

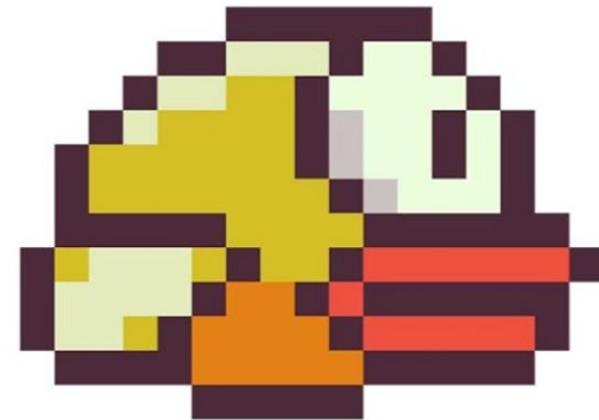
Mašinsko učenje u video igrama: Može li čovek pobediti AI?

Damjan Denić

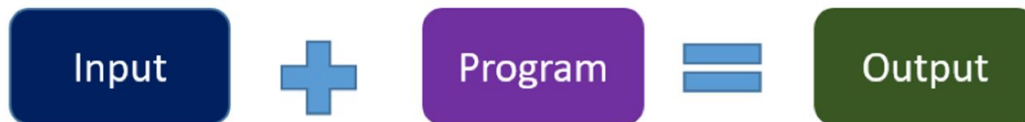
Matematička gimnazija

12. 04. 2022.

- Mašinsko učenje
- Uvod u Potkrepljenog učenje (RNN)
- Potkrepljenje u video igrama
- Metode potkrepljenog učenja (RNN)
 - Markov proces odlučivanja (MDP)
 - Q-learning
- Implementacija Q-learning metode u Flappy Bird
- Metode potkrepljenog učenja (RNN)
 - Neuro-evolucija promenljivih topologija (NEAT)
- Implementacija NEAT algoritma u Flappy Bird
- Zaključak



- Tradicionalno Programiranje:

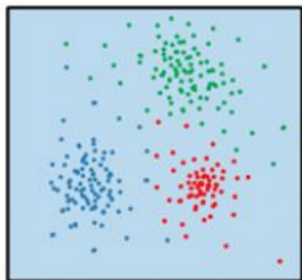


- Mašinsko učenje:

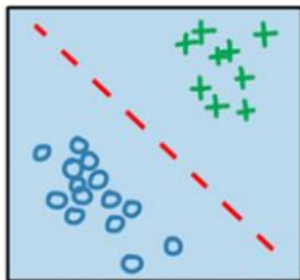


mašinsko učenje

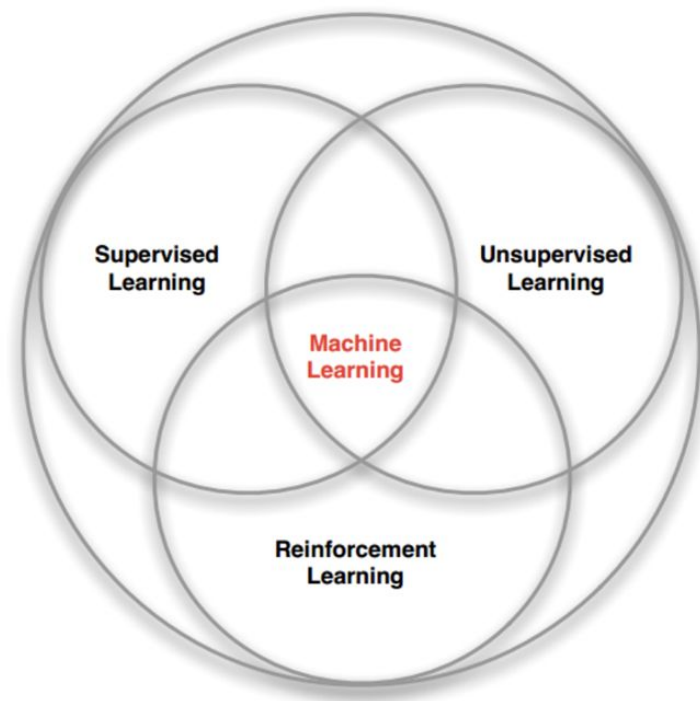
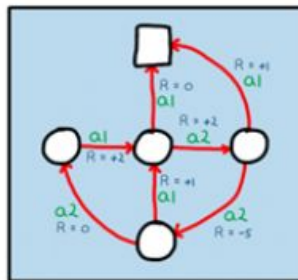
nenadgledano
učenje



