RL problems from academia and industry: learning goals and corresponding reward functions

Learning to play Chess

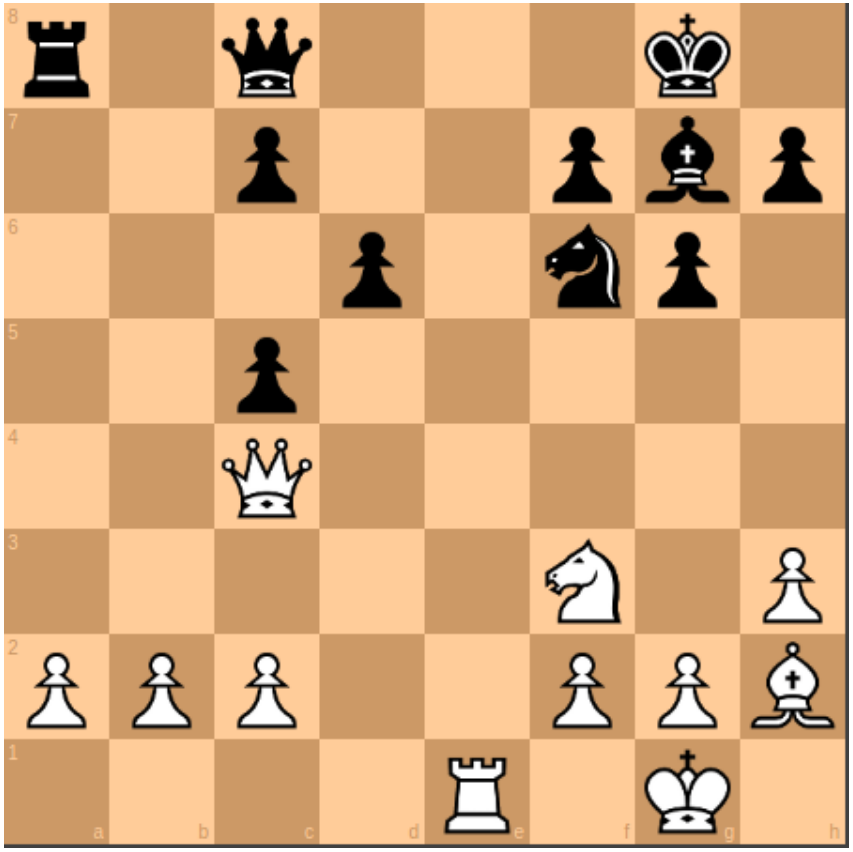


Figure: Typical chess position

Learning to play Chess

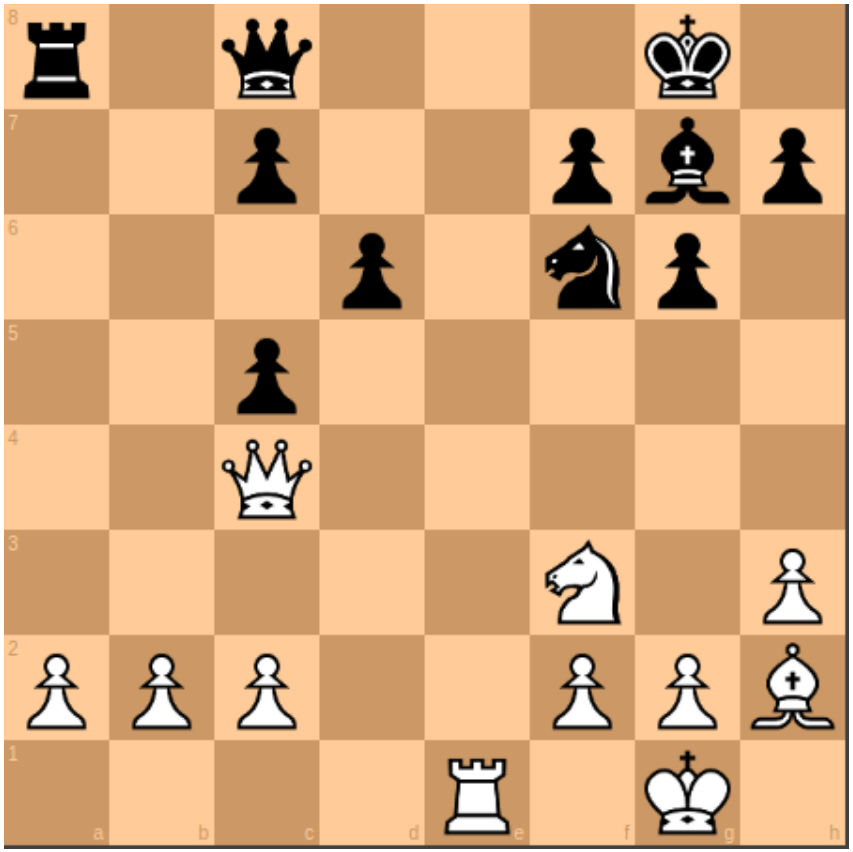


Figure: Typical chess position

Actions

- Allowed moves in a chess game

