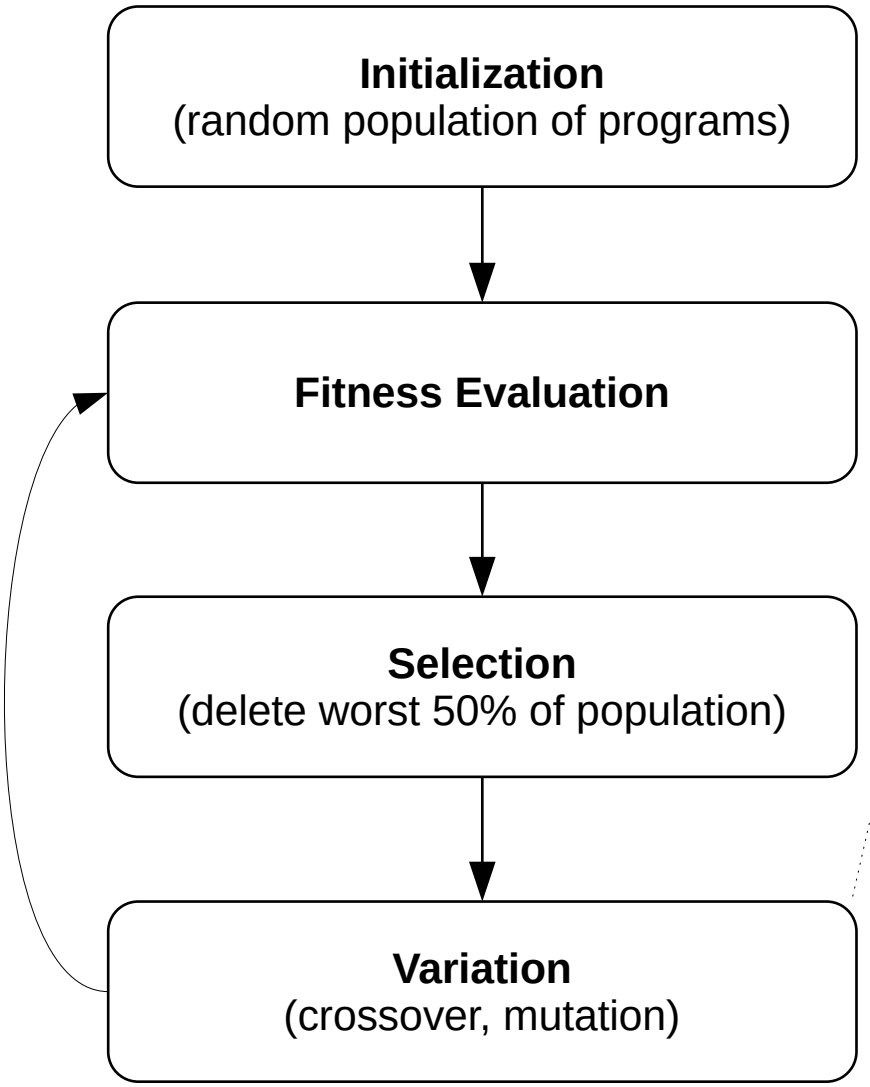
- Average total reward in 20 episodes with random initial conditions

