**Entrega final tarea 2:** Deep Q-Network

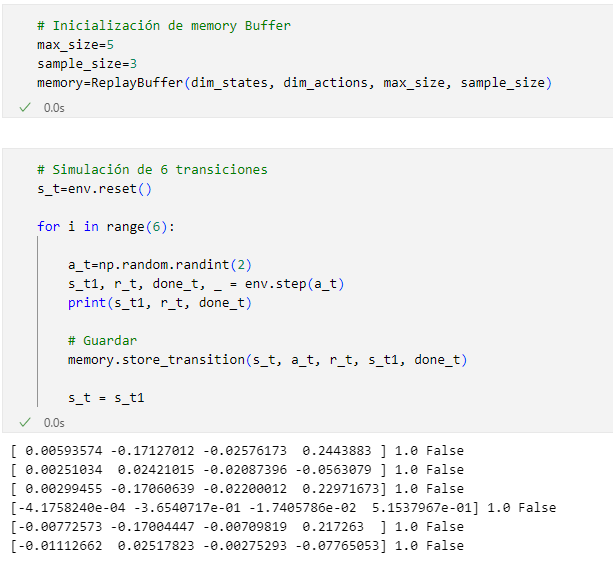
**Código:** EL7021-1

**Nombre:** José Luis Cádiz Sejas

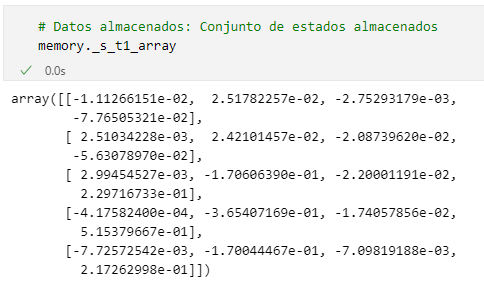
1. **Deep Q-Network (i):** Código adjunto.

1. **Replay Buffer**.
   1. Código adjunto.
   2. Código adjunto.
   3. Código adjunto.
   4. Corroboración de funcionamiento:

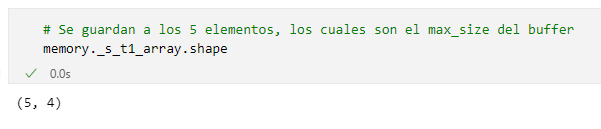
Se genera un Buffer de tamaño 5 en donde el muestreo será de tamaño 3. Se simula la obtención de 6 transiciones, las cuales se van almacenando dentro del Buffer.

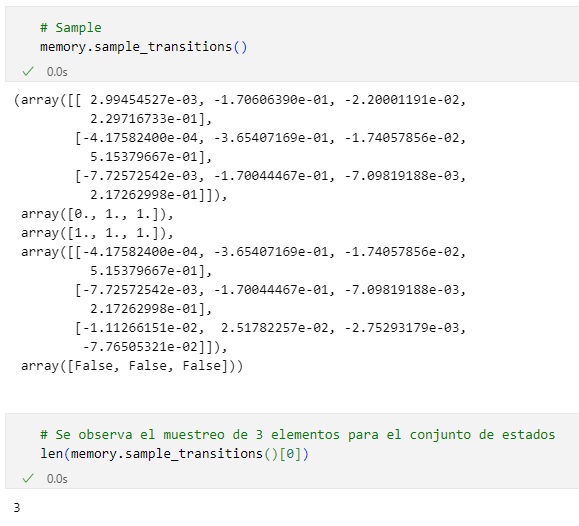


Se observa que el sexto elemento de la transición de estados coincide con el primer elemento del conjunto de estados almacenados. Esto debido a que la data se va sobrescribiendo en la medida de que hay nuevos registros.

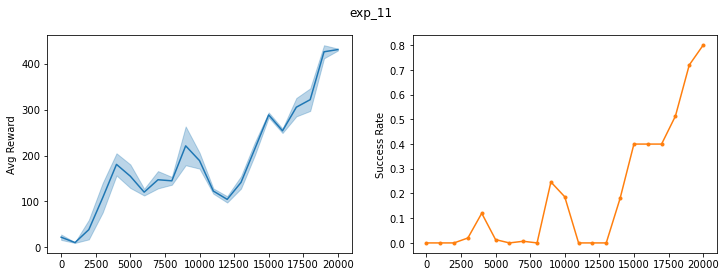


Se comprueba adicionalmente que el tamaño del muestreo es el mismo que el seteado para una variable en particular, en este caso para s\_t.

