

Razonamiento Automático

Machine Learning

Ivaró Jover Ivarez (aja10@alu.ua.es)
Jordi Amoros Moreno (jam80@alu.ua.es)
Cristian Garca Romero (cgr71@alu.ua.es)
Universidad de Alicante

January 5, 2019

Índice

1	Tecnologías implementadas y estado operativo.	3
1.1	Análisis de RAM	3
1.1.1	Modificando ALE	5
1.2	Bots básicos	8
1.2.1	Plantilla común	8
1.2.2	Breakout	10
1.2.3	Boxing	13
1.2.4	Demon Attack	14
1.3	Perceptrón	24
1.4	Redes neuronales	28
1.4.1	Algoritmo Genético (GANN)	29
2	Manual de utilización	44
2.1	Bots y bots naive	45
2.2	Red Neuronal	46
2.3	Red Neuronal y Genético: <i>GANN</i>	47
3	Experimentos realizados y resultados obtenidos	49
3.1	Experimento con el Perceptrón	49
3.2	Experimentos con la Red Neuronal	52
3.3	Experimentos con la Red Neuronal: <i>Dropout</i>	53
3.4	Experimentos con la Red Neuronal: <i>Cross Validation</i> 10	54
3.5	Experimentos con la Red Neuronal y Genético: <i>GANN</i>	55

4 Conclusiones	61
4.1 Valoracin personal de las prticas realizadas	61
4.2 Indicar qu ha echado de menos el alumno en la formacin recibida en la Universidad que considera le hubiera ayudado .	61
4.3 Posibles sugerencias para mejorar las prticas de empresa . .	61

1 Tecnologías implementadas y estado operativo.

1.1 Anlisis de RAM

Uno de los puntos ms importantes a la hora de implementar un bot con IA para un juego de Atari, es entender cmo est hecho. Utilizaremos el entorno *Arcade Learning Environment* (ALE) para extraer caractersticas de los juegos, el cual cuenta con una API que nos permite extraer informacin de los mismos. Para ello, se ha desarrollado un lector de RAM que nos ayuda a visualizar los 128 bytes de memoria de la Atari mientras se ejecuta un juego.

Adems, dicho lector implementa colores, lo cual permite que se puedan distinguir las posiciones de RAM que cambian de las que no en un step determinado (paso de ejecucin) como se puede ver en la figura 1.

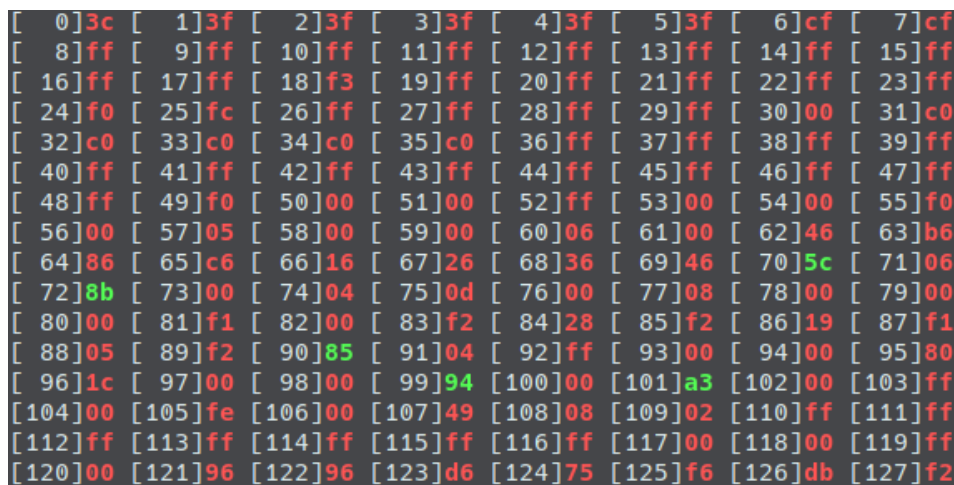


Figure 1: El color verde indica que el valor ha cambiado en este step.

Una de las caractersticas de este lector es que acompaa la ejecucin con un volcado de analytics para ver las posiciones de RAM que ms han cambiado en una ejecucin determinada.

Para extraer los datos mas interesantes de un juego en concreto, simplemente hay que observar las posiciones de RAM mas alteradas segn nuestro analytics. Una vez hecho esto, se pondr el juego en cmara lenta gracias a una feature del entorno ALE, lo cual nos permitir ver con qu sentido cambian estos valores. Como punto a destacar, no todos los valores que cambian mucho sern relevantes a la hora de sacar datos importantes del juego (un contador podra no ser relevante para un caso especfico).

