Entrega final tarea 3: Actor - Critic

Código: EL7021-1

Nombre: José Luis Cádiz Sejas

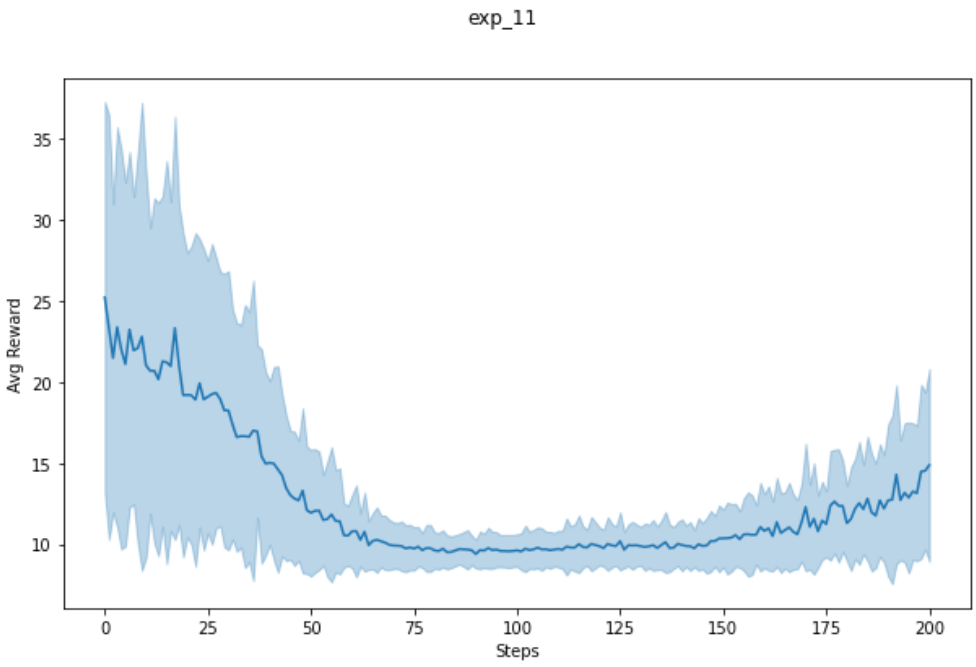
1. **Parametrización del actor**:
   1.1. Código adjunto.
   1.2. Código adjunto.

2. **Muestreo de trayectorias**.
   2.1. Código adjunto.
   2.2. Código adjunto.

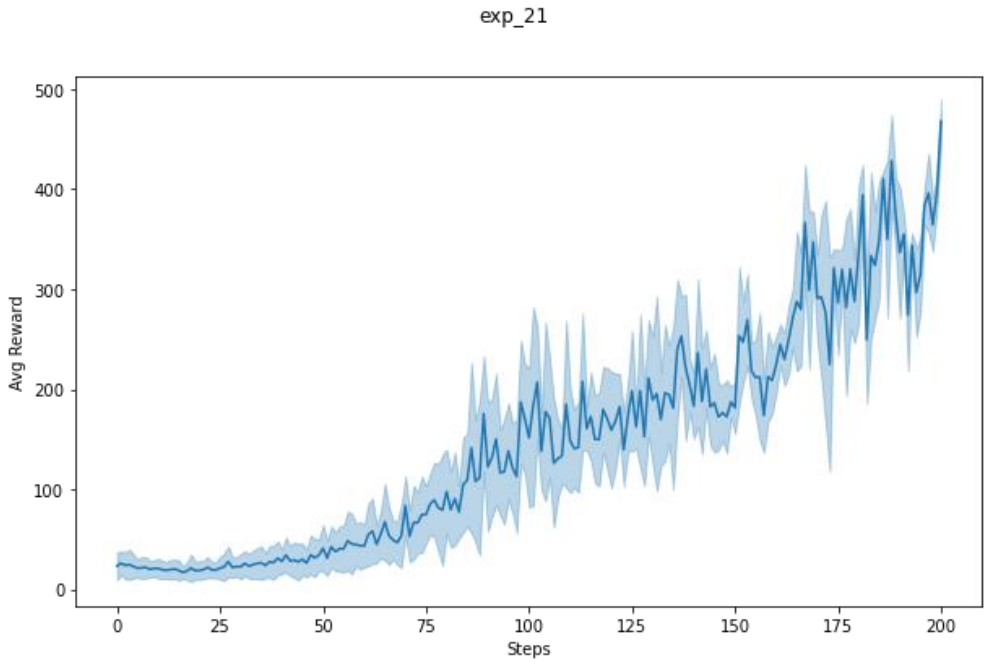
3. **Parametrización del crítico y actualización**:
   3.1. Código adjunto.
   3.2. Código adjunto.

