

## PRIMER ALGORITMO

<https://github.com/AppliedDataSciencePartners/DeepReinforcementLearning>

### PRIMER ENTRENAMIENTO

Adaptamos el tablero del juego de la siguiente manera: cada jugador tiene un tablero de 113 posiciones (226 en total). En cada iteración le indicamos nosotros al programa en qué posición de la matriz hay que poner el número que elija, hicimos esto por cuestiones de ahorro de espacio en el programa. El orden que le hemos indicado al programa es el siguiente:

2	3	9
7	1	8
6	4	5

La idea es que comience construyendo la matriz por el centro, después el elemento superior izquierdo y el superior central y a partir de ahí vaya construyendo por pares.

Por ejemplo, supongamos que tenemos que la acción = 0. Esto significa que el jugador 1 ha elegido la posición central para poner su valor numérico. Si por ejemplo tenemos la acción = 115 (la 114 es la primera acción del jugador 2), significa que el jugador ha elegido la posición de arriba a la izquierda de la matriz.

Después de elegir una posición, el jugador elige un número al azar entre 1 y 113, aprendiendo poco a poco cuáles debe poner y cuáles no.

### RESULTADOS ENTRENAMIENTO

Tras el entrenar durante 2 días, realizamos la inferencia, que consiste en que el modelo juegue varias partidas contra sí mismo. Los resultados fueron los siguientes:

#### EPISODE 1

2021-09-24 11:04:51,982 INFO =====

2021-09-24 11:04:51,985 INFO Matriz 1

2021-09-24 11:04:51,985 INFO ['105', '97', '5']

2021-09-24 11:04:51,985 INFO ['96', '73', '4']

2021-09-24 11:04:51,985 INFO ['3', '1', '2']

2021-09-24 11:04:51,985 INFO Score 1: 33.763074854429945

2021-09-24 11:04:51,985 INFO Matriz 2

2021-09-24 11:04:51,985 INFO ['99', '50', '51']

2021-09-24 11:04:51,986 INFO ['69', '75', '2']

2021-09-24 11:04:51,986 INFO ['83', '1', '32']

2021-09-24 11:04:51,986 INFO Score 2: 21.07431077481741

2021-09-24 11:04:51,986 INFO -----

2021-09-24 11:04:51,986 INFO player2 WINS!

2021-09-24 11:04:51,986 INFO =====

EPISODE 2

2021-09-24 11:06:28,849 INFO =====

2021-09-24 11:06:28,852 INFO Matriz 1

2021-09-24 11:06:28,852 INFO ['103', '1', '3']

2021-09-24 11:06:28,852 INFO ['101', '74', '105']

2021-09-24 11:06:28,852 INFO ['79', '2', '37']

2021-09-24 11:06:28,853 INFO Score 1: 41.46529707665217

2021-09-24 11:06:28,853 INFO Matriz 2

2021-09-24 11:06:28,853 INFO ['70', '1', '14']

2021-09-24 11:06:28,853 INFO ['20', '113', '2']

2021-09-24 11:06:28,853 INFO ['6', '75', '102']

2021-09-24 11:06:28,853 INFO Score 2: 19.236399956317683

2021-09-24 11:06:28,853 INFO -----

2021-09-24 11:06:28,853 INFO player1 WINS!

2021-09-24 11:06:28,853 INFO =====

EPISODE 3

2021-09-24 11:08:04,668 INFO =====

2021-09-24 11:08:04,671 INFO Matriz 1

2021-09-24 11:08:04,671 INFO ['100', '77', '73']

2021-09-24 11:08:04,671 INFO ['102', '72', '111']

2021-09-24 11:08:04,671 INFO ['105', '62', '61']

2021-09-24 11:08:04,671 INFO Score 1: 18.6708378234424

2021-09-24 11:08:04,672 INFO Matriz 2

2021-09-24 11:08:04,672 INFO ['1', '2', '7']

2021-09-24 11:08:04,672 INFO ['75', '106', '6']

2021-09-24 11:08:04,672 INFO ['5', '3', '4']

2021-09-24 11:08:04,672 INFO Score 2: 33.15902677360013

2021-09-24 11:08:04,672 INFO -----

2021-09-24 11:08:04,672 INFO player2 WINS!