Una vez hecho esto se puede desglosar la RAM de manera bastante precisa, sacando datos como los siguientes.

```
Para las coordenadas X hay que aplicar la formula que obtiene cada nibble por separado y, ademas,
invierte el primer nibble

Las coordenadas X de los enemigos van desde:
  Cuando las 2 moscas estan unidas:
    Parte Izquierda: desde 16 hasta 147
    Parte derecha: desde 24 hasta 155

  Cuando el enemigo muere, la coordenada X vale 0 y va aumentando hasta alcanzar el valor de donde
reaparece
  Cuando el enemigo muere definitivamente en la ronda actual, la coordenada X se queda congelada con
el ultimo valor que tuvo cuando el enemigo estaba vivo

La coordenada X del jugador va desde 21 hasta 138

0: El valor indica la ronda actual que se esta jugando. Se pasa de una ronda a otra cuando se
eliminan todos los enemigos en pantalla. Al terminar la ronda, el jugador se le transporta a la
posicion del medio de la pantalla (donde se empieza a jugar al principio de la partida). El valor
inicial es 0 y va aumentando
1: Modifica el contador de la puntuacion (valor * 10000)
2: * (parece tener el valor 0 todo el rato)
3: Modifica el contador de la puntuacion (valor * 100)
4: * (parece tener el valor 0 todo el rato)
5: Modifica el contador de la puntuacion (valor * 1)
6: * (parece tener el valor 0 todo el rato)
7: Parece seguir algun patron para el enemigo mas lejano. Cuando el enemigo mas lejano deja de
aparecer en la ronda actual, deja de cambiar de valor y se queda en un valor fijo
8: Igual que 7 pero para el enemigo de en medio
9: Igual que 7 pero para el enemigo mas cercano
10: Parece seguir algun patron para el enemigo mas lejano. Cuando el enemigo mas lejano reaparece en
la ronda actual, el valor empieza a cambiar, y cuando ya ha reaparecido, se congela
```

Figure 2: Demon Attack - Analisis de las primeras posiciones de RAM

Como se puede observar en la figura 2, para obtener la informacin correcta no solo basta con extraer las posiciones relevantes, en algunos casos ser necesario procesar esta informacin. Por ejemplo, en Demon Attack, las coordenadas X de las entidades aparecen ofuscadas de la siguiente manera:

```
Valor original coordenada X en RAM:
5A -> Primera conversi3n -> A5 -> Segunda conversi3n -> A2
```

Figure 3: Demon Attack - Coordenadas X de las entidades

Como podemos ver en la figura 3, los nibbles de las coordenadas estn invertidos, adems, el primer nibble requiere una operacin extra, una resta (7 - valor del nibble).

Una vez tenemos la informacin recogida y procesada, la podremos utilizar para crear una IA capaz de jugar al juego en concreto. Adems de eso, el entorno ALE cuenta con diversas funcionalidades que nos permiten recoger la informacin en pantalla en el caso que fuese necesario.

1.1.1 Modificando ALE

La primera aproximación que se hizo a la hora de analizar los datos de los diferentes juegos fue básica y til, pero presentaba varios problemas. El problema más serio que nos encontramos era la cantidad de datos que teníamos ante nuestros ojos, cosa que puede dificultar bastante el encontrar la información importante que necesitamos y, además, que dependemos totalmente de nuestra "intuición" para saber que significa cierta posición de la RAM. Intentando solventar un poco este problema, se nos ocurrió aislar las posiciones que hay en la RAM y, además, poder modificar esos valores aislados para no depender totalmente de nuestra intuición, sino saber más a ciencia cierta, a base de prueba y error, saber qué hace cierta posición de RAM, lo cual nos daría la ventaja de profundizar más en aquellas posiciones de RAM que por alguna razón, en el paso anterior, hemos visto que podían resultar relevantes.