4. **Experimentos CartPole:**

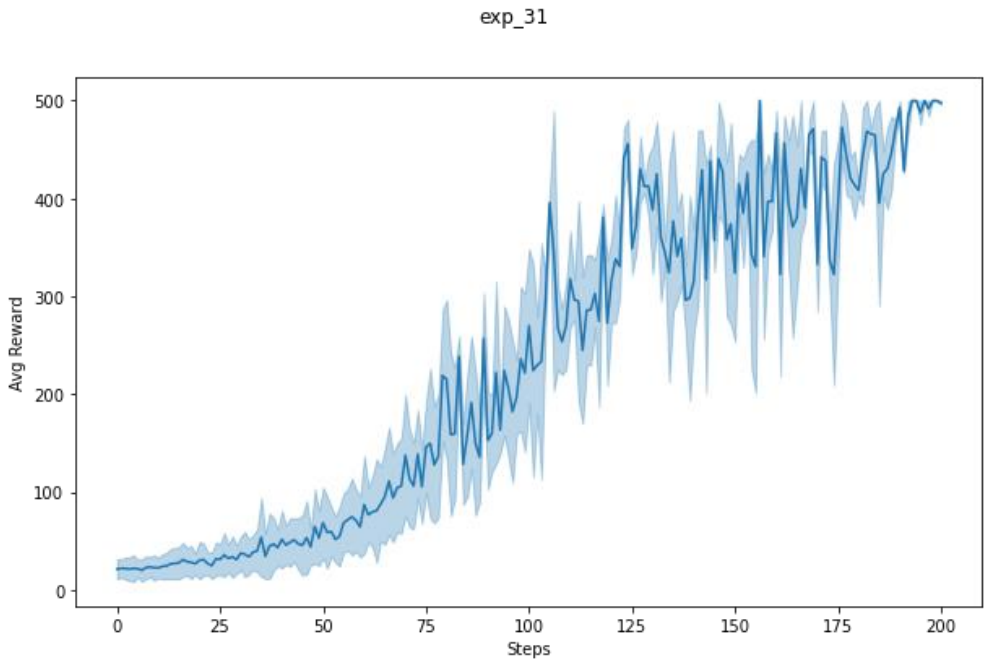
4.1.
```
exp_11={"name":"exp_11", "batch_size":500, "nb_critic_updates":1}
```



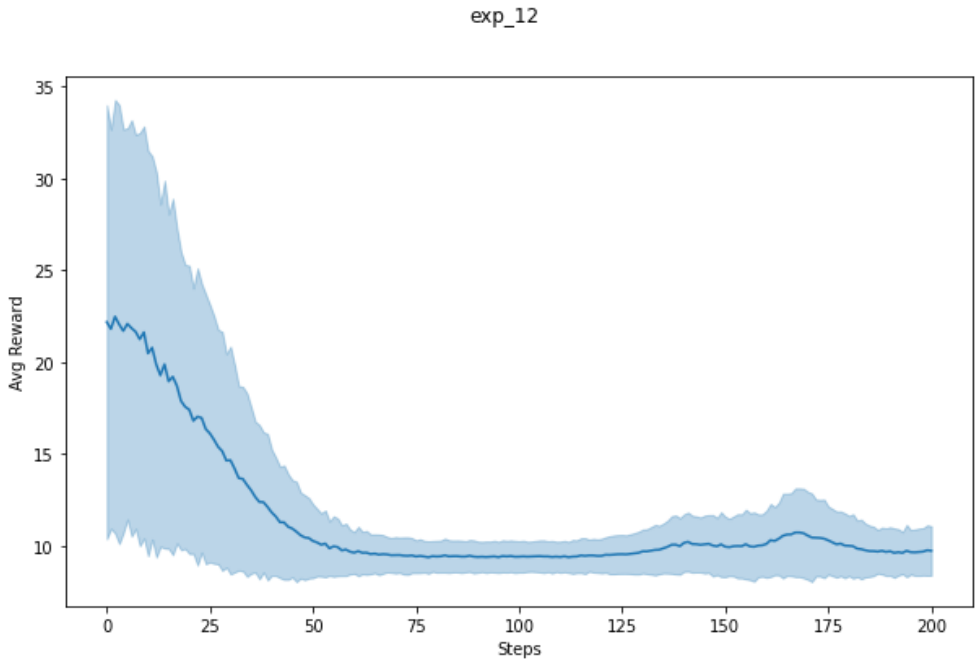
4.2. `exp_21={"name":"exp_21", "batch_size":500, "nb_critic_updates":10}`



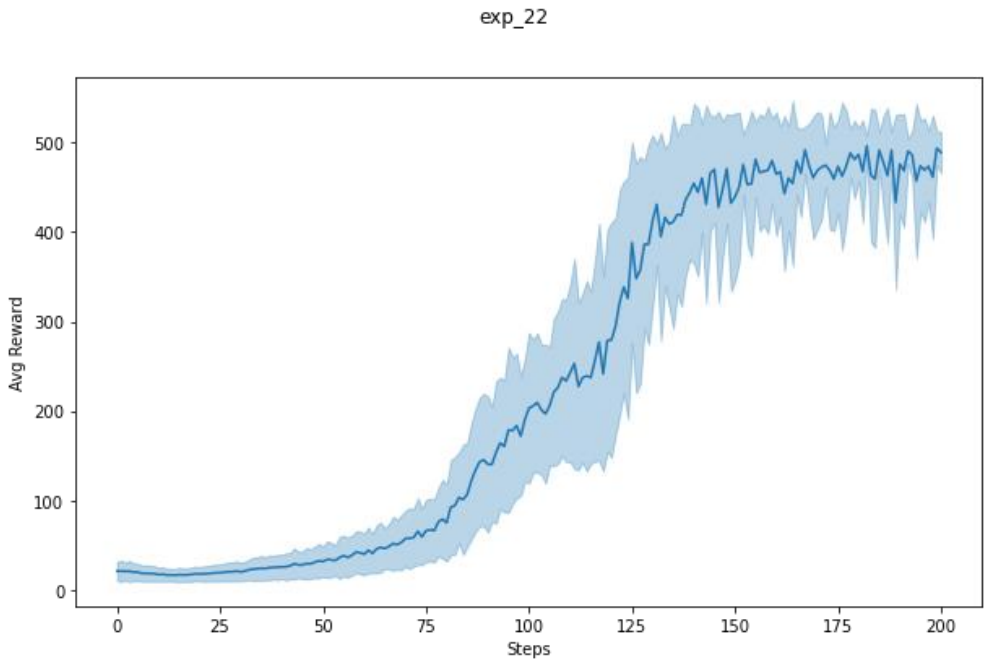
4.3. `exp_31={"name":"exp_31", "batch_size":500, "nb_critic_updates":100}`