Learning to play Chess

Reward function

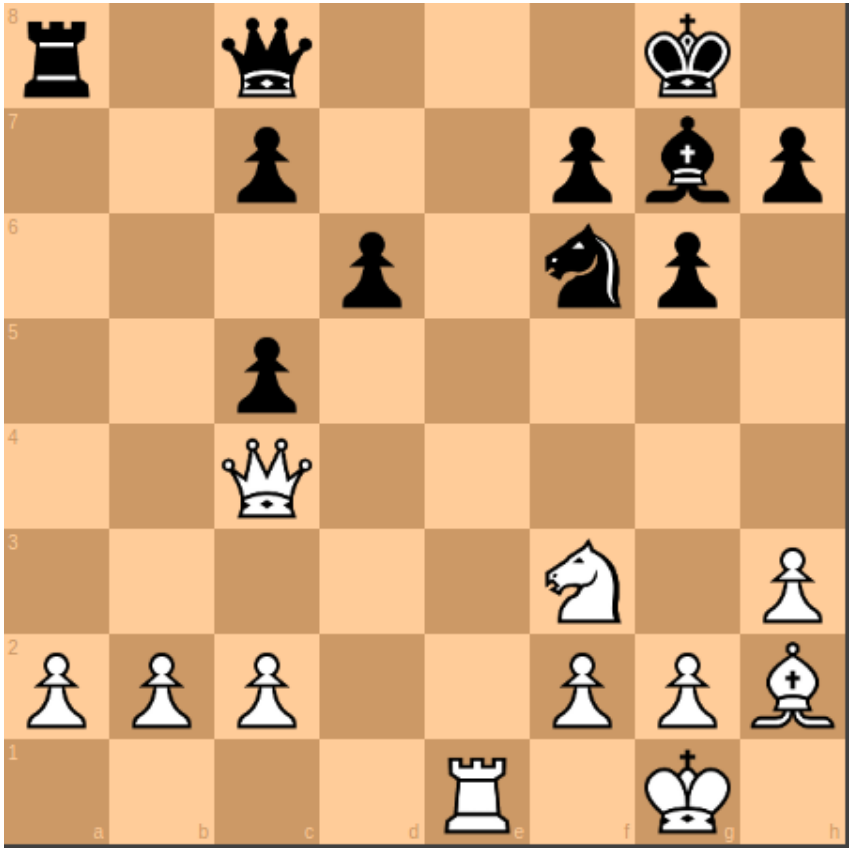


Figure: Typical chess position

Learning to play Chess

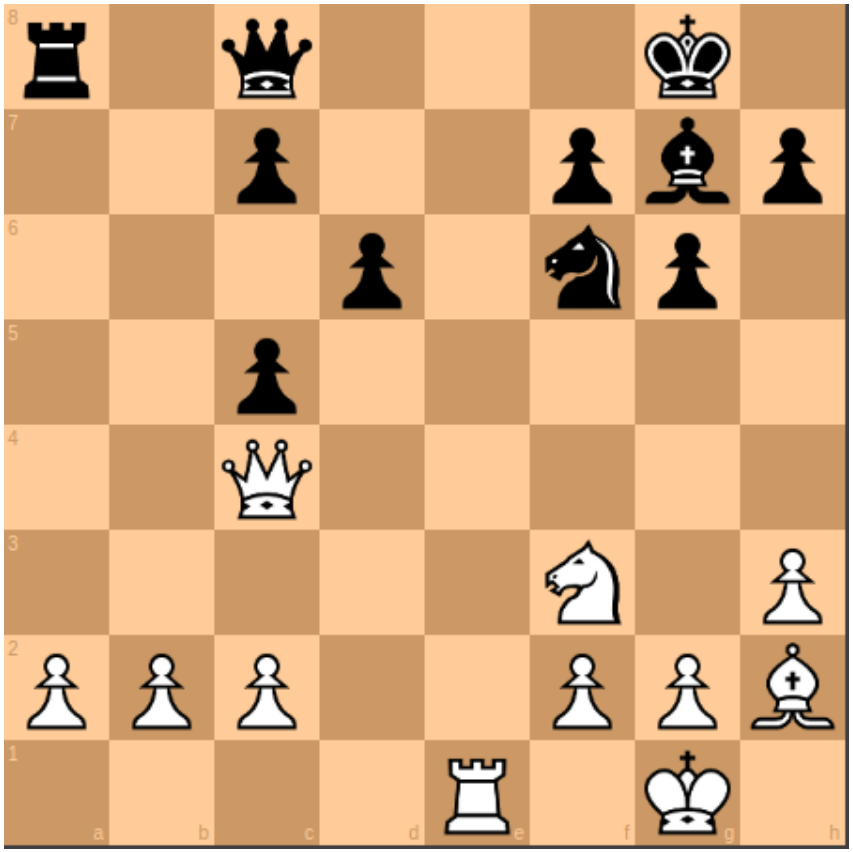


Figure: Neutral position

Reward function

- 0 when the game is ongoing

Learning to play Chess

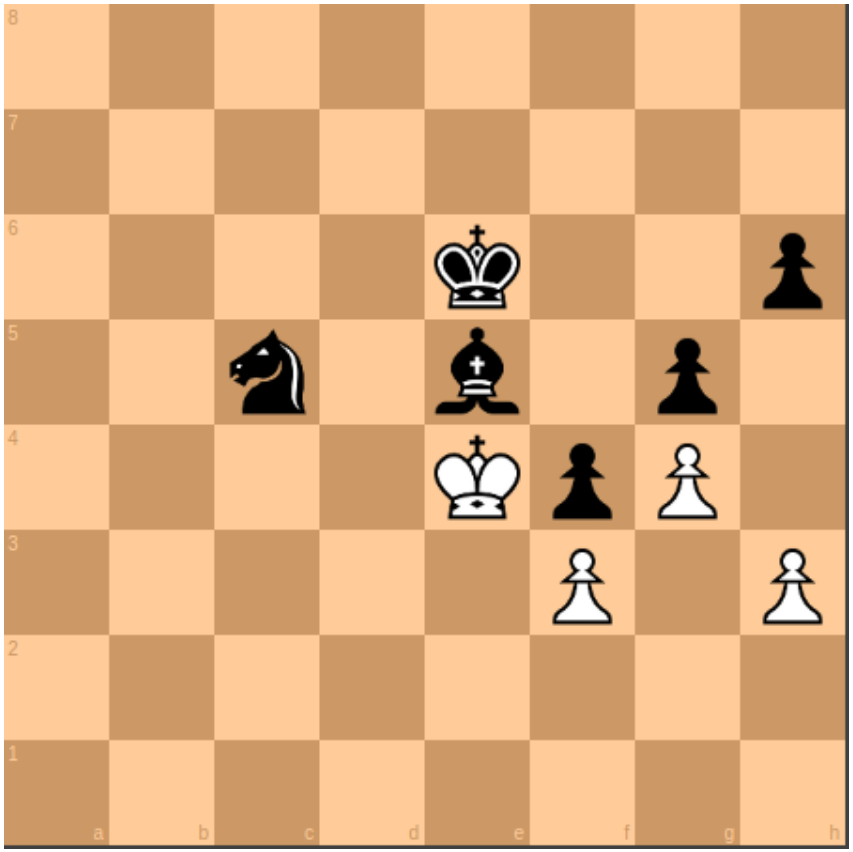


Figure: Black checkmates white

