- epsilon annealing: fit like old & make experience-based → Durchschnitte. vorkommen von Sa 1 Runde? + Median wegen Outliers

① reward shaping: not got. killed but for loops & too long waiting and maybe kamikaze modus

③ try to not show modes → modus 1: if sitting on coin = 1 (exclude own position)

② improve mode 3

- improve maps

→ train very long (until every reachable state-action pair occured ... times ?) (ask: 2nd computer ?)

→ finish Abgabe

Resultierende Ideen:

1. Idee:

if #collectable coins ≠ 0:
    Modus = 1

if #opponents > 0 & #coins = 0:
    Modus = 3
else:
    Modus = 2

2. Idee: in Modus 2 (coin Miner):

$$\text{reward speed} = \frac{\text{coin density} \cdot \text{crates destroyed} + h \cdot \#\text{opponents in bombing spread}}{\text{time to travel}}$$

h: kill probability · reward, heuristisch; z.B.

$$= \begin{cases} 0 & \text{für nicht own\_position} \\ \text{hyperparameter} \cdot \frac{4 - d(\text{opponent, me})}{3} \cdot 5 \end{cases}$$

→ weniger perfektionistisch & Wichtigkeit von Abstand
+ vermeide Opponent-Modell

Weitere Ideen dazu:
- score of agents (über runden!) ("kill small fishes")
- anzahl freier Felder drumherum bzw. um opponent
- Anzahl freier Felder die dann von Explosion bedeckt wären

3. Idee: in Modus 3 (Hunter)

if foe not in your bomb spread:

$f_5 = 0$ oder 1

$f_{1,2,3,4} = 2$ if towards nearest foe

else: compute 'local good spot' for foe, comparing own position and neighbors (and maybe further fields) by calculating something like

opponents in bomb spread . ?

↑

z.B. freie Felder etc.

→ $f_{1,2,3,4,5} = 2$ if best (local) spot