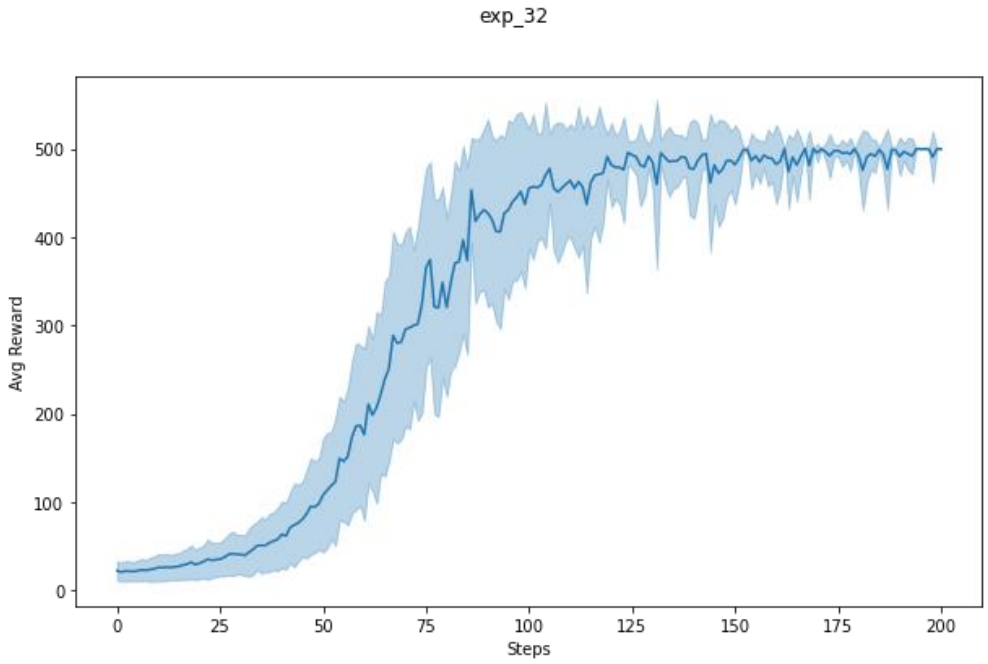
4.4. exp_12={"name":"exp_12", "batch_size":5000,"nb_critic_updates":1}