Reward function

- 0 when the game is ongoing
- -1 when black checkmates white (you lose)

Learning to play Chess

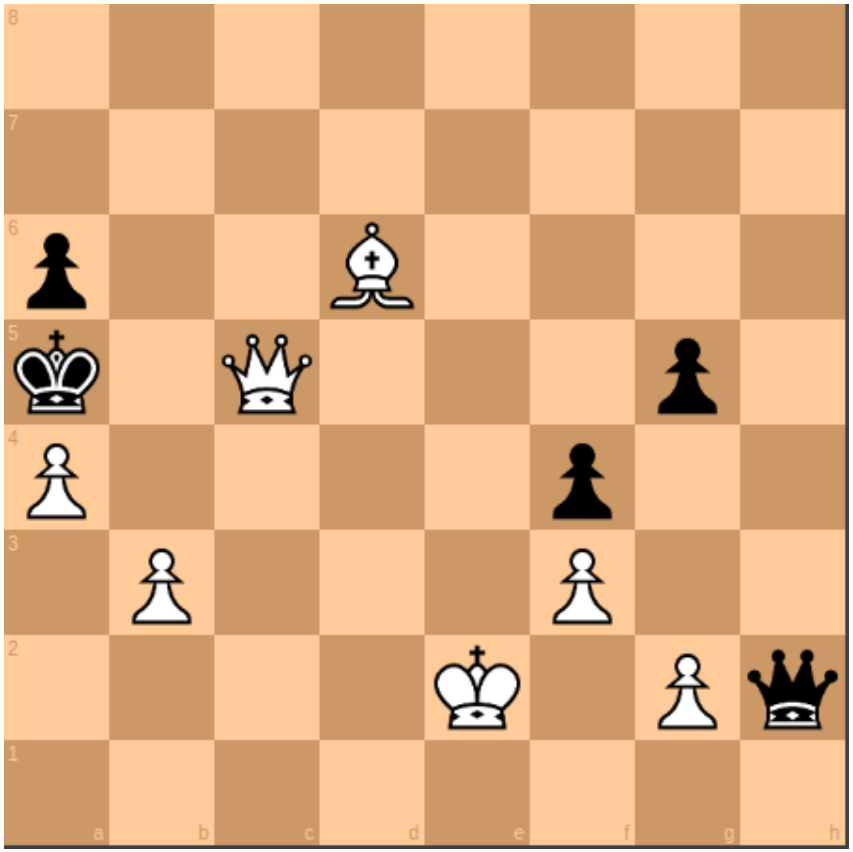


Figure: White checkmates black

Reward function

- 0 when the game is ongoing
- -1 when black checkmates white (you lose)
- $+1$ when white checkmates black (you win)