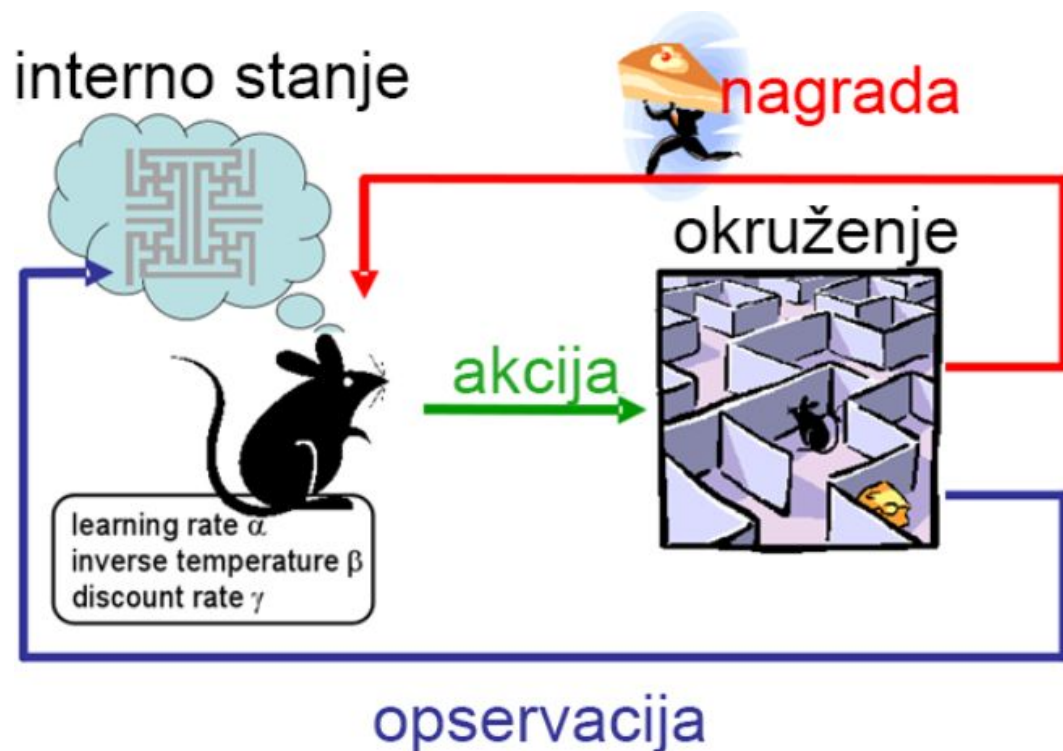
nadgledano
učenje



potkrepljeno
učenje



Krajnji cilj: predvideti najbolji sledeći korak da bismo zaradili najveću konačnu nagradu



- **DeepMind**
- **Atari igre - Agent57**
- Nisu preteške za rešavanje
- Bolje od prosečnog čoveka