4.5. exp_22={"name":"exp_22", "batch_size":5000,"nb_critic_updates":10}

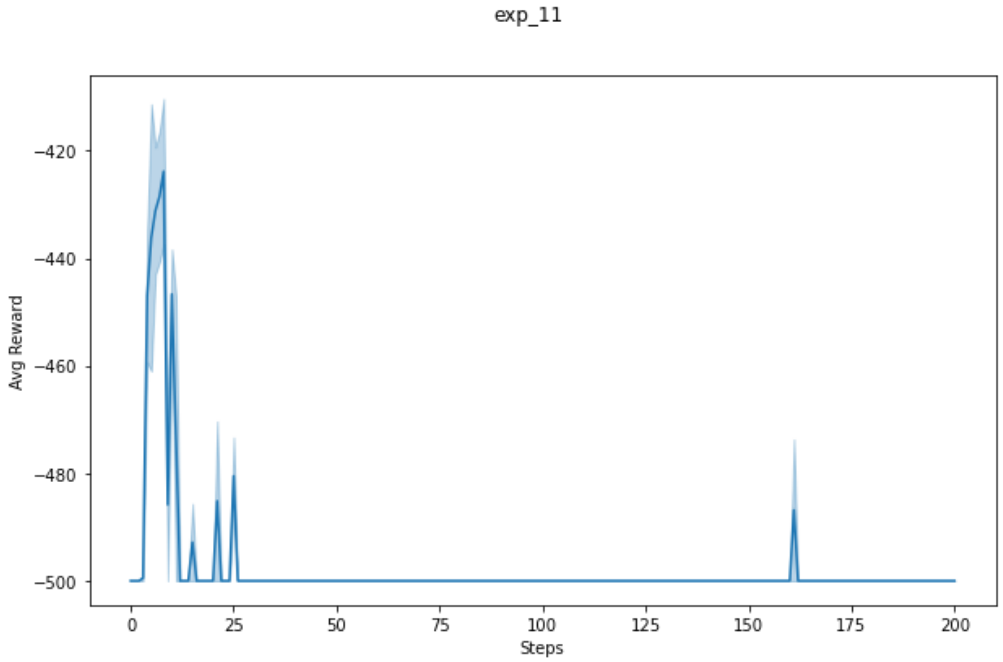


4.6. exp_32={"name":"exp_32", "batch_size":5000,"nb_critic_updates":100}

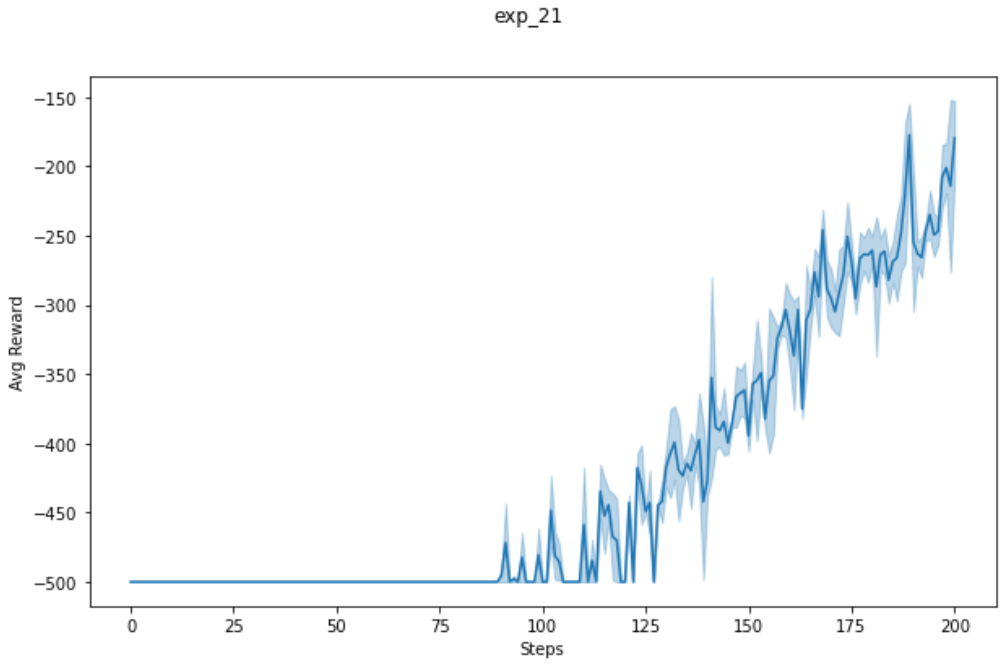


5. **Experimentos Acrobot:**

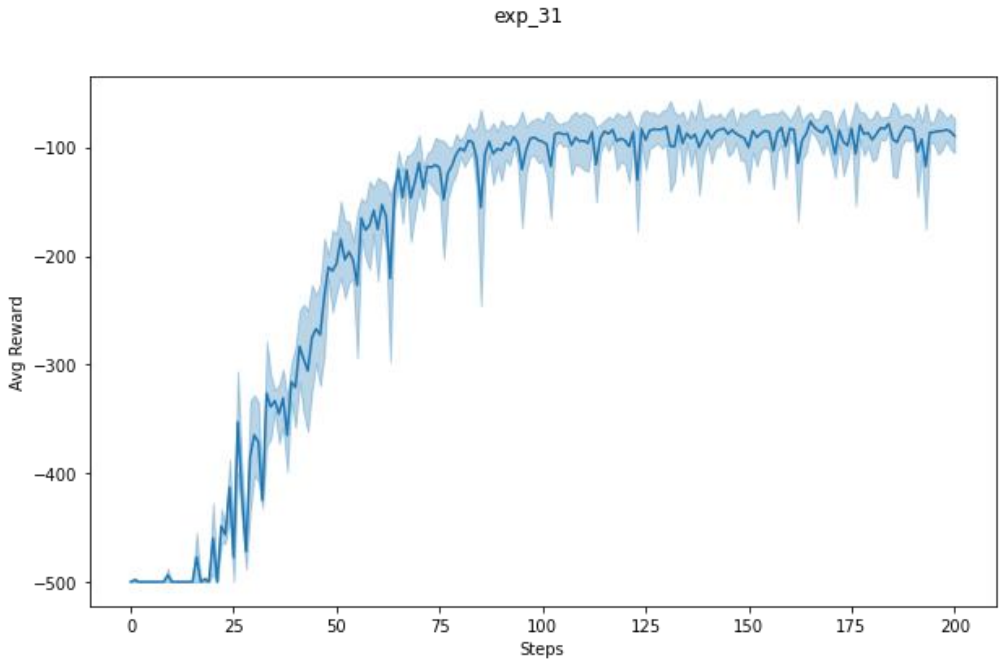
5.1. exp_11={"name":"exp_11", "batch_size":500, "nb_critic_updates":1}