2021-09-24 11:08:04,672 INFO =====

#### EPISODE 4

2021-09-24 11:09:40,356 INFO =====

2021-09-24 11:09:40,360 INFO Matriz 1

2021-09-24 11:09:40,360 INFO ['1', '2', '90']

2021-09-24 11:09:40,360 INFO ['61', '79', '5']

2021-09-24 11:09:40,360 INFO ['21', '3', '4']

2021-09-24 11:09:40,360 INFO Score 1: 40.72827276704114

2021-09-24 11:09:40,361 INFO Matriz 2

2021-09-24 11:09:40,361 INFO ['60', '71', '95']

2021-09-24 11:09:40,361 INFO ['70', '81', '89']

2021-09-24 11:09:40,361 INFO ['28', '54', '9']

2021-09-24 11:09:40,361 INFO Score 2: 14.575336452028893

2021-09-24 11:09:40,361 INFO -----

2021-09-24 11:09:40,361 INFO player1 WINS!

2021-09-24 11:09:40,361 INFO =====

#### EPISODE 5

2021-09-24 11:11:13,673 INFO =====

2021-09-24 11:11:13,676 INFO Matriz 1

2021-09-24 11:11:13,676 INFO ['1', '64', '3']

2021-09-24 11:11:13,676 INFO ['107', '96', '111']

2021-09-24 11:11:13,676 INFO ['83', '21', '2']

2021-09-24 11:11:13,676 INFO Score 1: 32.253786062189114

2021-09-24 11:11:13,677 INFO Matriz 2

2021-09-24 11:11:13,677 INFO ['42', '20', '6']

2021-09-24 11:11:13,677 INFO ['4', '106', '5']

2021-09-24 11:11:13,677 INFO ['3', '1', '2']

2021-09-24 11:11:13,677 INFO Score 2: 27.841269841269842

2021-09-24 11:11:13,677 INFO -----

2021-09-24 11:11:13,677 INFO player1 WINS!

2021-09-24 11:11:13,677 INFO =====

## EPISODE 6

2021-09-24 11:12:47,723 INFO =====

2021-09-24 11:12:47,726 INFO Matriz 1

2021-09-24 11:12:47,726 INFO ['103', '106', '2']

2021-09-24 11:12:47,726 INFO ['99', '97', '96']

2021-09-24 11:12:47,727 INFO ['1', '107', '100']

2021-09-24 11:12:47,727 INFO Score 1: 28.872450705930017

2021-09-24 11:12:47,727 INFO Matriz 2

2021-09-24 11:12:47,727 INFO ['93', '2', '51']

2021-09-24 11:12:47,727 INFO ['5', '1', '85']

2021-09-24 11:12:47,727 INFO ['4', '3', '20']

2021-09-24 11:12:47,727 INFO Score 2: 8.515321972332261

2021-09-24 11:12:47,727 INFO -----

2021-09-24 11:12:47,727 INFO player1 WINS!

2021-09-24 11:12:47,727 INFO =====

## EPISODE 7

2021-09-24 11:14:24,316 INFO =====

2021-09-24 11:14:24,320 INFO Matriz 1

2021-09-24 11:14:24,320 INFO ['1', '104', '73']

2021-09-24 11:14:24,320 INFO ['80', '72', '23']

2021-09-24 11:14:24,320 INFO ['108', '61', '40']

2021-09-24 11:14:24,320 INFO Score 1: 22.88888888888889

2021-09-24 11:14:24,321 INFO Matriz 2

2021-09-24 11:14:24,321 INFO ['1', '2', '65']

2021-09-24 11:14:24,321 INFO ['6', '106', '56']

2021-09-24 11:14:24,321 INFO ['5', '3', '4']

2021-09-24 11:14:24,321 INFO Score 2: 33.112042646616004

2021-09-24 11:14:24,321 INFO -----

2021-09-24 11:14:24,321 INFO player2 WINS!

2021-09-24 11:14:24,321 INFO =====

## EPISODE 8

2021-09-24 11:15:58,739 INFO =====

2021-09-24 11:15:58,741 INFO Matriz 1

2021-09-24 11:15:58,741 INFO ['71', '103', '1']

2021-09-24 11:15:58,741 INFO ['100', '80', '104']

2021-09-24 11:15:58,741 INFO ['113', '36', '24']

2021-09-24 11:15:58,741 INFO Score 1: 33.32984126984127

2021-09-24 11:15:58,742 INFO Matriz 2

2021-09-24 11:15:58,742 INFO ['93', '2', '73']

2021-09-24 11:15:58,742 INFO ['5', '72', '106']

2021-09-24 11:15:58,742 INFO ['4', '1', '3']

2021-09-24 11:15:58,742 INFO Score 2: 33.825693440266804

2021-09-24 11:15:58,742 INFO -----

2021-09-24 11:15:58,742 INFO player1 WINS!