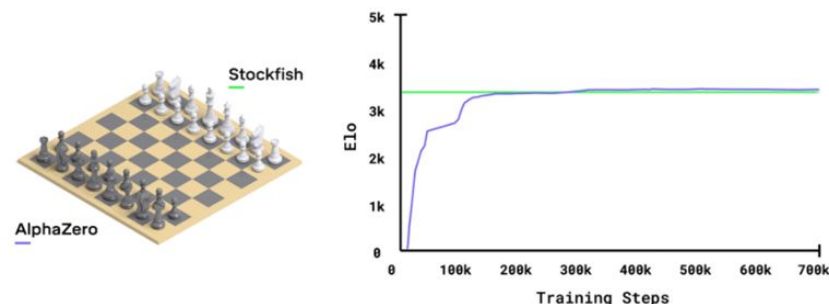
- **AlphaZero:**

- prvi koji je pobedio Stockfish u šahu posle samo 4 sata treninga
- prvi koji je pobedio Elmo u Šogiju posle 2 sata treninga

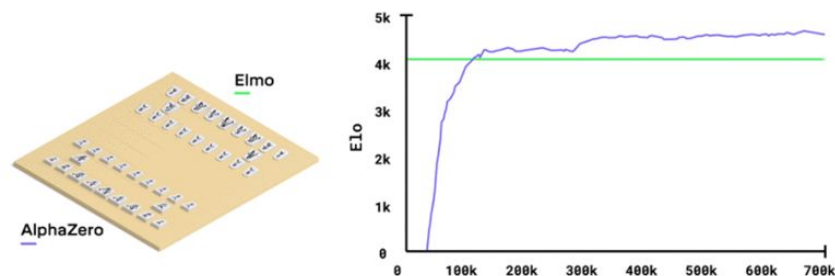
- **AlphaGo**

- Prvi koji je pobedio legendarnog Lee Sedola posle 30 sati treninga

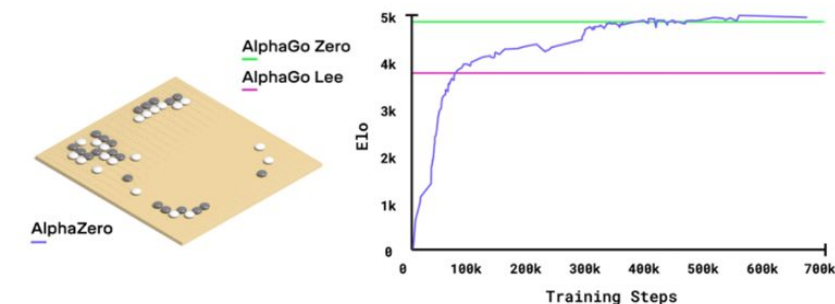
Chess



Shogi



Go



- **OpenAI Five**

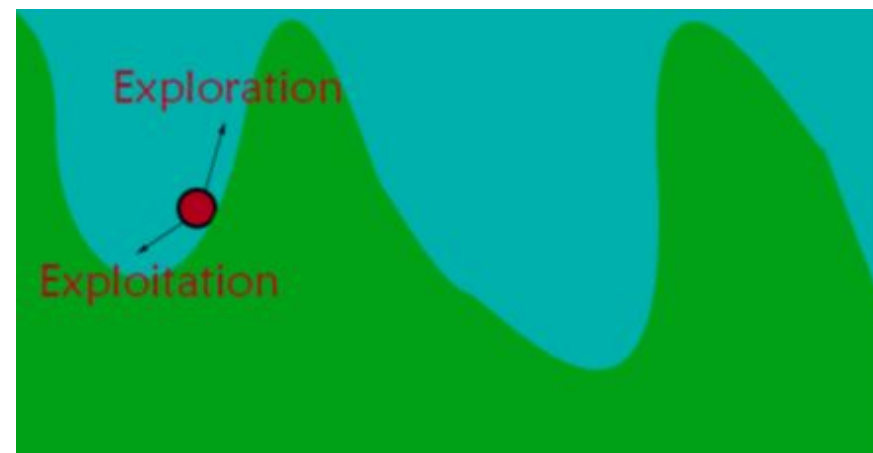
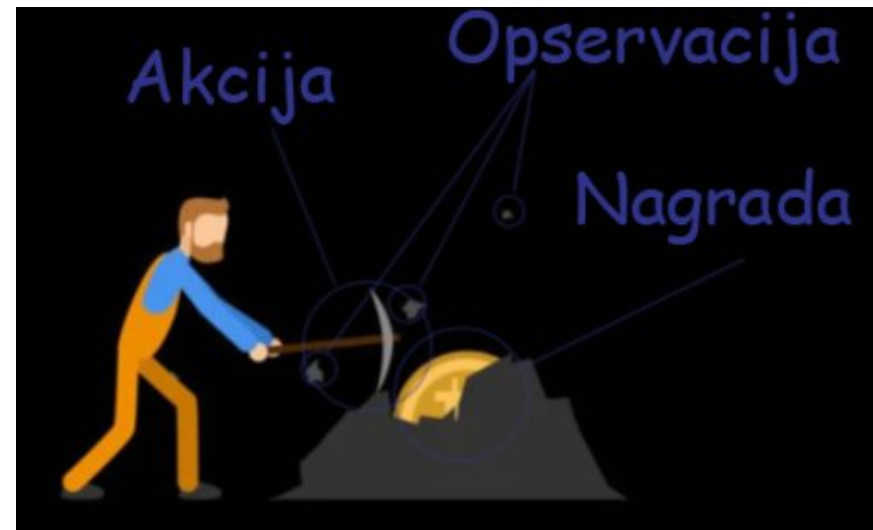
- 99.4% pobeda
- 33K igrača
- 122.000 procesorskih jezgara
- 2018 - 2019