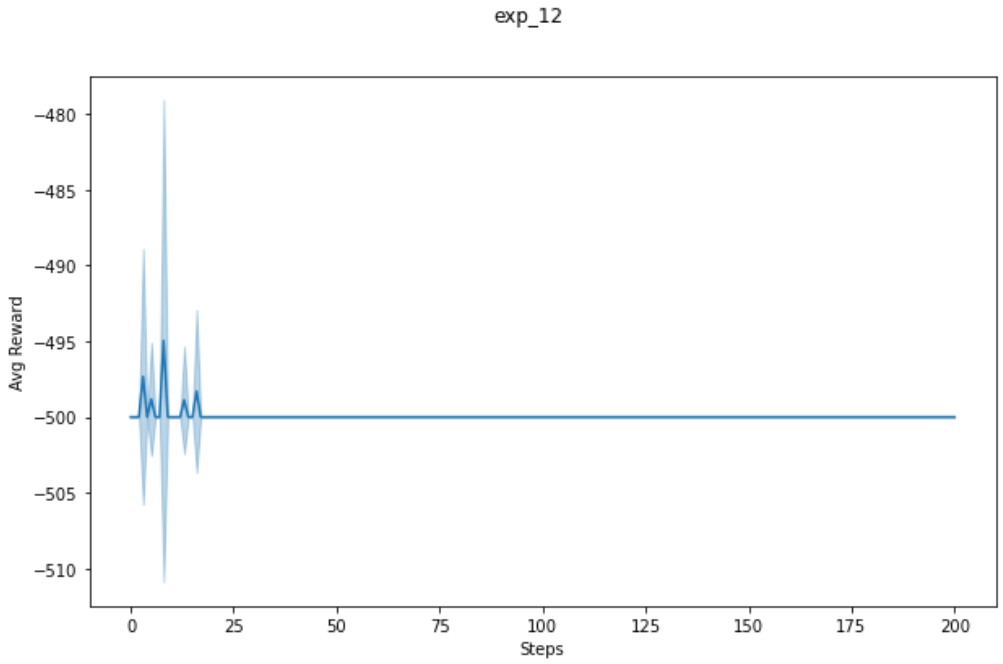
5.2. `exp_21={"name":"exp_21", "batch_size":500, "nb_critic_updates":10}`



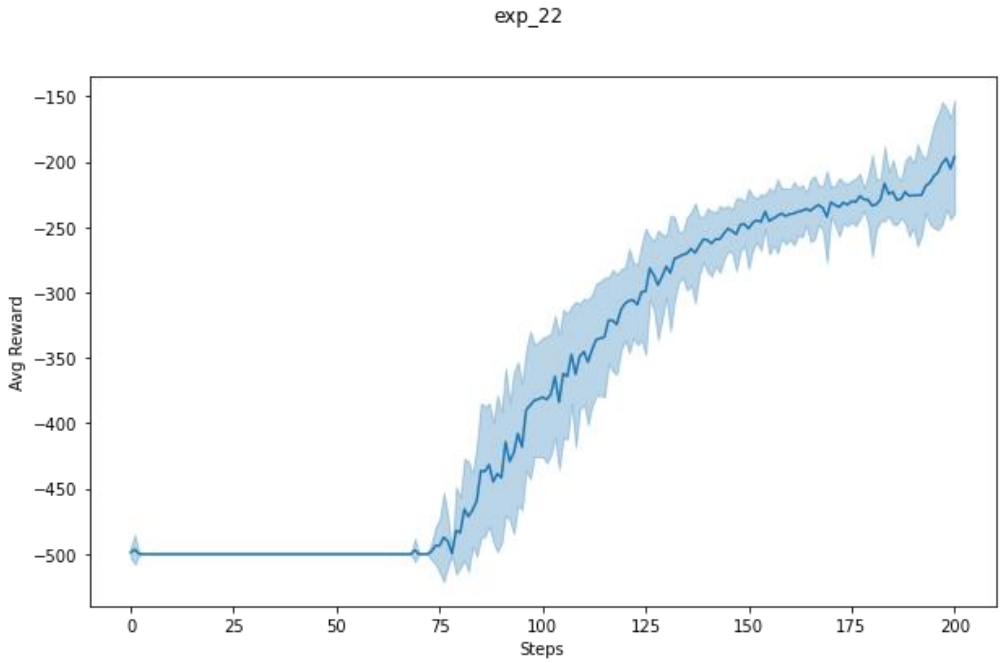
5.3. `exp_31={"name":"exp_31", "batch_size":500, "nb_critic_updates":100}`



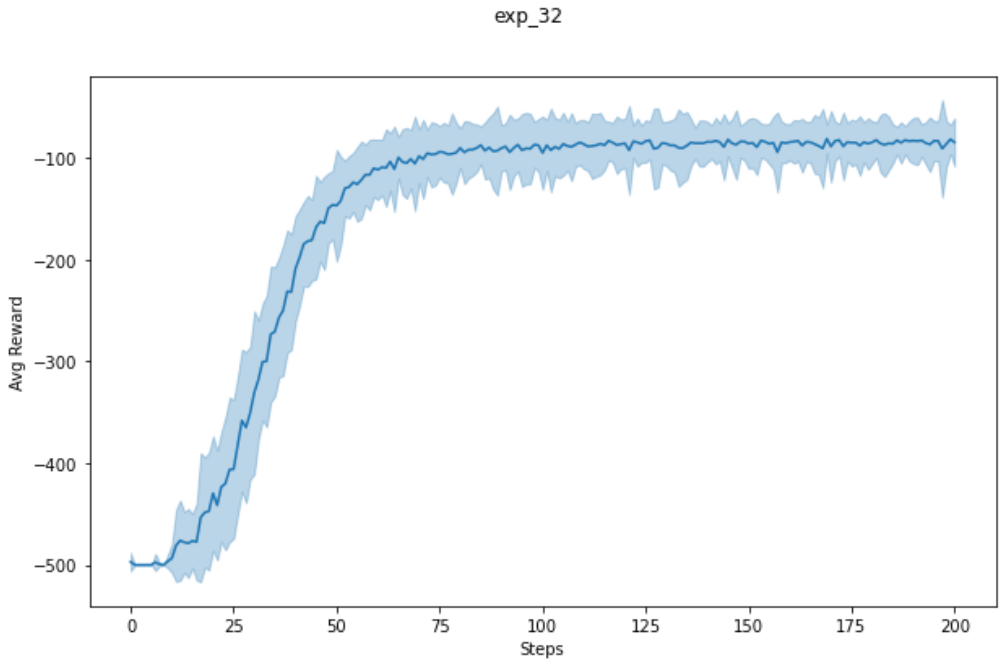
5.4. `exp_12={"name":"exp_12", "batch_size":5000,"nb_critic_updates":1}`