Learning to play Chess

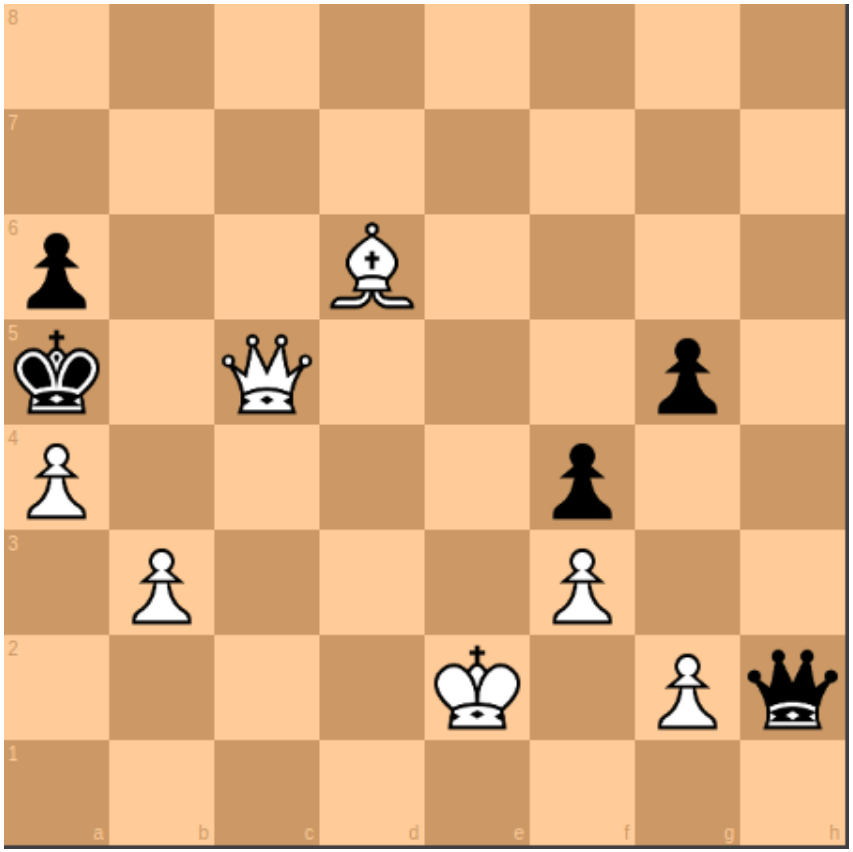


Figure: Good chess player

Reward maximization

- 0 when the game is ongoing
- -1 when black checkmates white (you lose)
- $+1$ **when white checkmates black** (you win)

Learning to trade stocks

Figure: Price of AAPL stock

Learning to trade stocks

Figure: Buy at $115.80

Actions

- Buy stock

Learning to trade stocks

Figure: Sell at $116.80

Actions

- Buy stock
- Sell stock

Learning to trade stocks

Figure: Profit of \$0.90

Profit

- sell price (\$116.80) − buy price (\$115.80) − broker fees (\$0.10) = \$0.90

Learning to trade stocks

Figure: Profit of $0.90

Reward function

- profit at every time step

Learning to trade stocks

Figure: Bought but not sold yet

Reward function

- profit at every time step
  - 0 for open trades

Learning to trade stocks

Figure: Profit of $-\$1.10$

Reward function

- profit at every time step
  - 0 for open trades
  - < 0 for losing trades

Learning to trade stocks

Figure: Profit of \$0.90

Reward function

- profit at every time step
  - 0 for open trades
  - < 0 for losing trades
  - > 0 for profitable trades

Learning to trade stocks

Figure: Profitable trading bot

Reward maximization

- profit at every time step
  - 0 for open trades
  - < 0 for losing trades
  - > 0 **for profitable trades**