- Akcija, Opservacija, Nagrada, Stanje
- A Stanje **S** se smatra Markovljevim ako i samo ako:

$$P[S_{t+1}|S_t] = P[S_{t+1}|S_1, S_2, \dots, S_t]$$

- Istraživanje ili Ekploatacija



Markovljev proces odlučivanja (Markov Decision Process MDP)



- MDP se definiše kao **($\mathbf{S}$, $\mathbf{A}$, $\mathbf{P}$, $\mathbf{R}$, γ)**
 - **$\mathbf{S}$** je skup stanja (states)
 - **$\mathbf{A}$** je skup akcija (actions)
 - **$\mathbf{P}$** je skup verovatnoća prelaza stanja (state transition probability matrix)

$$P_{ss'}^a = P[S_{t+1} = s' \mid S_t = s, A_t = a]$$

- **$\mathbf{R}$** je funkcija nagrade (reward)

$$R_s^a = E[R_{t+1} \mid S_t = s, A_t = a]$$

- γ – težinski faktor (discount factor) $\gamma \in [0, 1]$

Markovljev proces odlučivanja (Markov Decision Process MDP)



- **G_t** je ukupna nagrada u trenutku **t**

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots$$

- **G_t** je i dalje predikcija!
- **Polisa** (Policy)
- Polisa je raspodela akcija **A** za dato stanje **S**.

$$\pi(a|s) = P[A_t = a | S_t = s]$$

- Q-learning je algoritam potkrepljenog učenja koji uči vrednosti akcija za konačno stanje bez
- Ne zahteva model okruženja
- Radi odlično and konačnim Markovljevima procesima (FMDP).
- $Q(S_t, A_t)$ – očekivana nagrada za akciju **A_t** u stanju **S_t** .

