Instituto Tecnológico de Aeronáutica - ITA
Inteligência Artificial para Robótica Móvel - CT-213
Aluno: Danilo de Farias Matos

Relatório do Laboratório 12 - Deep Q-Learning

1. Breve Explicação em Alto Nível da Implementação

Iniciei implementando a função reward_engineering_mountain_car() extraindo a posição e velocidade do estado atual, depois o estado futuro e apliquei a fórmula proposta pelo exercício com duas alterações: transformei a recompensa pela velocidade em algo linear porque percebi que o valor da velocidade era menor que 0 e elevar isso ao quadrado gerava uma recompensa muito pequena, adicionei uma outra recompensa intermediária de pequeno valor para indicar que o agente estava chegando perto do objetivo, mantendo a ideia de recompensar bastante caso o carrinho alcançasse o objetivo.

Depois disso foi a vez de implementar as funções make_model() e act() do agente. Para a primeira função eu repliquei a proposta da rede neural contida na tabela 3 do roteiro do laboratório e depois implementei a política epsilon-greedy para a função act()

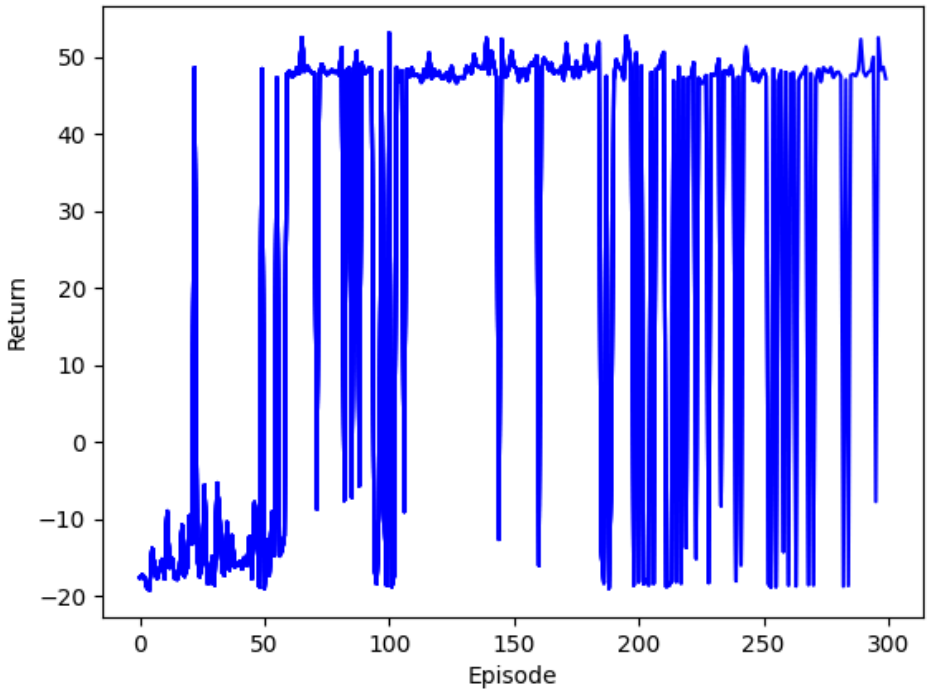
2. Figuras Comprovando Funcionamento do Código

Basta colocar as figuras.