2021-09-24 11:15:58,742 INFO =====

EPISODE 9

2021-09-24 11:17:33,452 INFO =====

2021-09-24 11:17:33,455 INFO Matriz 1

2021-09-24 11:17:33,455 INFO ['104', '83', '5']

2021-09-24 11:17:33,455 INFO ['4', '62', '97']

2021-09-24 11:17:33,455 INFO ['3', '1', '2']

2021-09-24 11:17:33,455 INFO Score 1: 34.66666666666667

2021-09-24 11:17:33,456 INFO Matriz 2

2021-09-24 11:17:33,456 INFO ['69', '72', '99']

2021-09-24 11:17:33,456 INFO ['2', '3', '33']

2021-09-24 11:17:33,456 INFO ['14', '66', '74']

2021-09-24 11:17:33,456 INFO Score 2: 7.7656147307413885

2021-09-24 11:17:33,456 INFO -----

2021-09-24 11:17:33,456 INFO player2 WINS!

2021-09-24 11:17:33,456 INFO =====

EPISODE 10

2021-09-24 11:19:08,255 INFO =====

2021-09-24 11:19:08,257 INFO Matriz 1

2021-09-24 11:19:08,257 INFO ['1', '70', '2']

2021-09-24 11:19:08,257 INFO ['63', '73', '45']

2021-09-24 11:19:08,258 INFO ['98', '74', '95']

2021-09-24 11:19:08,258 INFO Score 1: 20.768624331123878

2021-09-24 11:19:08,258 INFO Matriz 2

2021-09-24 11:19:08,258 INFO ['1', '2', '53']

2021-09-24 11:19:08,258 INFO ['5', '111', '39']

2021-09-24 11:19:08,258 INFO ['13', '3', '4']

2021-09-24 11:19:08,258 INFO Score 2: 40.48648447488753

2021-09-24 11:19:08,258 INFO -----

2021-09-24 11:19:08,258 INFO player2 WINS!

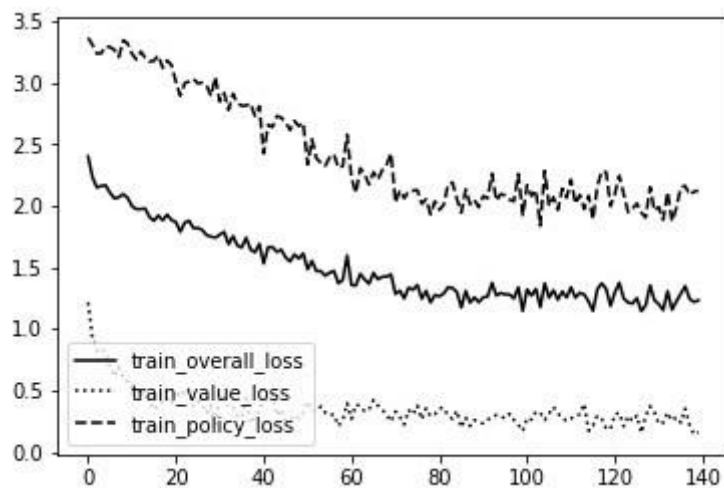
Las puntuaciones que obtiene no superan el 50/100 en ningún caso.

Si calculamos qué porcentaje de cada instrucción proviene de cada tipo de regla, vemos que el programa atiende mucho más a las reglas materiales que a las formales. Y dentro de las formales es más probable que se cumpla la de los múltiplos de 6 que cualquier otro.

Episodio	Matriz	Puntos	Reglas Materiales(%)			Reglas formales(%)					
			Primos	Impares	Total	Sumas	Serie pares	Pareja pares	Múltiplo de 6	Trios asimétricos	Total
1	1	33,76	98,72	1,27	<b>100</b>	0	0	0	0	0	<b>0</b>
	2	21,07	79,09	8,26	<b>87,35</b>	0	0	0	12,65	0	<b>12,65</b>
2	1	41,46	93,79	1,04	<b>94,82</b>	0	0	0	1,61	3,57	<b>5,18</b>
	2	19,24	86,64	2,20	<b>88,84</b>	0	0	0	3,47	7,69	<b>11,16</b>
3	1	18,67	59,51	6,82	<b>66,33</b>	22,95	7,14	0	3,57	0	<b>33,67</b>
	2	33,16	83,77	1,29	<b>85,06</b>	12,92	0	0	2,01	0	<b>14,94</b>
4	1	40,73	81,84	1,09	<b>82,93</b>	10,52	0	0	3,27	3,27	<b>17,06</b>
	2	14,58	76,23	9,04	<b>85,27</b>	0	0	0	4,57	10,15	<b>14,73</b>
5	1	32,25	86,12	2,63	<b>88,75</b>	0	0	0	2,07	9,18	<b>11,24</b>
	2	27,84	79,81	0	<b>79,82</b>	15,39	4,79	0	0	0	<b>20,18</b>
6	1	28,87	96,20	1,48	<b>97,69</b>	0	0	0	1,31	0	<b>2,31</b>
	2	8,52	86,99	5,18	<b>92,17</b>	0	0	0	7,83	0	<b>7,83</b>