5.5. `exp_22={"name":"exp_22", "batch_size":5000,"nb_critic_updates":10}`

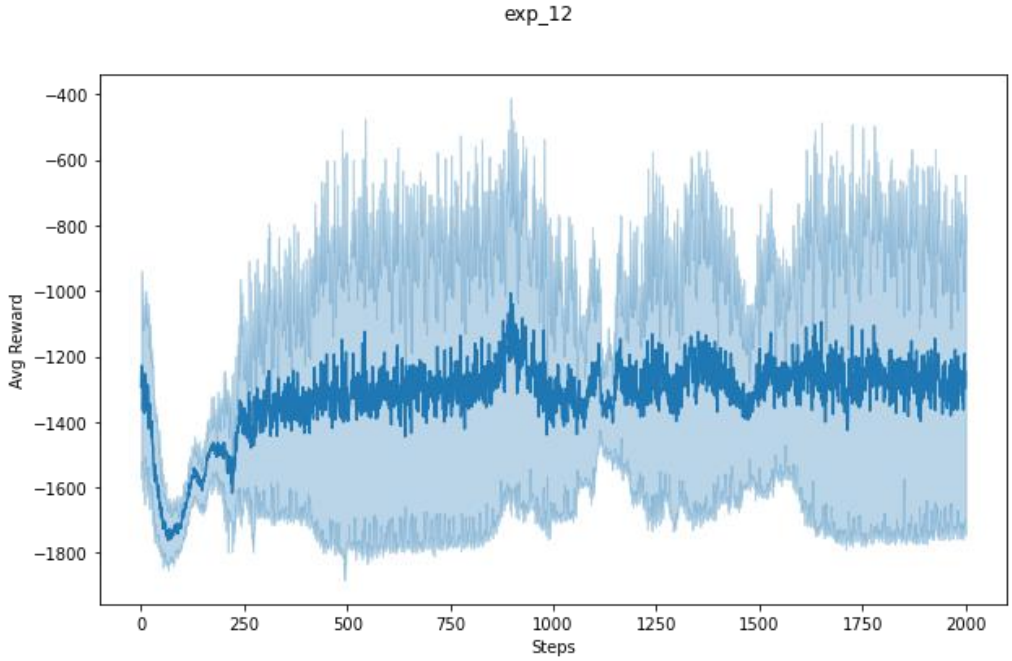


5.6. exp_32={"name":"exp_32", "batch_size":5000,"nb_critic_updates":100}

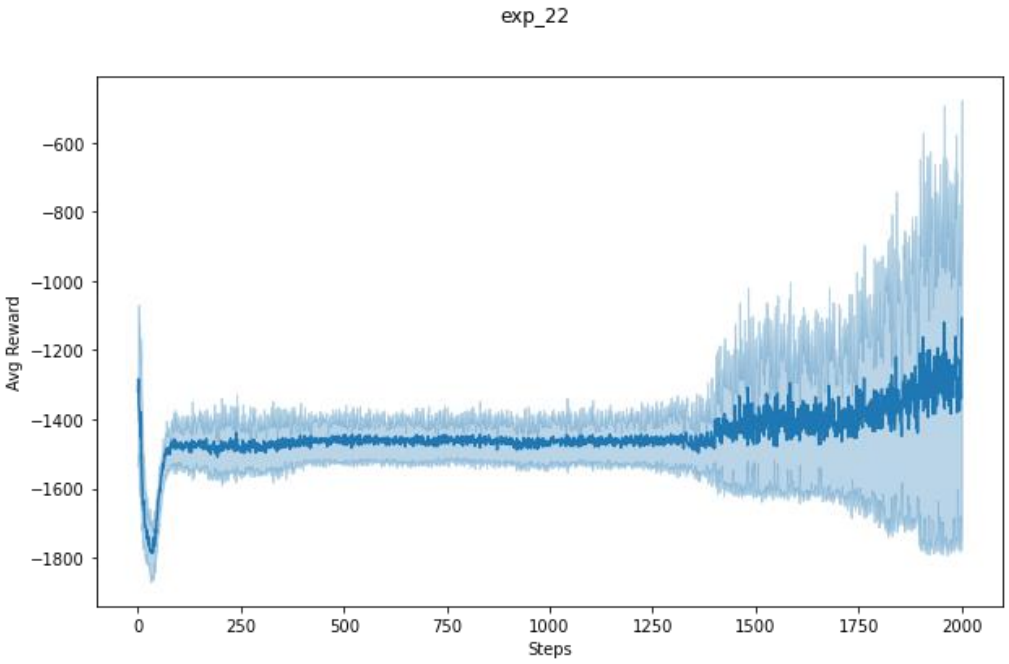


6. **Experimentos Pendulum:**