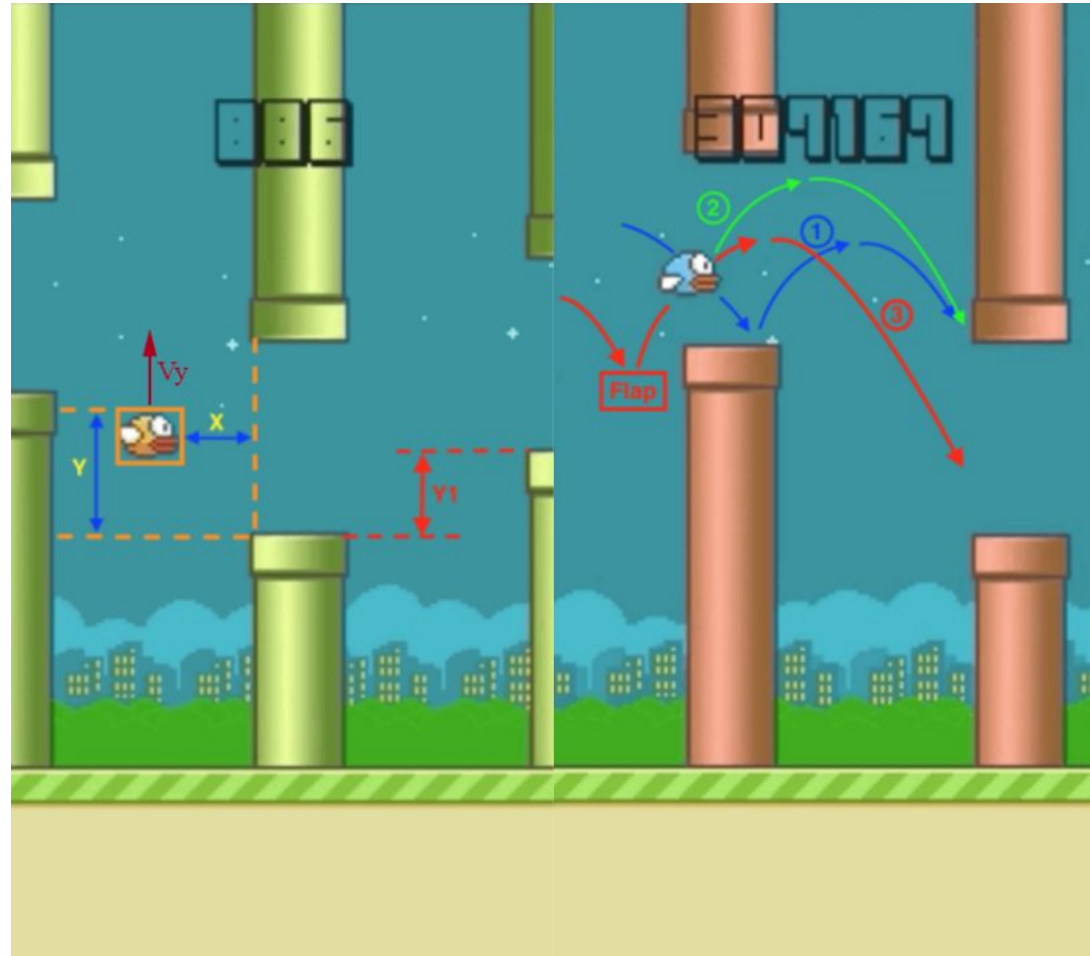
$$Q^{new}(s_t, a_t) \leftarrow \underbrace{Q(s_t, a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \underbrace{\left(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}} - \underbrace{Q(s_t, a_t)}_{\text{old value}} \right)}_{\text{temporal difference}}$$

new value (temporal difference target)