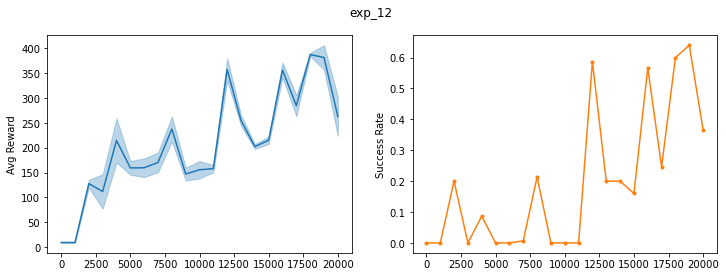




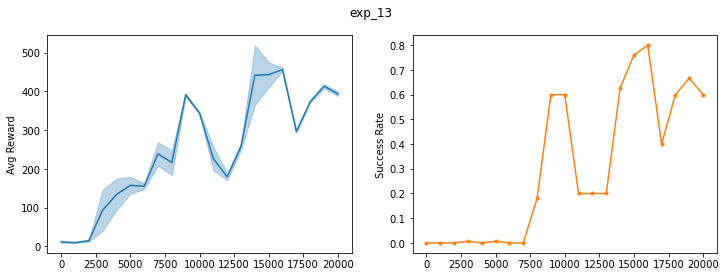
1. **Deep Q-Network (ii):** Código adjunto.
2. **Experimentos**.
   1. Código adjunto.
   2. Experimentos y análisis:
   * Exp\_11: {"name":"exp\_11","replay\_buffer\_size":1000, "batch\_size":16, "nb\_steps\_target\_replace":100}



* + Exp\_12: {"name":"exp\_12","replay\_buffer\_size":2500, "batch\_size":16, "nb\_steps\_target\_replace":100}

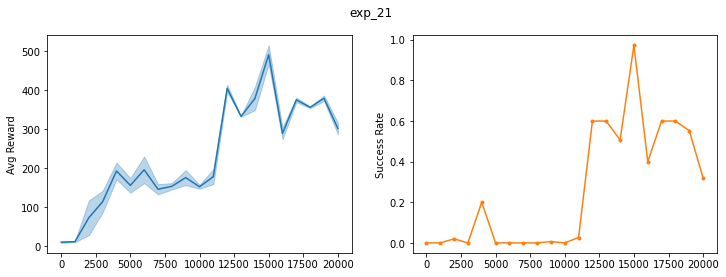


* + Exp\_13: {"name":"exp\_13","replay\_buffer\_size":5000, "batch\_size":16, "nb\_steps\_target\_replace":100}

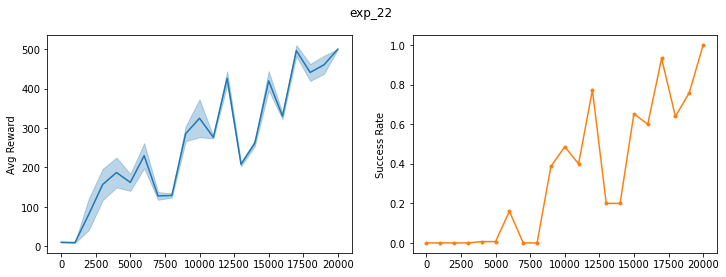


A partir del experimento 1, se puede observar que se obtiene un mejor Success Rate en un número menor de episodios cuando el Buffer posee un mayor tamaño (ver experimento 13). Esto se asocia a que es posible almacenar una mayor variabilidad de información, la cual es valiosa para el entrenamiento de la red neuronal.