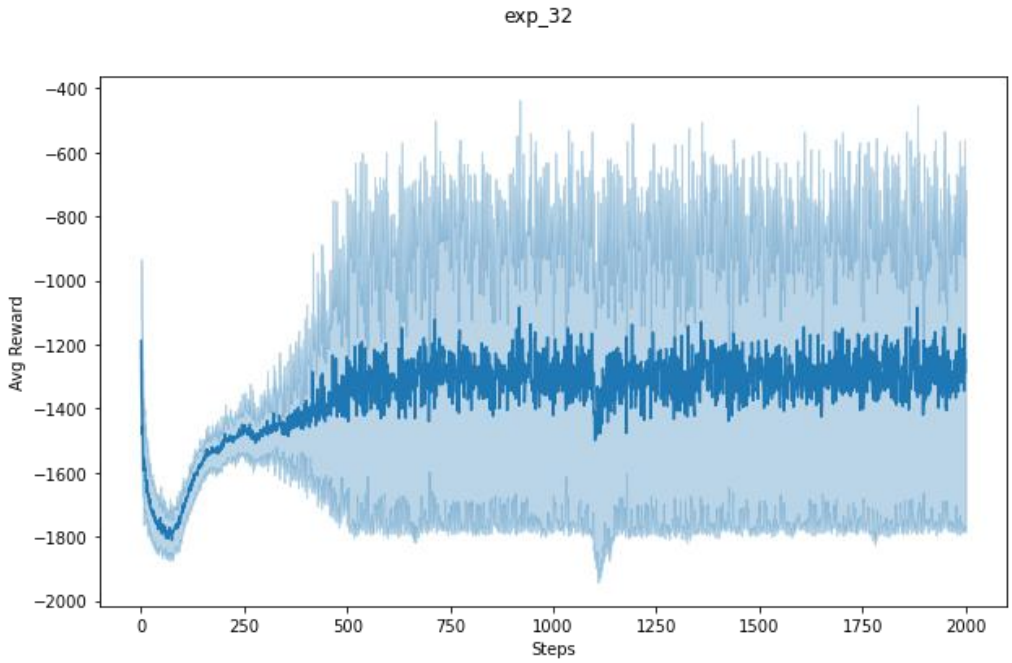
6.1. exp_12={"name":"exp_11", "batch_size":5000,
"nb_critic_updates":1,"critic_lr":0.001}



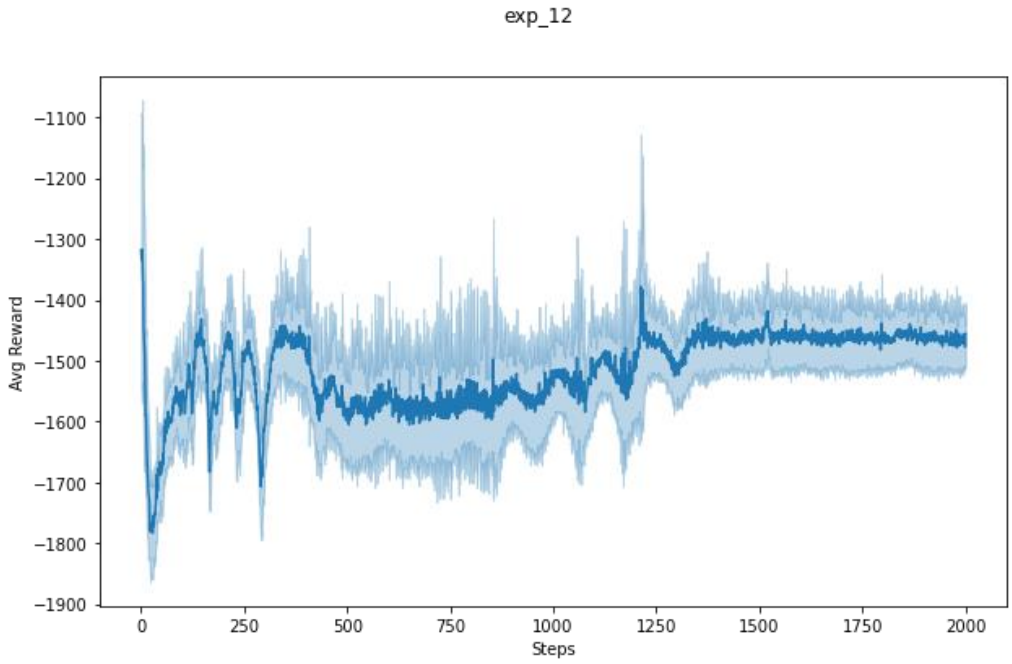
6.2. exp_22={"name":"exp_21", "batch_size":5000, "nb_critic_updates":10,"critic_lr":0.001}