No se le da ningún punto al programa por los números 1,2,3, ya que son valores que no nos interesa que aparezcan si quiera.

La función de aprendizaje, para medir lo que ha aprendido el programa es la siguiente:



## SEGUNDO ENTRENAMIENTO

Aumentamos el tiempo de entrenamiento del modelo a 4 días. Las matrices que generó el modelo eran de puntuación baja. Sin embargo, sí que generó alguna matriz que estaban relativamente bien construidas. Por ejemplo:

Matriz:

['107', '61', '59']

['101', '103', '79']

['13', '89', '73']

Puntos totales: 73.39428571428573

Materiales:

- Primos: 50
- Impares: 0
- Total: 50

Formales:

- Sumas: 8.571428571428571
- Serie pares: 1.3333333333333333
- Pareja pares: 6.666666666666667
- Múltiplos de 6: 6.0
- Tríos asimétricos: 0.8228571428571427
- Total: 23.394285714285715

Esta es la matriz de mayor puntuación que generó el modelo en este entrenamiento. Como podemos ver, todos los números de dicha matriz son primos, sin embargo, las reglas de las sumas de tríos, las diferencias entre pares, etc, le cuesta más. La conclusión fue que el modelo

entiende bastante bien las reglas formales, pero no tanto las materiales, puesto que son más complejas. Otro ejemplo:

Matriz:

['83', '61', '89']

['17', '103', '79']

['67', '73', '13']

Puntos totales: 66.28571428571429

Materiales:

- Primos: 50
- Impares: 0
- Total: 50

Formales:

- Sumas: 4.285714285714286
- Serie pares: 0.0
- Pareja pares: 6.666666666666667
- Múltiplos de 6: 5.333333333333333
- Tríos asimétricos: 0.0
- Total: 16.285714285714285

Se observa lo que habíamos comentado anteriormente, las reglas formales las entiende bastante bien, llegando a tener la puntuación máxima en varias ocasiones, pero no es el caso de las reglas materiales, que al ser más complejas le cuesta más a la red neuronal entenderlas.

### TERCER ENTRENAMIENTO

Después volvimos a realizar otro entrenamiento, pero corrigiendo un error: de la forma en la que estaba estructurado el programa, las matrices que contenían ceros obtenían puntos. De forma que la máquina iba perdiendo puntos a medida que rellenaba la matriz.

Al entrenarlo de esta manera, hubo una mejora en los resultados, a pesar de que, por un fallo de memoria, no fue capaz de entrenar durante el mismo tiempo que en el caso anterior.

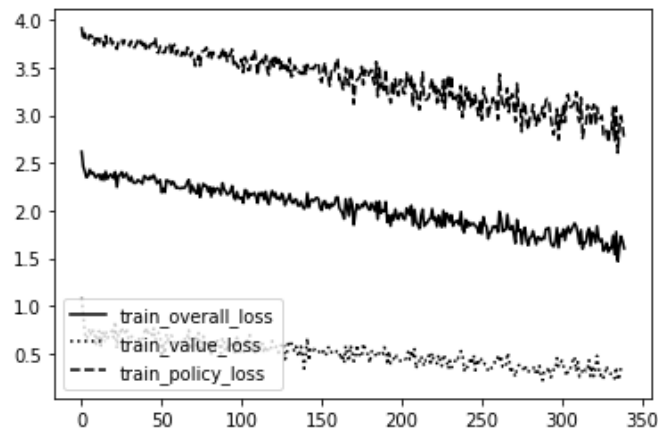
La mejor matriz que generó durante todo el entrenamiento fue esta:

103	67	61
113	89	97
73	83	107

Que tiene una puntuación de 64,06. Pese a todo, está lejos de la máxima puntuación.

La curva de aprendizaje fue la siguiente.