Lane following in autonomous driving

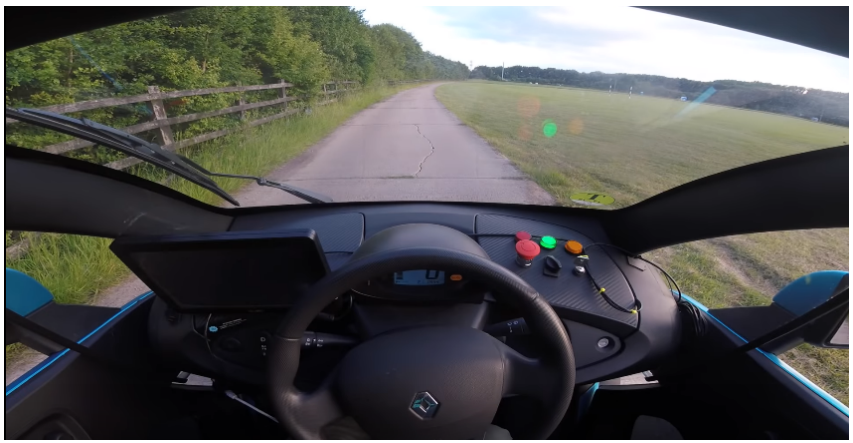


Figure: Car to follow the road

Lane following in autonomous driving

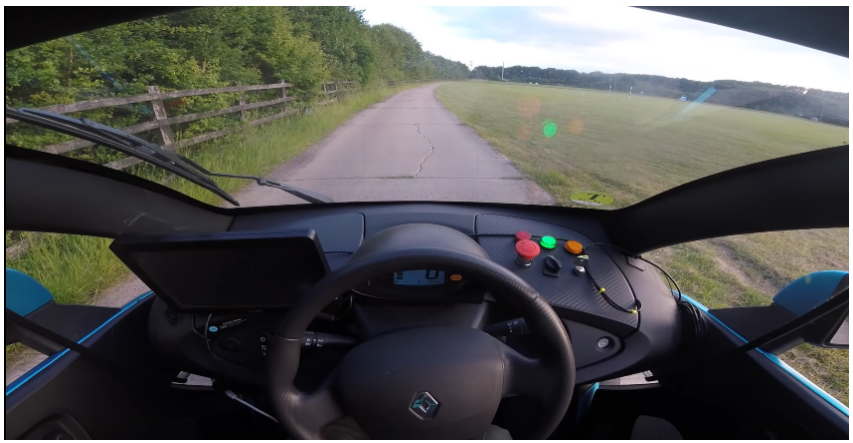


Figure: Following the road

Reward function

- Distance in meters traveled along the road without human intervention in a given time step

Lane following in autonomous driving

Figure: Going off road!

Reward function

- Distance in meters traveled along the road without human intervention in a given time step
- -1 for all time steps requiring human intervention

Lane following in autonomous driving

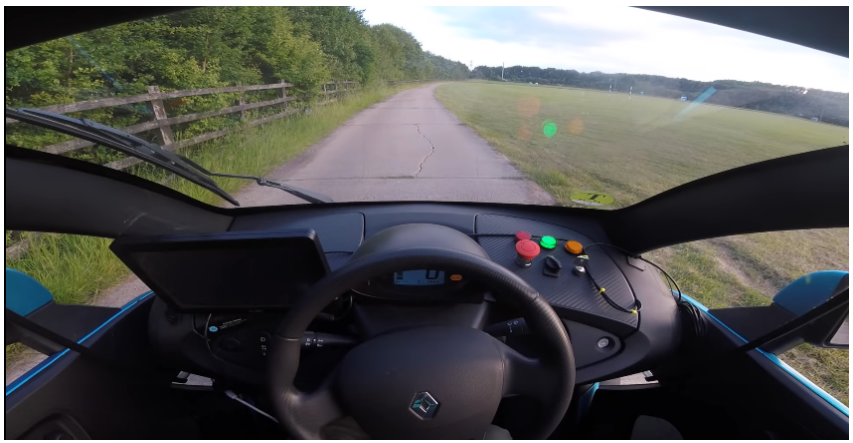


Figure: Autonomous lane follower

Reward maximization

- **Distance in meters traveled along the road without human intervention in a given time step**
- -1 for all time steps requiring human intervention