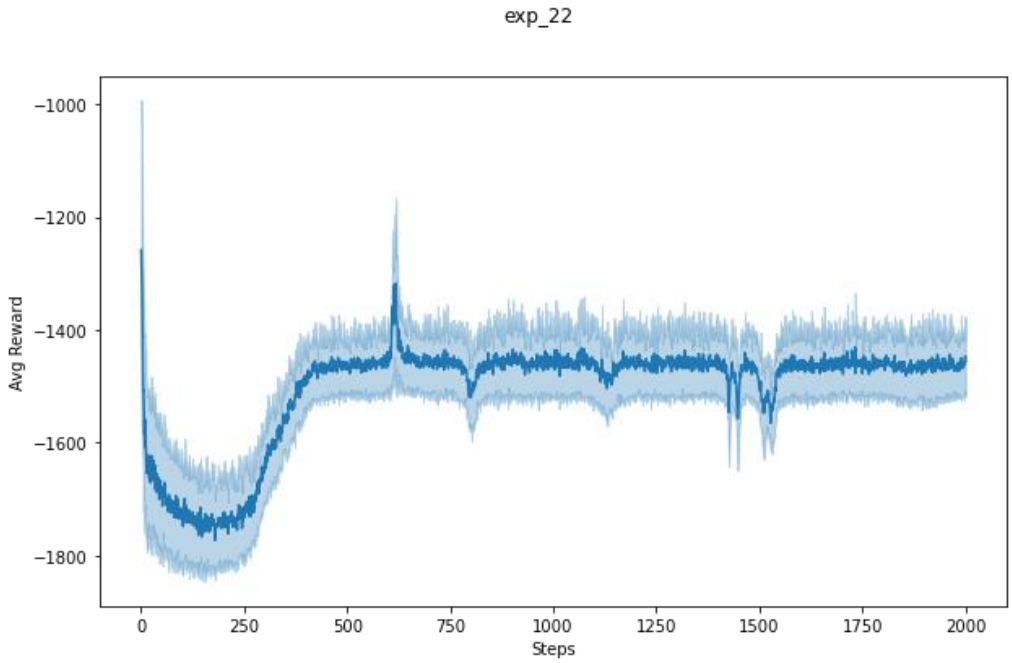
6.3. exp_32={"name":"exp_31", "batch_size":5000, "nb_critic_updates":100,"critic_lr":0.001}



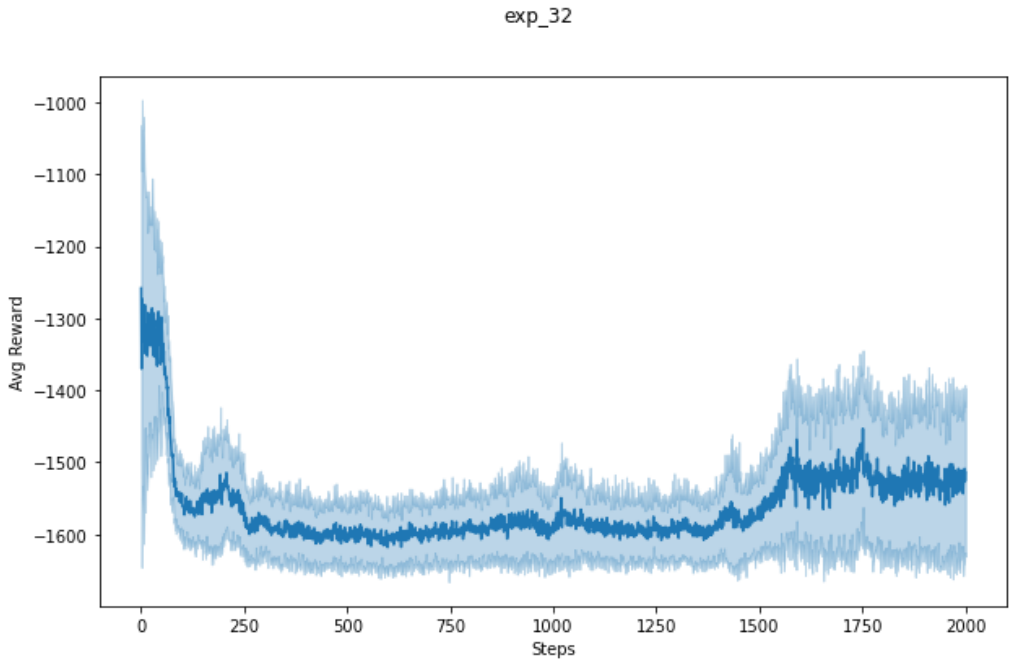
6.4. exp_12={"name":"exp_11", "batch_size":5000, "nb_critic_updates":1,"critic_lr":0.01}



6.5. exp_22={"name":"exp_21", "batch_size":5000, "nb_critic_updates":10,"critic_lr":0.01}



6.6. `exp_32={"name":"exp_31", "batch_size":5000, "nb_critic_updates":100,"critic_lr":0.01}`



7. **Análisis CartPole y Acrobot:**

A partir de los experimentos Exp_{i}1, se observa que con solo una actualización del crítico por iteración de entrenamiento no es posible obtener un buen reward que refleje un buen comportamiento en la toma de decisiones del agente. En la medida de que haya una mayor cantidad de actualizaciones, se esperaría obtener un mejor rendimiento en un número menor de steps.

Por otro lado, al comparar los experimentos Exp_{i}1 y Exp_{i}2 se observa que con un mayor tamaño de batch es posible reducir la varianza en la toma de decisiones.

8. **Análisis Pendulum:**

A partir de los Exp_{i}1 no se observa una clara mejora, esto se asocia a algún potencial error en la programación del actor que no logra aprender las acciones continuas óptimas, pero se esperaría a que el desempeño del actor mejore en la medida de que el número de actualizaciones aumente.

Por otro lado, a partir de la comparación de los Exp_{i}1 y Exp_{i}2, se observa una disminución en la varianza de los resultados al aumentar critic_lr, esto se asocia a que la red que estima la función de retorno logra barrer valores de sus pesos de una forma más rápida, encontrando valores óptimos de manera más eficaz.