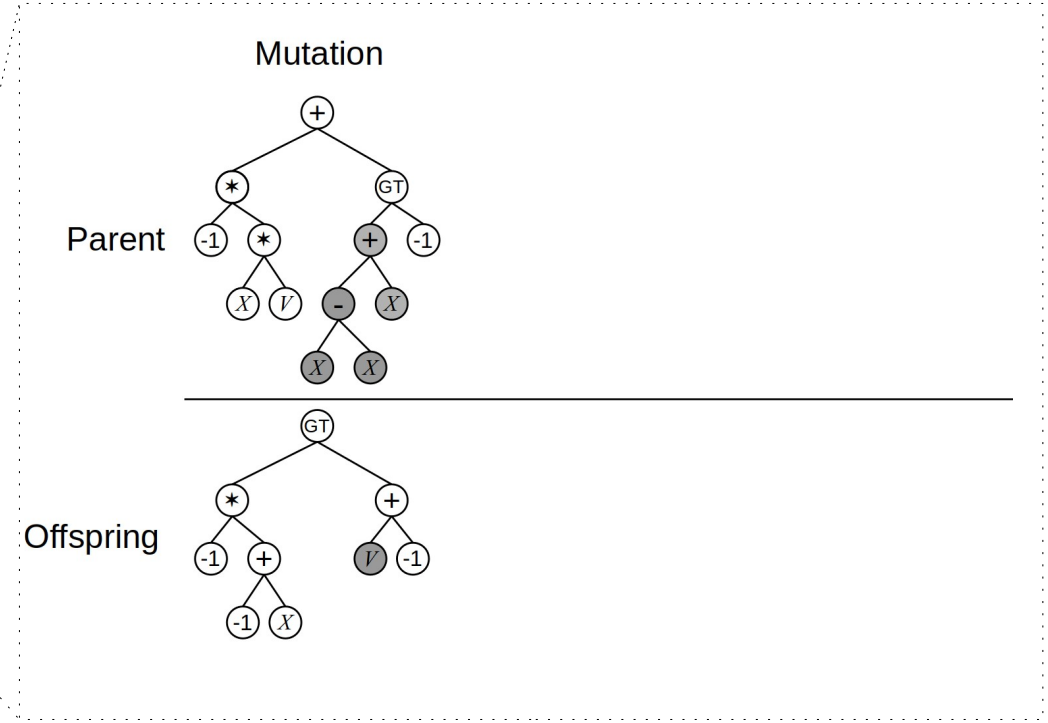
Genetic Programming



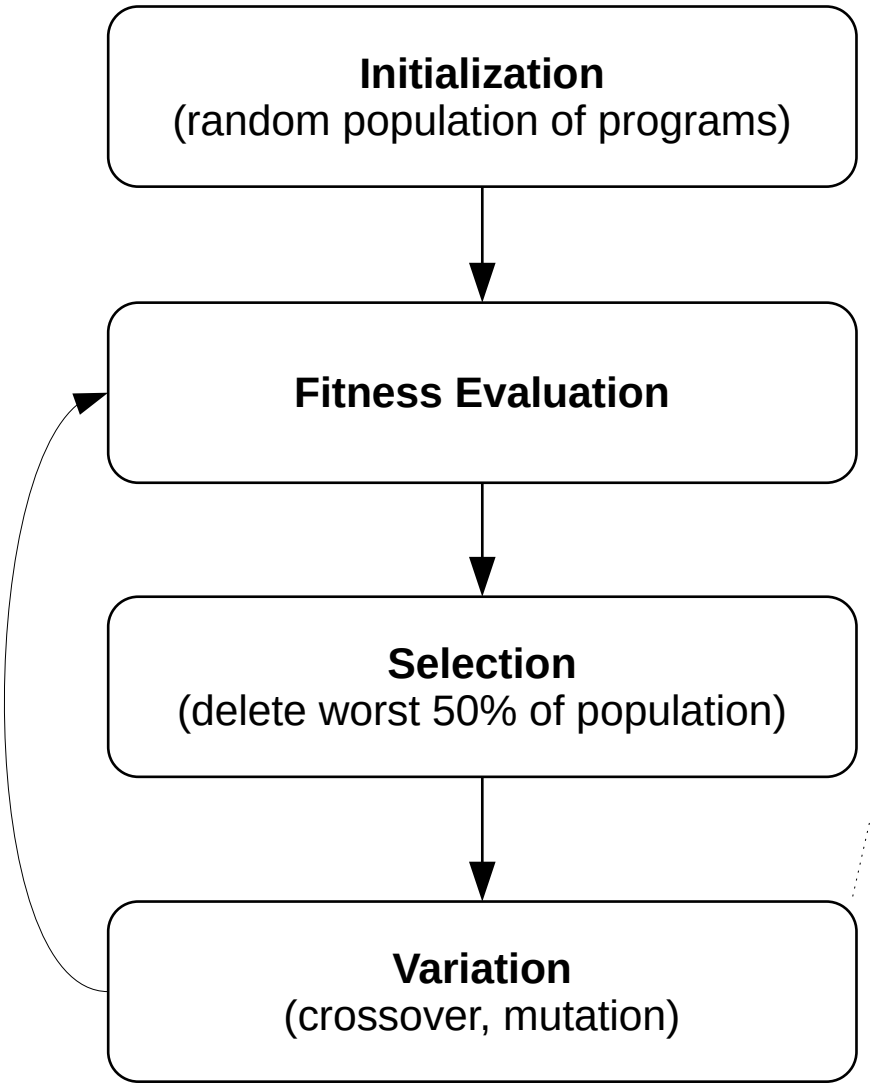


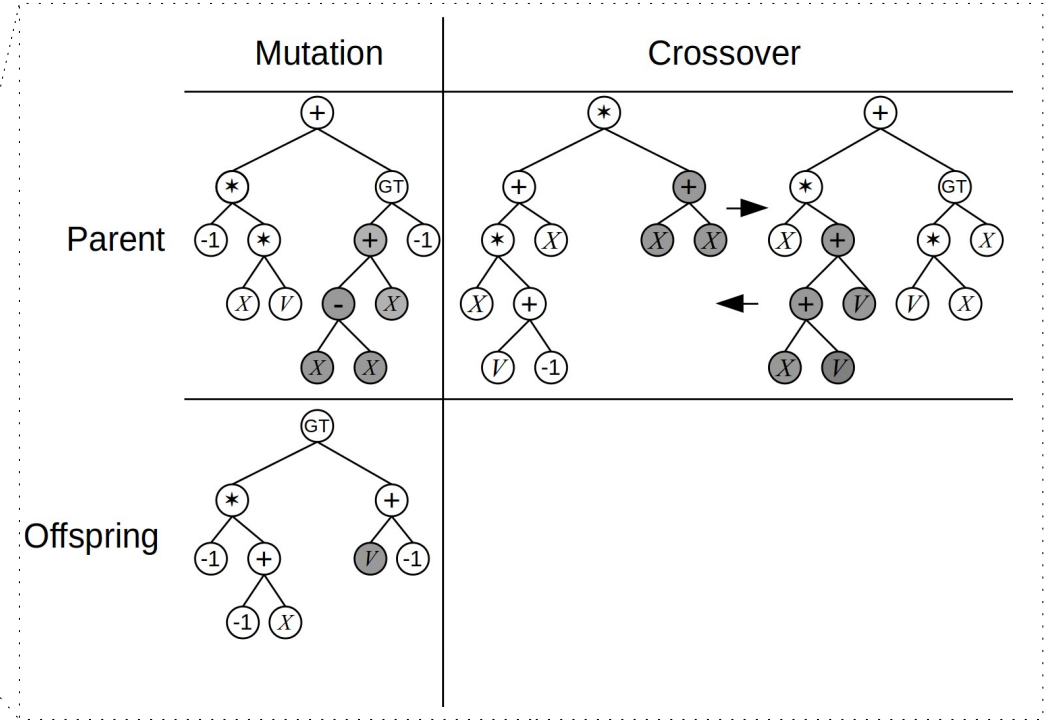
Genetic Programming



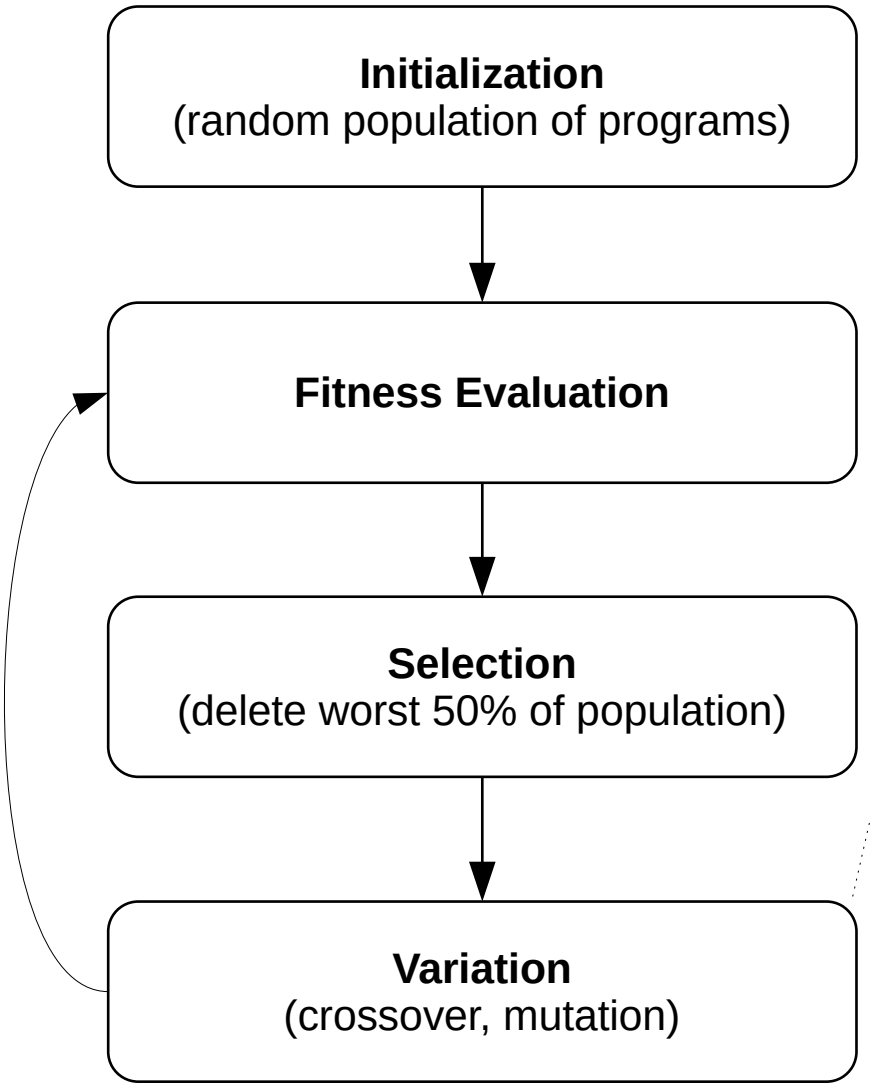