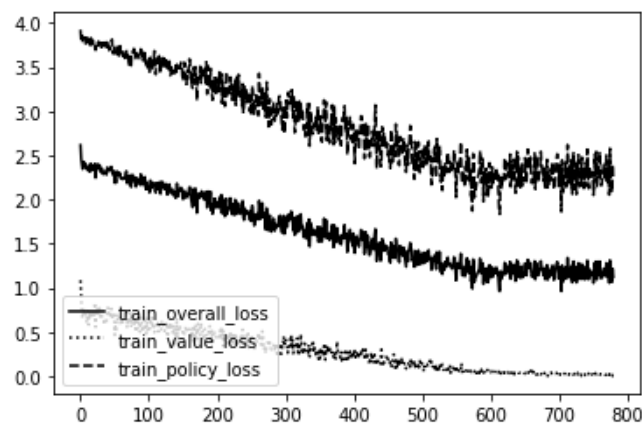
## ENTRENAMIENTO SOLAMENTE CON SUMAS

<https://github.com/bpm246/biobloques-ML-investigation/tree/main/DeepReinforcementLearning>

A la vista de los resultados, en el siguiente entrenamiento decidimos simplificar el objetivo del entrenamiento: ver si el modelo es capaz de sumar y obtener un mismo valor en la suma de las filas, columnas y diagonales de las matrices; independientemente de si los números son primos o no. Para ello, establecimos un nuevo sistema de puntuación mucho más simple que el anterior.

Las nuevas reglas del juego son las siguientes: como tenemos un total de 8 sumas posibles (3 de las filas, 3 de columnas y 2 de diagonales) daremos 1 punto a la matriz si tiene dos sumas iguales, 2 puntos si tiene 3 sumas iguales, 3 puntos si tiene 4 iguales y así consecutivamente hasta un máximo de 7 puntos. Además, solo se puntuará un único valor de los sumatorios. Es decir, si, por ejemplo, una matriz tiene tres sumas iguales con un valor X y dos sumas iguales, pero con un valor Y distinto de X, le daremos sólo los puntos por las tres sumas, por tanto, tendrá 2 puntos en este caso.

Su curva de aprendizaje después de 5 días entrenando fue la siguiente:



El programa no alcanzó los objetivos esperados. Por tanto, la conclusión es que este algoritmo no es eficaz para nuestro problema: si no es capaz de entender la regla de la suma va a ser imposible que sea capaz de construir matrices anidadas de números primos.

## SEGUNDO ALGORITMO

<https://github.com/suragnair/alpha-zero-general>

### PRIMER ENTRENAMIENTO

Después de ver estos resultados, decidimos repetir la prueba (buscar matrices anidadas independientemente de la presencia de primos), pero utilizando otro modelo de alphazero.

Dado que este modelo nos da una libertad mayor para implementar las reglas, introdujimos nuestro juego de una manera un poco distinta:

1. El tablero son dos matrices de 3x3, una para cada jugador (a diferencia del tablero anterior que consistía en un array unidimensional de tamaño el doble de todos los movimientos posibles y dividido en dos mitades, una para cada jugador)
2. Cada jugador juega en una de las matrices
3. Cada casilla se rellena con el número correspondiente (a diferencia del caso anterior en el que el número estaba determinado por la posición en el tablero)
4. No existe limitación en los movimientos, es decir, pueden repetirse números, aunque esto supone una penalización en la puntuación, evidentemente.

Aplicamos solamente la regla de la suma, con una pequeña diferencia respecto del caso anterior: en esta prueba a las distintas sumas les damos puntos también (es decir, si una matriz tiene dos tríos que suman 100 por ejemplo y otros dos que suman 70, la matriz obtiene 2 puntos, uno por cada par)

Además, por temas de memoria y simplicidad, el máximo número que puede jugar este algoritmo es 113.

Los resultados de este entrenamiento están en lacarpeta  
logger\_segundo\_algoritmo\_primer\_entrenamiento

### SEGUNDO ENTRENAMIENTO

En una segunda prueba solo dimos puntuación a un único valor de la suma (siguiendo el ejemplo anterior: si una matriz tiene dos tríos que suman 100 por ejemplo y otros dos que suman 70, la matriz solo obtendrá un punto en total) Esta es la misma regla que aplicamos en la prueba de la semana pasada.

El lunes veremos los resultados.

### TERCER ENTRENAMIENTO

Para intentar simplificar al máximo, hicimos una última versión que solo permitía al programa jugar, exclusivamente, números del 1 al 9. Además, dejamos que obtuviera puntos incluso si existen números repetidos en la matriz.

