

Instrucciones:

- Esta es una actividad en grupos de 3 personas máximo
- No se permitirá ni se aceptará cualquier indicio de copia. De presentarse, se procederá según el reglamento correspondiente.
- Tendrán hasta el día indicado en Canvas.

Task 1

Responda a cada de las siguientes preguntas de forma clara y lo más completamente posible.

1. Defina y explique qué "expected sarsa"
 - a. ¿Cómo se diferencia de "sarsa"?
 - b. ¿Para qué sirven las modificaciones que se hacen sobre "sarsa"?
2. Defina y explique qué es "n-step TD"
 - a. ¿Cómo se diferencia de TD(0)?
 - b. ¿Cuál es la utilidad de esta modificación?
 - c. ¿Qué usa como objetivo?
3. ¿Cuál es la diferencia entre SARSA y Q-learning?

Task 2

En este laboratorio, compararán el rendimiento de SARSA y Q-Learning, dos algoritmos de aprendizaje de refuerzo populares, utilizando el entorno CliffWalking-v0 de la biblioteca Gymnasium. Analizará y graficará las recompensas por episodio y responderá preguntas para profundizar su comprensión de las diferencias entre estos algoritmos.

Instrucciones

1. Instalar las bibliotecas necesarias:
 - a. Asegúrese de tener Gymnasium y matplotlib instalados en su entorno de Python.
 - b. Instale estas bibliotecas usando pip si es necesario.
2. Importe las bibliotecas necesarias:
 - a. Importe las bibliotecas de Python necesarias para configurar el entorno y representar gráficamente los resultados.
3. Inicialice el entorno:
 - a. Cree el entorno CliffWalking-v0 usando Gymnasium.
 - b. Establezca parámetros comunes como la tasa de aprendizaje, el factor de descuento, epsilon para la política epsilon-greedy y la cantidad de episodios.
4. Implementar SARSA:
 - a. Escriba una función para implementar el algoritmo SARSA.
 - b. Asegúrese de que la función devuelva la recompensa acumulada por episodio.
5. Implementar Q-Learning:
 - a. Escriba una función para implementar el algoritmo Q-Learning.
 - b. Asegúrese de que esta función también devuelva la recompensa acumulada por episodio.
6. Ejecutar ambos algoritmos:
 - a. Ejecute las funciones SARSA y Q-Learning que implementaste.
 - b. Recopilar y almacenar las recompensas por episodio para cada algoritmo.
7. Gráfico de los resultados:
 - a. Crear un gráfico que represente las recompensas por episodio tanto para SARSA como para Q-Learning en el mismo gráfico.
8. Analice los resultados:

Reinforcement Learning - Laboratorio 4 -

Semestre II - 2024

-
- a. Observe los resultados graficados y compare el rendimiento de SARSA y Q-Learning a lo largo de los episodios.
 - b. Considere cómo cada algoritmo equilibra la exploración y la explotación.
9. Responda las siguientes preguntas:
- a. P1: ¿Qué diferencias observa en los patrones de recompensa entre SARSA y Q-Learning?
 - b. P2: ¿Por qué Q-Learning tiende a ser más optimista en sus actualizaciones de valor Q en comparación con SARSA?
 - c. P3: ¿Cómo afecta la naturaleza "on-policy" de SARSA a su proceso de aprendizaje en comparación con la naturaleza "off-policy" de Q-Learning?
 - d. P4: Según las penalizaciones y recompensas del entorno, ¿qué algoritmo parece aprender el camino más seguro? ¿Por qué?
 - e. P5: ¿Cómo podría afectar la disminución de ϵ con el tiempo al rendimiento de cada algoritmo?

Notas adicionales:

- Considere experimentar con diferentes hiperparámetros (p. ej., α , γ , ϵ) para ver cómo afectan el rendimiento de ambos algoritmos.
- Reflexione sobre cómo las diferencias teóricas entre SARSA y Q-Learning se traducen en diferencias prácticas en su rendimiento en este entorno.

Preguntas para responder:

1. ¿Cuál es el valor estimado de mantener diferentes niveles de existencias para cada producto?
2. ¿Cómo afecta el valor ϵ en la política blanda al rendimiento?
3. ¿Cuál es el impacto de utilizar el aprendizaje fuera de la política en comparación con el aprendizaje dentro de la política?

Entregas en Canvas

1. Documento PDF con las respuestas a cada task
 - a. Pueden exportar el JN como PDF si trabajan con esto.
2. Código de la implementación del Task 2
 - a. Si trabaja con JN deje evidencia de la última ejecución
 - b. Caso contrario, deje en comentarios el valor resultante

Evaluación

1. [1.5 pts] Task 1 (0.5 cada pregunta)
2. [3.5 pts] Task 2