**Pruebas Agente coche**

Base:

Partimos de que el coche consigue recompensas en las siguientes situaciones:

1. Si alcanzara el checkpoint correcto, ganará 0.25+un plus conseguido por el menor tiempo que haya tardado ((maxTiempo-tiempoEmpleado)/maxTiempo)
2. Si Llega al checkpoint final se le sumará 0.5 más el plus del tiempo
3. Se le suma entre 0 y 0.00006 por lo bien alineado que esté con el siguiente checkpoint, en función de su rotación Y.

Perderá recompensa en las siguientes ocasiones:

1. Cada acción que pase pierde una recompensa de 0.0006
2. Cada vez que pase el tiempo máximo y no haya alcanzado el siguiente checkpoint el agente morirá y perderá 1 de recompensa
3. Cada vez que dé más de 30 pasos hacia atrás morirá y perderá 1 de recompensa

**Cambios aplicados:**

Se han añadido más observaciones, entre ellas la velocidad en el eje Z del vehículo, su rotación Y junto a la del checkpoint. Estás se añaden a las que ya había de la distancia al siguiente checkpoint.

Se intentará aplicar imitation learning, hasta ahora solo se han conseguido peores resultados.

Se ha disminuido la recompensa negativa proporcionada al perder por tiempo de 1 a 0.5, par aque el agente prefiera perder por tiempo a perder por ir marcha atrás

Se ha añadido una recompensa si se detecta que el checkpoint está girado hacia un lado y el coche está girando hacia ese lado

Cambio el espacio de acciones a discretas. De esta manera sigue funcionando mal, se vuelve al anterior espacio de decisiones continuas.

**Observaciones:**

A pesar de la gran recompensa negativa que da el ir para atrás el agente sigue queriendo ir marcha atrás nada más aparecer.

Se está usando un espacio de acciones continuo porque tanto el giro como la aceleración/freno es progresiva, no obstante, para la grabación de demos se stá usando teclado, con lo que, a pesar de que aparezca como continuo, solo se usan tres valores, -1, 0 y 1. No sé si esto puede estar dando lugar a fallos.