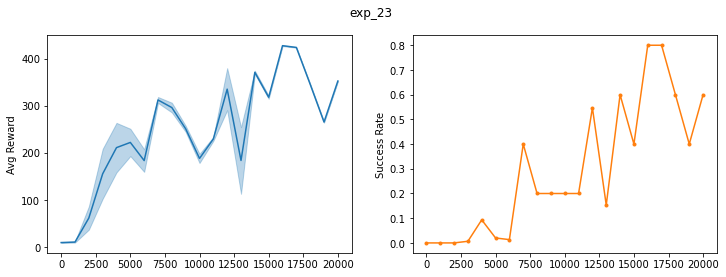
* + Exp\_21: {"name":"exp\_21","replay\_buffer\_size":5000, "batch\_size":32, "nb\_steps\_target\_replace":100}



* + Exp\_22: {"name":"exp\_22","replay\_buffer\_size":5000, "batch\_size":64, "nb\_steps\_target\_replace":100}

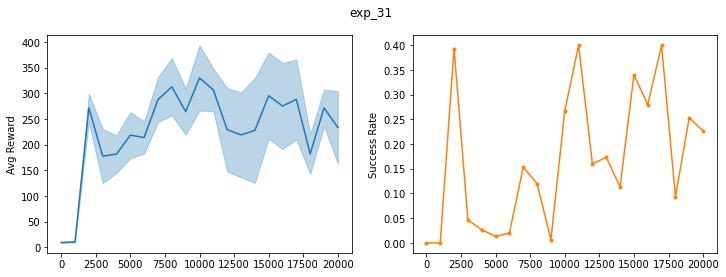


* + Exp\_23: {"name":"exp\_23","replay\_buffer\_size":5000, "batch\_size":128, "nb\_steps\_target\_replace":100}

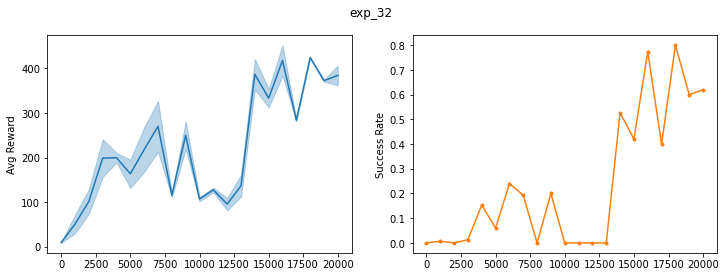


A partir del experimento 2, no se observa una mayor variabilidad en el rendimiento de la política óptima, pero si observa una disminución en la varianza del Reward obtenido cuando hay un tamaño mayor del batch (ver experimento 23), esto en la medida de que van avanzados los episodios de entrenamiento.

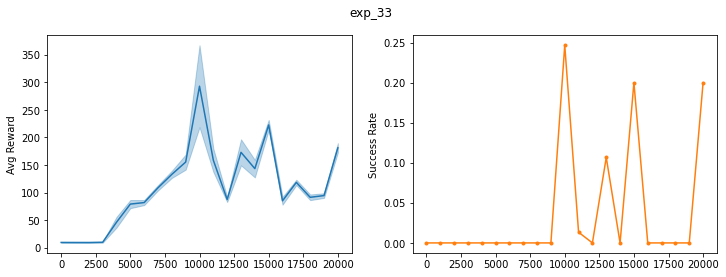
* + Exp\_31: {"name":"exp\_31","replay\_buffer\_size":5000, "batch\_size":128, "nb\_steps\_target\_replace":1}



* + Exp\_32: {"name":"exp\_32","replay\_buffer\_size":5000, "batch\_size":128, "nb\_steps\_target\_replace":100}



* + Exp\_33: {"name":"exp\_33","replay\_buffer\_size":5000, "batch\_size":128, "nb\_steps\_target\_replace":1000}



A partir del experimento 3, se observa que para un nb\_steps\_target\_replace

igual 1 (ver experimento 31) se genera una inestabilidad en el aprendizaje de la red neuronal, esto debido a que estamos evaluando la red con las predicciones de esta misma. La inestabilidad del aprendizaje de la red se puede ver en la alta varianza del Reward obtenido.

Por otro lado, si nb\_steps\_target\_replace posee un valor muy alto, el proceso de aprendizaje tampoco es bueno (ver experimento 33). A partir de esto, se concluye que para lograr un buen Success Rate nb\_steps\_target\_replace no puede ser un valor ni muy bajo ni tampoco muy alto (ver experimento 32).

**En definitiva, para que el agente pueda aprender las políticas óptimas del ambiente en cuestión, es importante tener un buen tamaño de Replay Buffer y un valor intermedio para la actualización de los parámetros de la red neuronal target. Por otro lado, el tamaño del batch podría afectar en la varianza del Reward pero no de manera significativa en el rendimiento en termino de Success Rate.**