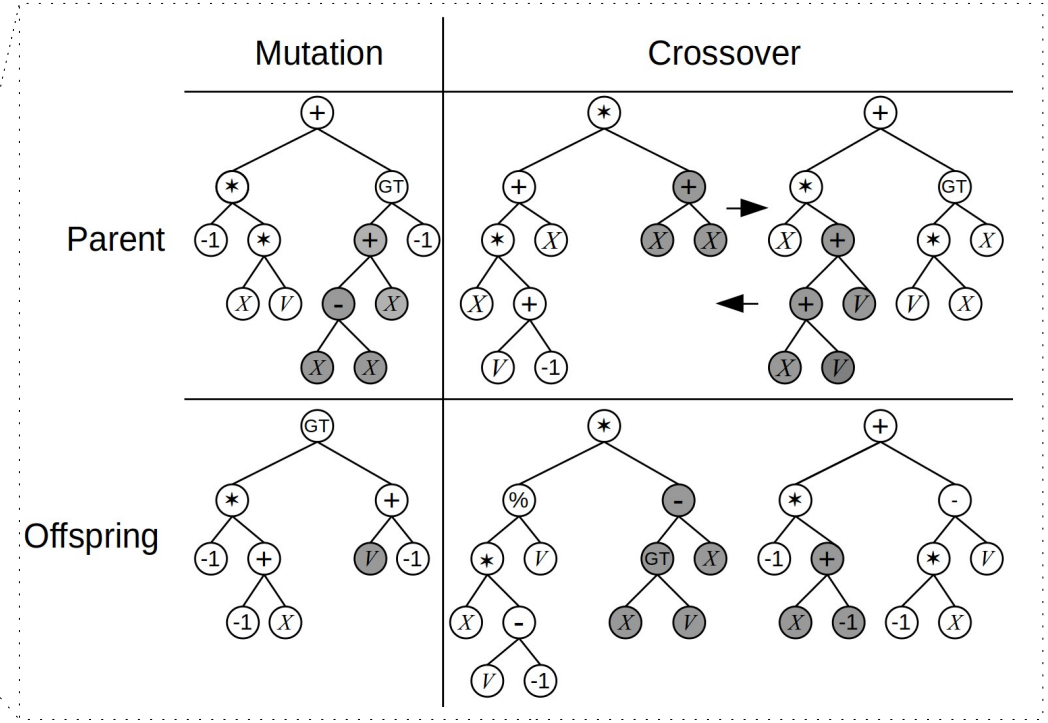
Genetic Programming



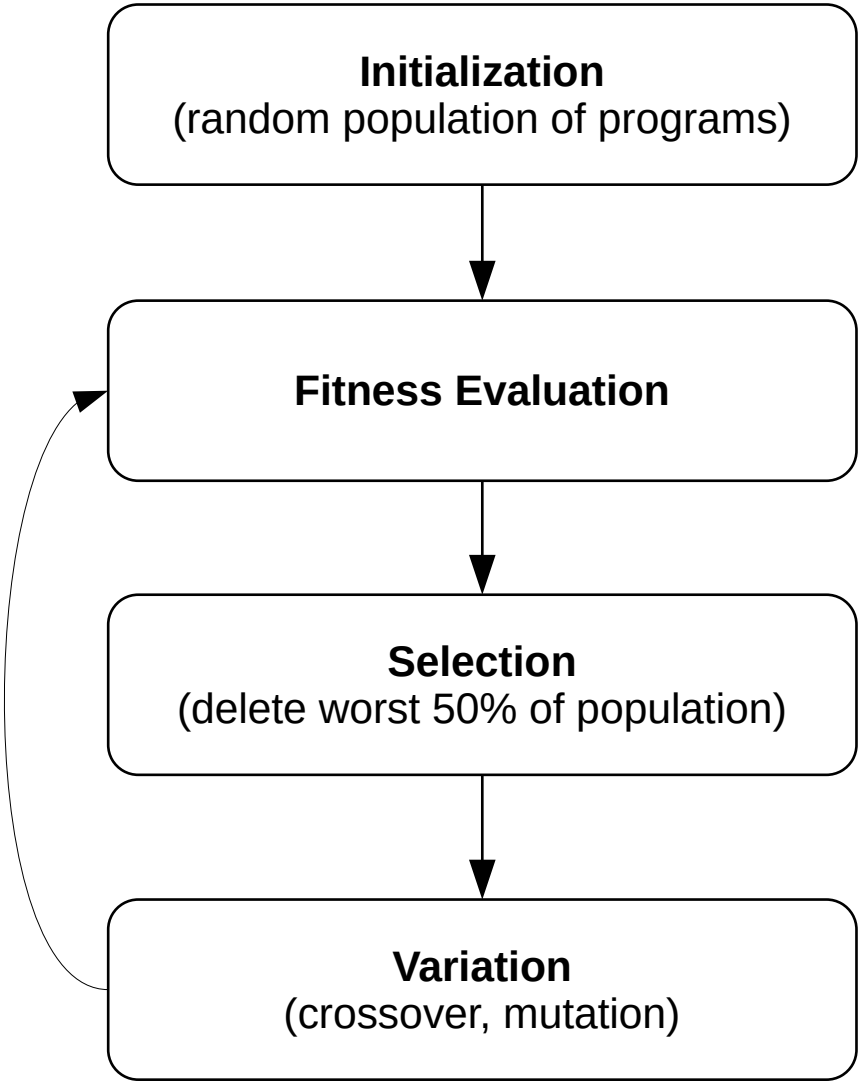


Genetic Programming





Genetic Programming



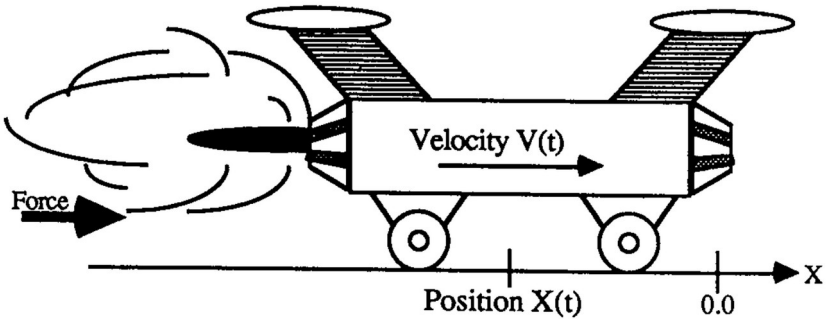
- 1 loop = 1 ***generation***

Genetic Programming Experiment

Cart Centering (isotropic rocket) Problem

Setting up a run:

- Terminal set: *X*, *V (i.e. function inputs, system state)*
- Function Set: +, -, *, %, GT, ABS
- Reward per step: -1
- Fitness measure: sum of reward averaged over 20 episodes
- Parameters: Population size 500 , 50 generations



Genetic Programming Experiment

Cart Centering (isotropic rocket) Problem