Flappy Bird Implementacija



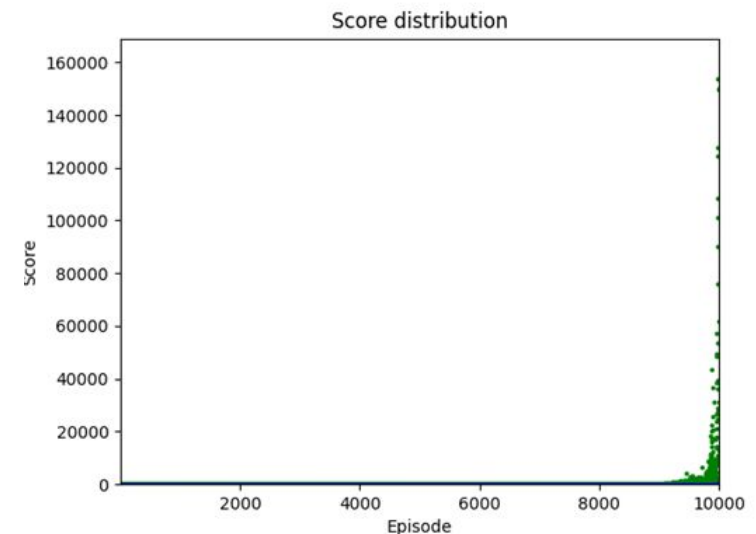
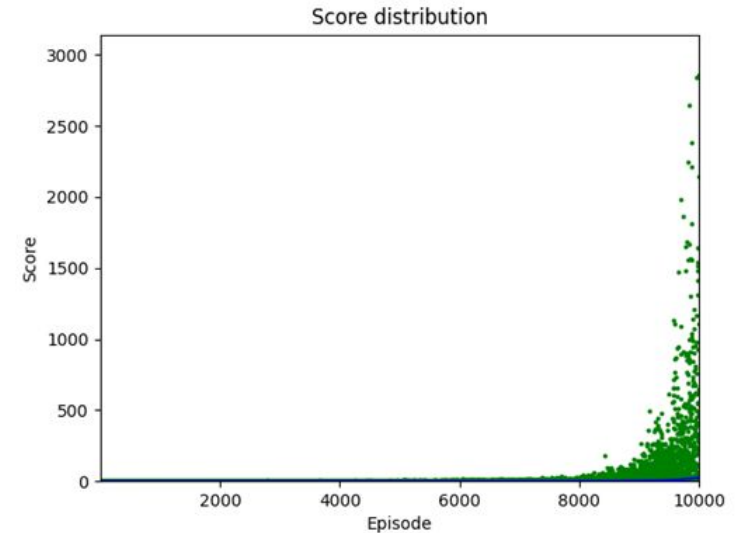
- Markovljev Proces?
- Stanja
 - $X, Y, Y1, V_y$
 - $V_x, X1$?
- Akcija
 - Skoči/Nemoj da skočiš
- Opservacija
 - Sledeće stanje
 - Da li sam udario u šipku
- Nagrada
 - Živ: +1 / 0
 - Sudar sa šipkom: -1000



Flappy Bird Posle Ažuriranja

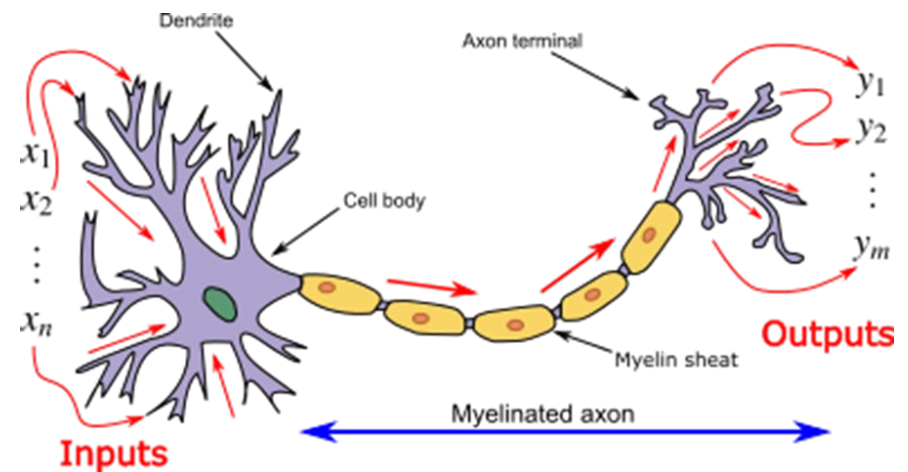
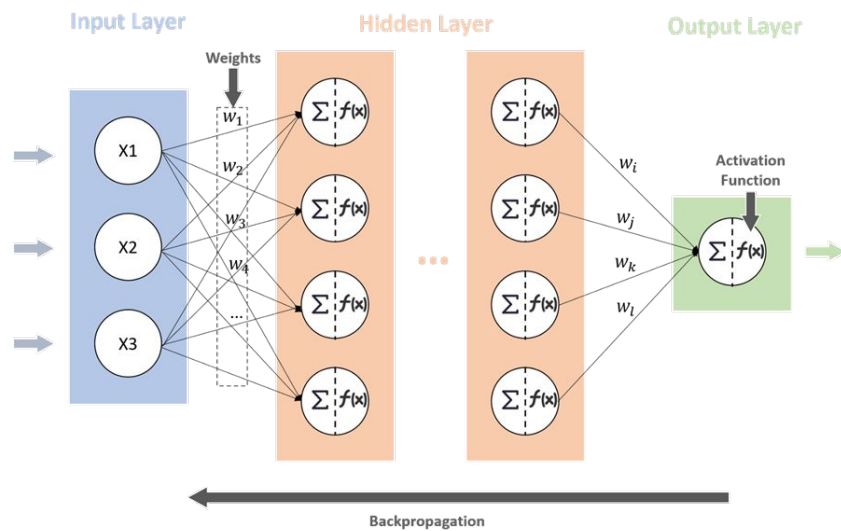