Learning to play video games

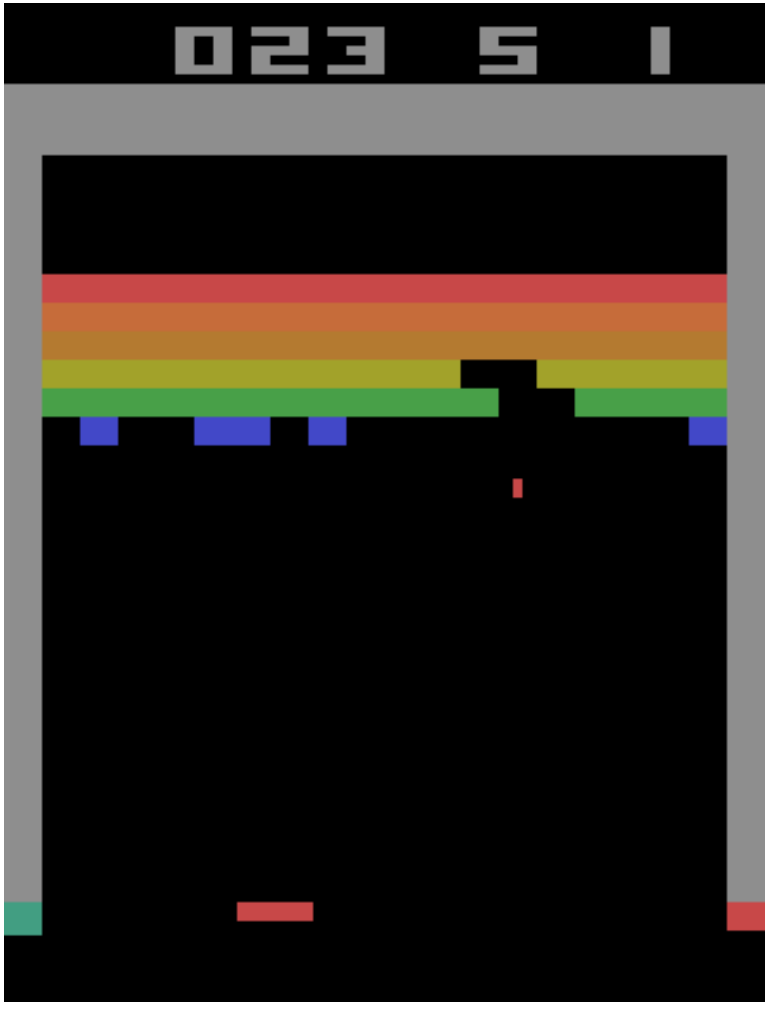


Figure: Atari Breakout

Learning to play video games

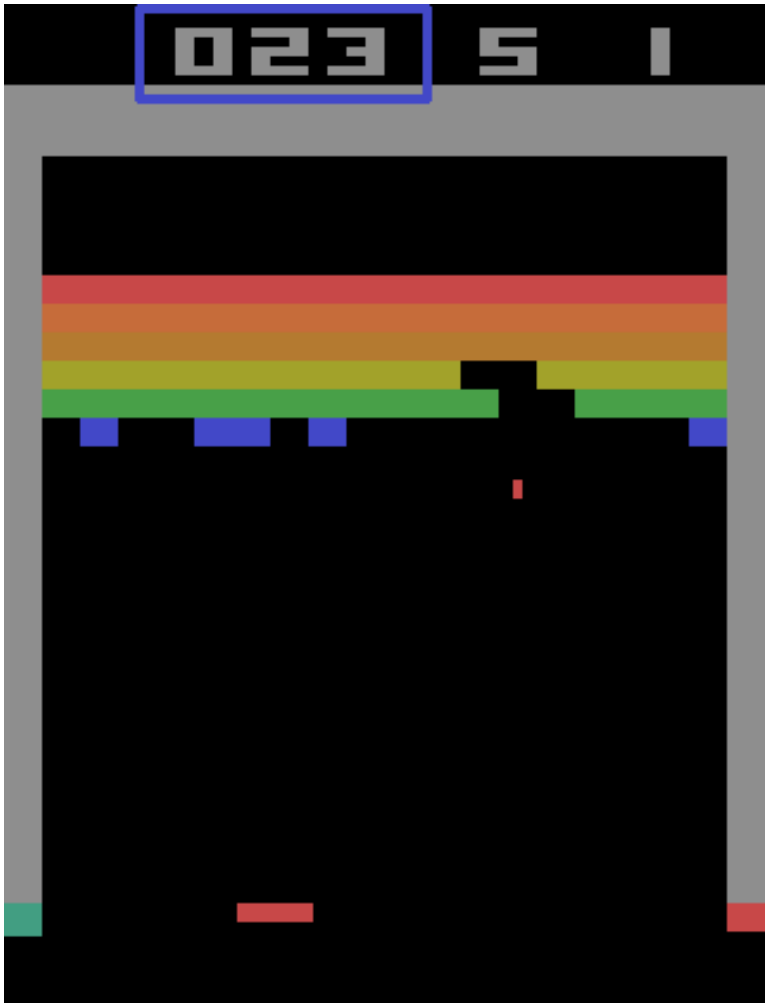


Figure: Video games have scores!

Learning to play video games

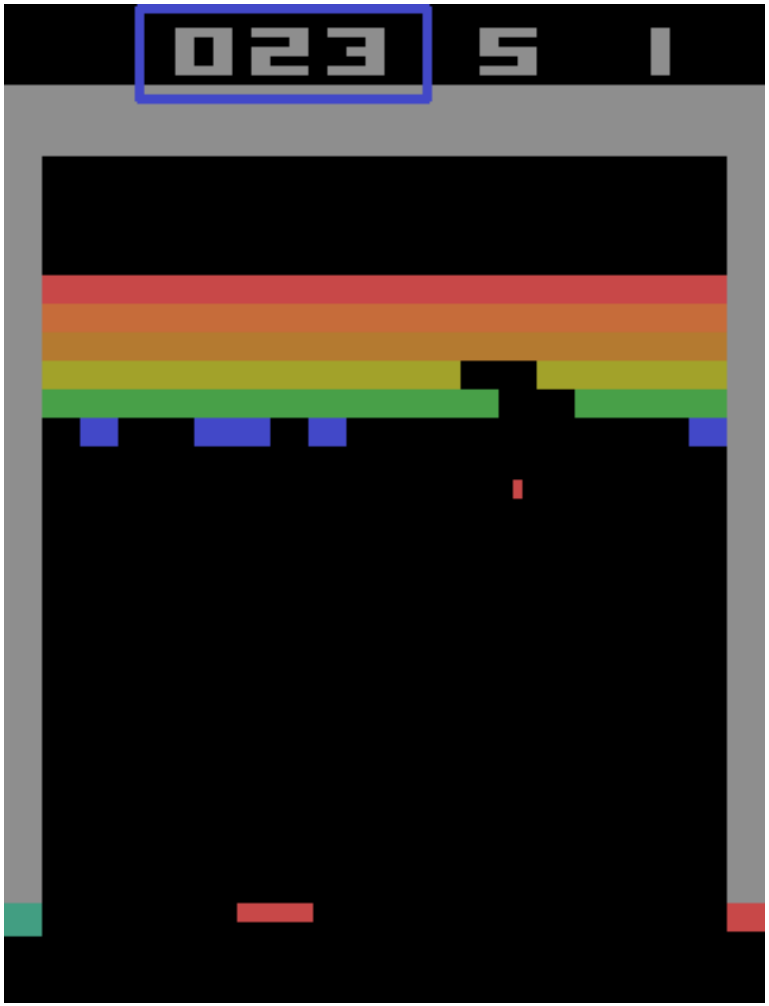


Figure: Video games have scores!

