**GUIÓN**

* ¿Qué es el aprendizaje automático?
* ¿Qué es el aprendizaje por refuerzo?
  + Estados.
  + Acciones.
  + Agente.
  + Definición.
  + Tipos.
* Q-Learning.
  + Tabla Q.
    - ¿Qué la compone?
    - ¿Cómo se usa?
  + Conceptos de explotación y exploración.
    - Definición.
    - La exploración reduce conforme aprendemos.
* Algoritmo.
  + Matemáticas de explotación (se aprende también).
    - Función de recompensa.
    - Parámetro y.
    - Parámetro alfa. Función reducción de aprendizaje.
  + Matemáticas de exploración.
    - Método voraz.
    - Función de exploración
* Conclusiones.
  + Necesita mucha exploración.
  + Aunque actuemos de forma óptima se pueden cometer errores.
  + Aunque no actúe de forma óptima, también puede tener éxito.
  + Tarda mucho tiempo.

**¿Qué es el aprendizaje automático? (17 s)**

El aprendizaje automático es un tipo de inteligencia artificial que proporciona a las computadoras (agentes) la capacidad de aprender sin ser programadas explícitamente. Decimos que una máquina aprende una **tarea** si, en base a una **experiencia,** consigue mejorar su **rendimiento** en ella.

**¿Qué es el aprendizaje por refuerzo? (45 s)**

El aprendizaje por refuerzo es un método de aprendizaje automático en el cual el agente trata de aprender una tarea mediante prueba y error. Involucra a un agente, un conjunto de estados (S) y un conjunto de acciones (A) por cada estado.

Cuando el agente realiza una acción, puede cambiar de estado y recibe una recompensa o un castigo, en función de lo bien o lo mal que lo haya hecho.

Hay diferentes tipos de aprendizaje por refuerzo:

* En el aprendizaje activo, el agente ejecuta las acciones y aprende de las consecuencias de esas acciones.
* Mientras que en el aprendizaje pasivo, el agente no actúa, sino que aprende de las acciones que se le indican.

**Q-Learning (26 s)**

El Q-Learning es un método de aprendizaje activo que se basa en una tabla Estados-Acción que llamamos tabla-Q. El contenido de la tabla Q es una estimación de lo buena o mala que resulta realizar una acción en cada estado.

(Imagen de tabla Q)

Dado un estado concreto, la acción que ejecuta el agente es aquella que le da la mayor recompensa, es decir, aquella que tiene la mejor estimación.

**Conceptos de explotación y exploración (30 s)**

Cuando el agente está aprendiendo, puede actuar de dos maneras: explotación o exploración.

* La explotación consiste en utilizar lo que hemos aprendido.
* Y la exploración consiste en probar cosas nuevas, que pueden ser acciones diferentes a las que creemos que son mejores.

Al principio, el agente no sabe nada y necesita explorar mucho. Sin embargo, a medida que aprende, la exploración se reduce y se explotan los conocimientos adquiridos.

**Matemáticas de explotación (1 m 06 s)**

* Explotar los conocimientos adquiridos consiste en lo siguiente:
* Inicialmente, el agente se encuentra en un estado que llamamos estado actual. Lo primero que debe hacer es ejecutar la acción que tiene la mejor estimación en la Tabla-Q.
* Esto va a provocar una serie de consecuencias en su entorno y provocará un cambio de estado en el agente, es lo que llamamos estado siguiente.
* Según el estado actual, la acción ejecutada y el estado al que llega, la función de recompensa R devuelve un valor. Será mayor cuanto mejor haya sido la decisión del agente.
* De esta forma, se recompensan las acciones buenas y se penalizan las acciones malas.
* Desde el estado al que ha llegado el agente se evalúan todas las posibles acciones y se elige la que tenga mayor Q-valor para tratar de predecir cómo de buenas van a ser las acciones futuras.
* Con la recompensa obtenida y la estimación de la mejor acción futura, se obtiene el valor de la muestra.
* La tabla Q se actualiza combinando lo que se sabía previamente y el valor de la muestra.
* Para que el algoritmo funcione correctamente, es necesario ponderar algunos de los sumandos de la fórmula.
* Gamma es un valor que indica el grado de confianza en nuestras predicciones. Cuanto mayor sea, más importancia le daremos a las acciones futuras frente a las recompensas que estamos obteniendo actualmente.
* Alfa es el grado de aprendizaje o la confianza que tenemos en los conocimientos previos. Un alfa elevado indica que confiamos más en lo que acabamos de aprender que en lo que sabíamos de antes.
* Para un aprendizaje óptimo, al principio, alpha debe tener un valor muy alto e ir decreciendo conforme el agente vaya aprendiendo. De esta forma, el agente empezará a confiar más en lo que sabe que en lo que experimenta. Llegará un momento en el que tenga suficiente experiencia y no necesitará aprender.
* Esto se puede implementar con una función matemática que reduzca de forma gradual. Una posible opción sería función exponencial, como la que se muestra en la imagen, que decremente a medida que pasa el tiempo.
* Aplicar Alpha consiste en hacer la suma ponderada del valor que se ha obtenido de la muestra y el q-valor que había almacenado en la tabla-q. El resultado será el nuevo q-valor para ese estado y esa acción.

**Matemáticas de exploración (1 m 10 s)**

La exploración se puede introducir de muchas formas, dos de ellas son el método voraz y las funciones de exploración.

* Método voraz: el método voraz consiste en utilizar una probabilidad para decidir se explotamos o exploramos. Esta probabilidad también tiene que decrementar con el tiempo para asegurar un aprendizaje óptimo.
* Método de exploración pondera el Q-valor con el número de veces que se ha realizado la acción. De manera que las acciones que se hayan realizado menos veces tienen una ponderación mayor.

**Conclusiones (10 s)**

Finalmente, para un funcionamiento óptimo, se requiere de mucha exploración y tiempo. Además, aunque el agente no actúe de forma óptima puede tener éxito, o incluso habiendo aprendido de forma racional, puede equivocarse.

En definitiva, cometer errores es de humanos y agentes.