- Stari algoritam
 - $Q(x,y,vel,y1) \leftarrow (0, 0, 0)$
 - Broj epoha: 10000
 - Najviši rezultat: 2857
 - Prosek: 24.0664
- Novi algoritam:
 - $Q(x,y,vel,y1) \leftarrow Q(x, y, ?, ?)$
 - Broj epoha: 10000
 - Najviši rezultat: 153635
 - Prosek: 292.2415



Flappy bird implementacija





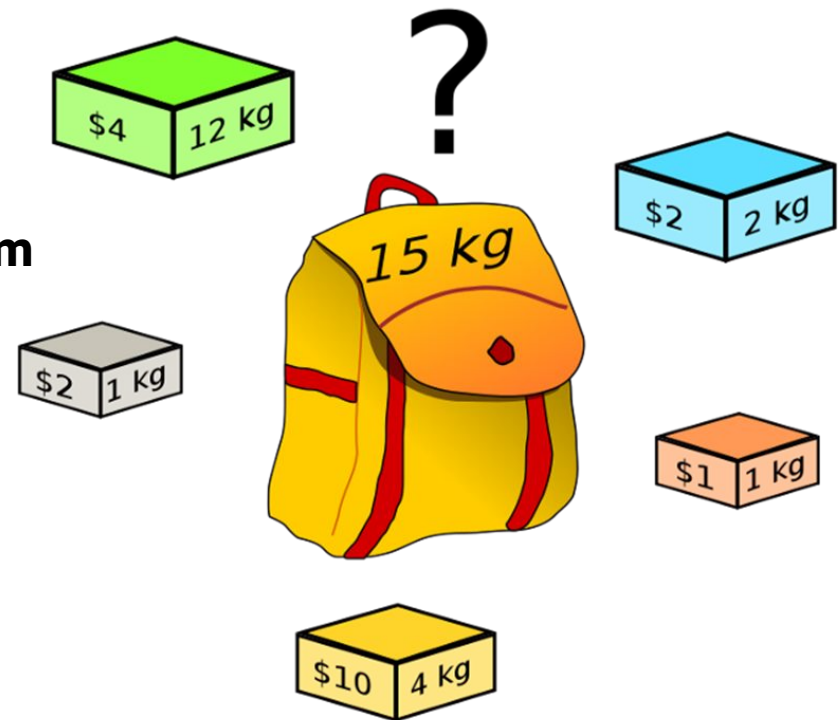
Neuroevolucija promenljivih topologija

NeuroEvolution of Augmenting Topologies (NEAT)