Reward function

- $\Delta(\text{score})$ at each time step

Remarks (Reward functions)

1. *Reward functions, whose maximization is equivalent to the learning goal, are intuitive and easy to design for a wide variety of problems.*

Remarks (Reward functions)

1. *Reward functions, whose maximization is equivalent to the learning goal, are intuitive and easy to design for a wide variety of problems.*
2. *Reinforcement Learning agents that learn to maximize these reward functions become extremely good at the tasks.*

Learning to play Chess

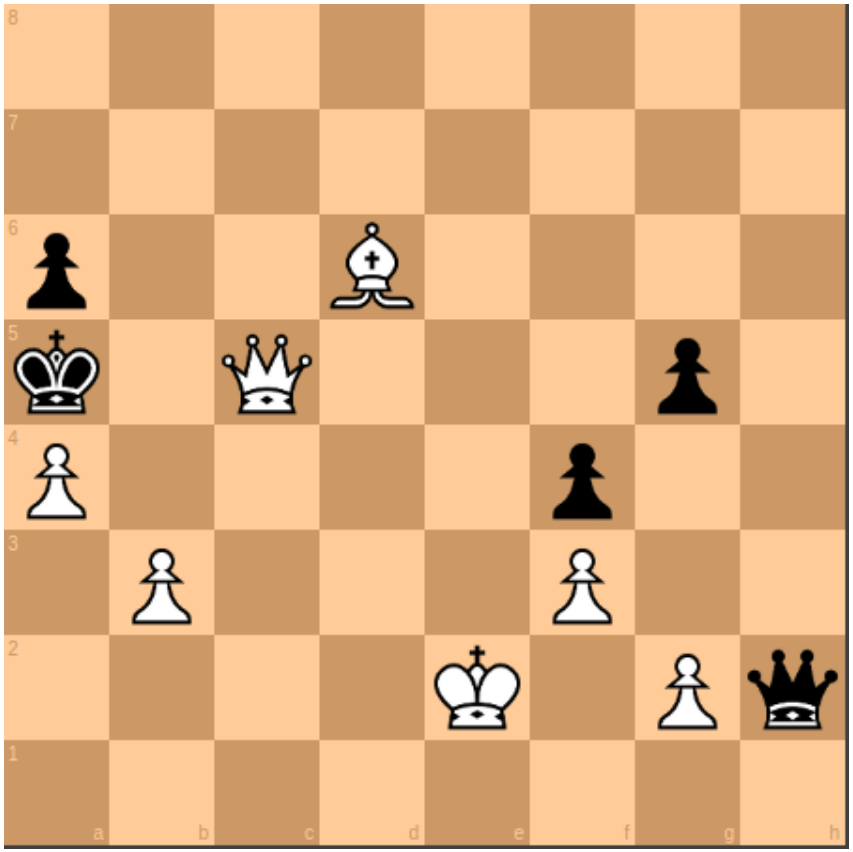


Figure: White checkmates black

Reward function

- 0 when the game is ongoing
- -1 when black checkmates white (you lose)
- $+1$ when white checkmates black (you win)

Learning to trade stocks

Figure: Profit of \$0.90

Reward function

- profit at every time step
  - 0 for open trades
  - < 0 for losing trades
  - > 0 for profitable trades

Lane following in autonomous driving

Figure: Going off road!

Reward function

- Distance in meters traveled along the road without human intervention in a given time step
- -1 for all time steps requiring human intervention