Los resultados fueron los siguientes:

1. La máxima puntuación que obtuvo fue de 6
2. Estas matrices de 6 puntos están compuestas en su mayoría por 9 y 8. Es decir, que abusa de la repetición.
3. La mayor parte de las matrices que obtiene siguen teniendo una puntuación mínima.

Las mejores matrices que obtuvo el entrenamiento fueron las siguientes:

-----

Matrices de puntuación 6

2021-10-20 07:06:49,558 INFO Matriz 1

2021-10-20 07:06:49,558 INFO ['9', '8', '8']

2021-10-20 07:06:49,559 INFO ['8', '8', '9']

2021-10-20 07:06:49,559 INFO ['8', '9', '8']

-----

Matrices de puntuación 5

2021-10-20 08:06:40,496 INFO Matriz 1

2021-10-20 08:06:40,496 INFO ['9', '8', '8']

2021-10-20 08:06:40,497 INFO ['8', '9', '9']

2021-10-20 08:06:40,497 INFO ['9', '9', '8']

2021-10-20 07:36:40,808 INFO Matriz 1

2021-10-20 07:36:40,808 INFO ['9', '8', '8']

2021-10-20 07:36:40,808 INFO ['8', '9', '9']

2021-10-20 07:36:40,808 INFO ['9', '9', '8']

2021-10-20 07:26:11,205 INFO Matriz 1

2021-10-20 07:26:11,205 INFO ['8', '9', '8']

2021-10-20 07:26:11,205 INFO ['8', '8', '8']

2021-10-20 07:26:11,205 INFO ['8', '7', '8']

2021-10-20 07:18:18,708 INFO Matriz 1

2021-10-20 07:18:18,708 INFO ['9', '9', '8']

2021-10-20 07:18:18,708 INFO ['8', '9', '9']

2021-10-20 07:18:18,708 INFO ['9', '9', '8']

2021-10-20 07:18:16,313 INFO Matriz 1

2021-10-20 07:18:16,313 INFO ['8', '6', '9']

2021-10-20 07:18:16,314 INFO ['6', '8', '9']

2021-10-20 07:18:16,314 INFO ['9', '9', '5']

2021-10-20 07:08:29,044 INFO Matriz 1

2021-10-20 07:08:29,044 INFO ['9', '8', '8']

2021-10-20 07:08:29,044 INFO ['8', '8', '9']

2021-10-20 07:08:29,045 INFO ['9', '9', '8']

2021-10-20 07:05:10,225 INFO ['9', '9', '8']

2021-10-20 07:05:10,225 INFO ['8', '9', '9']

2021-10-20 07:05:10,225 INFO ['9', '9', '8']

2021-10-20 06:35:37,186 INFO Matriz 1

2021-10-20 06:35:37,186 INFO ['9', '9', '8']

2021-10-20 06:35:37,187 INFO ['8', '9', '9']

2021-10-20 06:35:37,187 INFO ['9', '9', '8']

2021-10-20 03:42:24,573 INFO Matriz 2

2021-10-20 03:42:24,573 INFO ['3', '1', '7']

2021-10-20 03:42:24,573 INFO ['6', '7', '3']

2021-10-20 03:42:24,573 INFO ['2', '8', '6']

2021-10-20 03:12:34,734 INFO Matriz 1

2021-10-20 03:12:34,734 INFO ['8', '6', '9']

2021-10-20 03:12:34,734 INFO ['6', '8', '9']

2021-10-20 03:12:34,734 INFO ['9', '9', '5']

2021-10-20 02:40:55,884 INFO Matriz 1

2021-10-20 02:40:55,884 INFO ['9', '8', '8']

2021-10-20 02:40:55,884 INFO ['8', '9', '9']

2021-10-20 02:40:55,884 INFO ['9', '9', '8']

2021-10-20 01:03:57,842 INFO Matriz 2

2021-10-20 01:03:57,842 INFO ['3', '8', '5']

2021-10-20 01:03:57,843 INFO ['4', '5', '7']

2021-10-20 01:03:57,843 INFO ['9', '3', '4']

Los resultados completos de este entrenamiento están en la carpeta  
logger\_segundo\_algoritmo\_tercer\_entrenamiento

#### **CUARTO ENTRENAMIENTO**

Por último, hicimos un programa que impedía las repeticiones para evitar los problemas. Pero el programa no llegó a construir ninguna matriz con una puntuación por encima de 3.