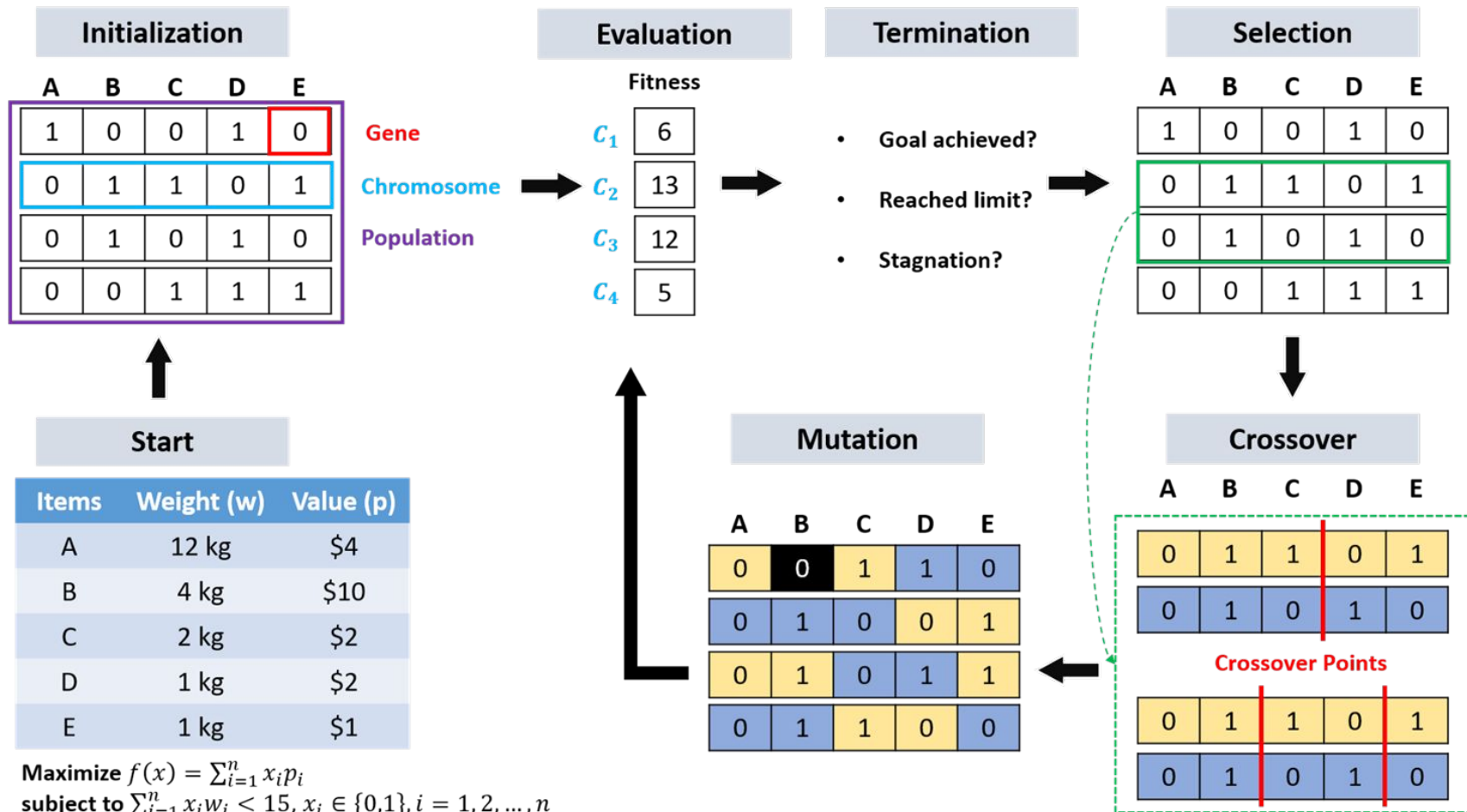
- **Vraćanje unazad vs. Genetski algoritam**

- “Svedoci smo divnih adaptacija u svakom ćošku organskog sveta.”
— Čarls Darwin, Postanak vrsta.



Neuroevolucija promenljivih topologija

NeuroEvolution of Augmenting Topologies (NEAT)



Flappy bird NEAT implementacija



Hvala na pažnji!

Pitanja?