Learning to play video games

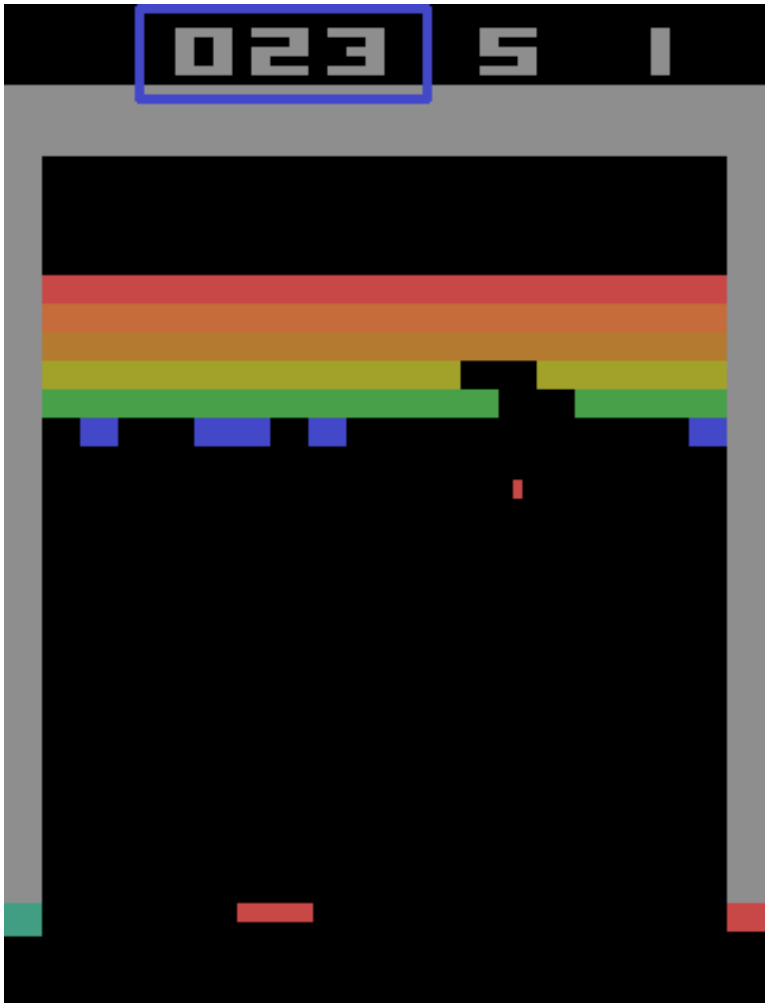


Figure: Video games have scores!

Reward function

- $\Delta(\text{score})$ at each time step

Remarks (The procedure is circural)

- *Set the learning goal.*

Remarks (The procedure is circural)

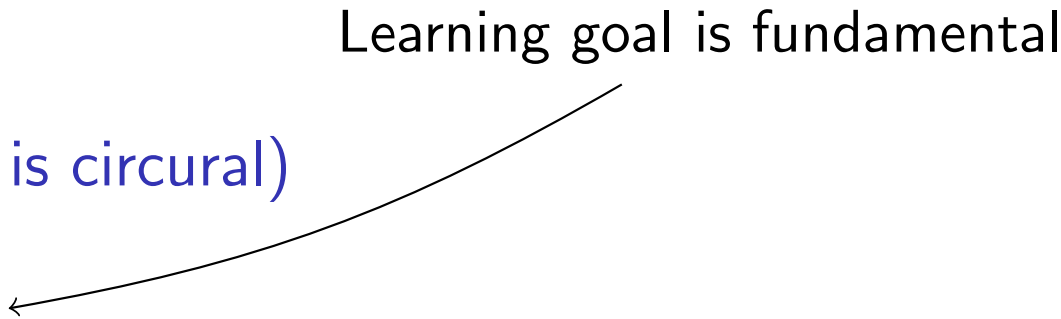
- *Set the learning goal.*
- *Create reward function, whose maximization is equivalent to the learning goal.*

Remarks (The procedure is circural)

- *Set the learning goal.*
- *Create reward function, whose maximization is equivalent to the learning goal.*
- *Reinforcement Learning agents learn to maximize this reward function.*

Remarks (The procedure is circural)

- *Set the learning goal.*
- *Create reward function, whose maximization is equivalent to the learning goal.*
- *Reinforcement Learning agents learn to maximize this reward function.*
- *They reach learning goal.*



Remarks (The procedure is circural)

- *Set the learning goal.*
- *Create reward function, whose maximization is equivalent to the learning goal.*
- *Reinforcement Learning agents learn to maximize this reward function.*
- *They reach learning goal.*

Learning goal is fundamental

Remarks (The procedure is circural)

- *Set the learning goal.*
- *Create reward function, whose maximization is equivalent to the learning goal.*
- *Reinforcement Learning agents learn to maximize this reward function.*
- *They reach learning goal.*

Reward is fundamental

Learning goal is fundamental

Remarks (The procedure is circural)

- *Set the learning goal.*
- *Create reward function, whose maximization is equivalent to the learning goal.*
- *Reinforcement Learning agents learn to maximize this reward function.*
- *They reach learning goal.*

Reward is fundamental

- General AI

Learning goal is fundamental

Remarks (The procedure is circural)

- *Set the learning goal.*
- *Create reward function, whose maximization is equivalent to the learning goal.*
- *Reinforcement Learning agents learn to maximize this reward function.*
- *They reach learning goal.*

Reward is fundamental

- General AI
- Applicable in humans/animals

Pursuit of Happyness

Pursuit of ~~Happyness~~ Happiness

The fundamental human desire to maximize happiness in our lives.

Pursuit of ~~Happyness~~ Happiness

The fundamental human desire to maximize happiness in our lives.

Reward function

Pursuit of ~~Happyness~~ Happiness

The fundamental human desire to maximize happiness in our lives.

Figure: Sadness is undesired

Reward function

- < 0 when we are sad.

Pursuit of ~~Happyness~~ Happiness

The fundamental human desire to maximize happiness in our lives.

Figure: Look at that elephant!

Reward function

- < 0 when we are sad.
- > 0 when we are happy.

Pursuit of ~~Happyness~~ Happiness

The fundamental human desire to maximize happiness in our lives.

Figure: Happiness is desired

Reward function

- < 0 when we are sad.
- > 0 when we are happy.

Remarks (Pursuit of happiness)

1. *In humans, rewards are fundamental and the drive to maximize it is in-built.*

Remarks (Pursuit of happiness)

1. *In humans, rewards are fundamental and the drive to maximize it is in-built.*
2. *A lot of complex intelligent behavior emerges from the "maximization of happiness" drive.*

Remarks (Pursuit of happiness)

1. *In humans, rewards are fundamental and the drive to maximize it is in-built.*
2. *A lot of complex intelligent behavior emerges from the "maximization of happiness" drive.*
3. *Worth thinking about: How much of human intelligent behavior is a result of this?*