Änderungen zu normalen Carcode:

Action space nicht mehr diskret sondern Box zwischen -1 und 1.

* Andere Berechnung der Größe von Q Tabelle und Berechnung der Action jetzt ein Randomwert zwischen low und high (-1 und 1).
* Q Wert aus Q Tabelle: action = np.argmax(Q[state\_adj[0], state\_adj[1]]) action=action\*np.array([1])-1

Macht eigentlich keinen Unterschied, da eh nur die Integerwertegenommen werden können

Problem erster Durchlauf: Wert immer 0 oder -1, da Wert für rechts immer kleiner ist

Action muss um 1 erhöht werden, da ansonsten rechts in der mitte gespeichert wird etc.

Bewegung bringt negativen reward-> man endet in der Mitte mit reward 0

Das liegt daran das links und rechts schlechter bewertet werden->

Bei gleich großen werten wird stehen bleiben bevorzugt, außerdem wird der wert nicht schlechter, während die anderen Richtungen immer schlechter bewertet werden

Kein Schwung sodass auch keine neuen Werte erreicht werden -> bei allen wird mit der Zeit stehen bleiben gewählt, dessen wert immer mehr auf 0 geht

Bei komplett Zufall immer -50 Average