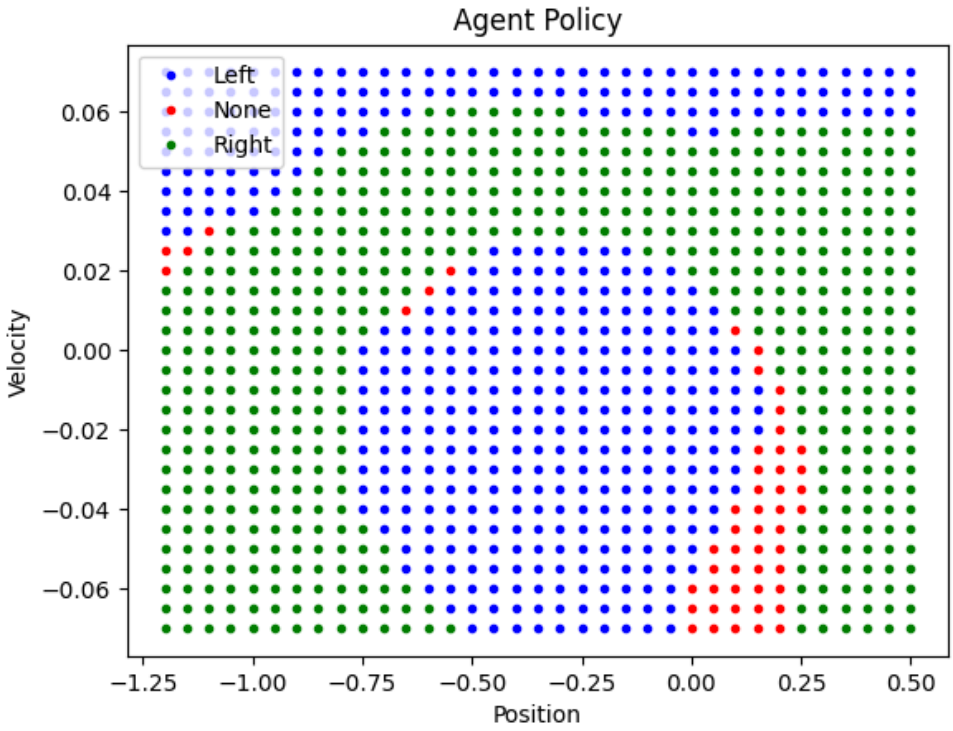
2.1. Sumário do Modelo

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 24)                72

 dense_1 (Dense)             (None, 24)                600

 dense_2 (Dense)             (None, 3)                 75

=================================================================
Total params: 747 (2.92 KB)
Trainable params: 747 (2.92 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
Loading weights from previous learning session.
```

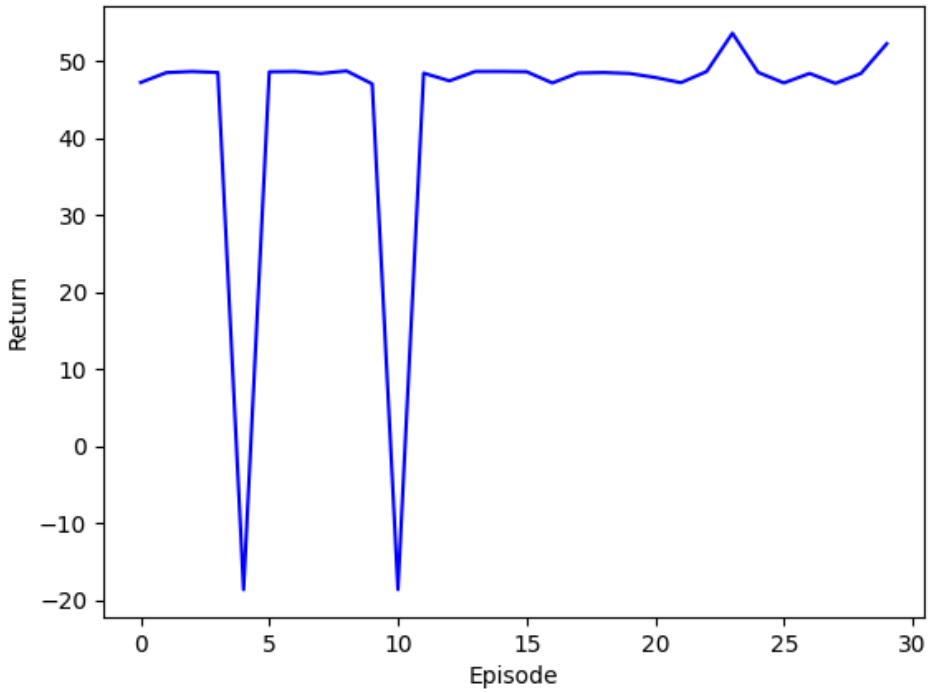
2.2. Retorno ao Longo dos Episódios de Treinamento



2.3. Política Aprendida pelo DQN



2.4. Retorno de 30 Episódios Usando a Rede Neural Treinada.



3. Discussão dos Resultados

Percebi que a implementação de recompensa intermediária antes da chegada no objetivo incentivou o agente a desacelerar quando chegava em uma velocidade muito grande próxima do objetivo final (pois este comportamento gerava mais recompensas) e que punir o agente por ir demais para trás fazia com que ele não aprendesse a subir a rampa (pois ele tentava acelerar sempre numa mesma direção). O aprendizado não apresentou uma convergência clara de recompensas, mas foi bom o suficiente para o agente conseguir finalizar o circuito proposto a maioria das vezes durante o teste.