A la hora de intentar poner esta idea en práctica lo primero que intentamos es obtener una referencia de la RAM desde el objeto que hemos ido empleando en todo momento: **ALEInterface**. Este objeto nos dota del método *getRAM()* que hemos ido utilizando en todo momento para el análisis de los datos (y para posteriormente el tratamiento de los mismos) y que está en el archivo *src/ale_interface.cpp*. La implementación de dicho método es la siguiente:

```
// Returns the current RAM content
const ALERAM& ALEInterface::getRAM() {
    return environment->getRAM();
}
```

Figure 4: Método *getRAM()*

Como observamos, el método *getRAM()* nos devuelve una referencia de un objeto de tipo **ALERAM**, por lo que vamos a ver que métodos nos pueden servir para modificar la RAM. Cuando investigamos el archivo *src/environment/ale_ram.hpp* encontramos la definición del objeto **ALERAM** y la implementación de sus métodos. Contiene 2 métodos que pueden servirnos para nuestro propósito (modificar la RAM), los cuales se muestran en la figura 5.

```

/** Returns a reference to a concrete byte. */
byte_t *byte(unsigned int x);

/** Returns the whole array (equivalent to byte(0)). */
byte_t *array() const { return (byte_t*)(m_ram); }

```

Figure 5: Mtodos de acceso a la RAM de ALEInteface

Una vez ya encontramos estos mtodos, intentamos hacer una pequena prueba, la cual falla, ya que al intentar hacer uso del mtodo *array()* vemos que modificamos un valor y que, efectivamente, al volver a comprobar el valor, se ha modificado, pero nada sucede en el juego (la prueba la realizamos sobre el videojuego *breakout*, el cual tiene la coordenada X de la barra en la posicin 72). Lo que estaba sucediendo, despues de indagar un poco por el cdigo de ALE era que el mtodo *getRAM()* de *ale_interface.cpp* era un mtodo que a su vez llamaba a otro mtodo llamado *getRAM()* pero del objeto **StellaEnvironment**, objeto que es el encargado de manejar todo el entorno de ALE. Este objeto, de nuevo, devuelve otra vez un objeto de tipo ALERAM por referencia, por lo que de nuevo parece que no hay problema, pero este objeto, *StellaEnvironment*, contiene una copia de la RAM del objeto **OSystem**, objeto que es el ncleo de la Atari 2600 (del emulador). El problema consiste en que este objeto ALERAM con el que constantemente trabajamos es una copia del que el objeto *OSystem* contiene, es decir, cada vez que llamados al mtodo *getRAM()* desde *ale_interface.hpp*, se hace otra llamada al mtodo *getRAM()* de *src/environment/stella_environment.cpp* y este devuelve su copia interna, no la del ncleo de la consola. Cada vez que el objeto *StellaEnvironment* actualiza su copia hace uso del mtodo *processRAM()*.

```

void StellaEnvironment::processRAM() {
    // Copy RAM over
    for (size_t i = 0; i < m_ram.size(); i++)
        *m_ram.byte(i) = m_osystem->console().system().peek(i + 0x80);
}

```

Figure 6: Mtodo *processRAM()* de StellaEnviroment

Una vez llegado a este punto, lo que se puede hacer si se quiere modificar la RAM que contiene el objeto *OSystem* es implementar un mtodo que

lo que haga es actualizar la RAM del objeto *OSystem* del mismo mtodo que podemos observar que se puede obtener una copia de la RAM con su mtodo *console().system().peek()*, pero tenemos que comprobar si existe el mtodo que nosotros necesitamos, y es que en vez de *peek()* necesitamos un mtodo que nos permita actualizar. De nuevo, buscando en el cdigo, encontramos el archivo *src/emucore/OSystem.hxx*, el cual nos lleva al archivo *src/emucore/Console.hxx*, el cual nos lleva, finalmente, al archivo *src/emucore/System.hxx*. En este ltimo encontramos el mtodo *poke(uInt16, uInt8)*, el cual nos indica a travs de su descripcin que modifica un valor de la RAM de la consola Atari 2600. Parece ser que ya lo tenemos todo, pero nos queda un ltimo problema, y es que este mtodo no es accesible desde el objeto *StellaEnvironment*, ya que su referencia del objeto *OSystem*, a diferencia del objeto *ALEInterface*, est en su parte privada y no nos da ningn mtodo pblico para obtener una copia de la referencia. El ltimo paso que nos queda para solucionar este ltimo problema es modificar el cdigo y aadir un mtodo que nos permita actualizar la RAM de la Atari 2600 a partir de la copia que tiene internamente el objeto *StellaEnvironment*, ya que esta ltima s que podemos modificarla.

```
/**
 * Change the byte at the specified address to the given value.
 * No masking of the address occurs before it's sent to the device
 * mapped at the address.
 *
 * @param address The address where the value should be stored
 * @param value The value to be stored at the address
 */
void poke(uInt16 address, uInt8 value);
```

Figure 7: Mtodo *poke(uInt16, uInt8)*

El mtodo que hemos implementado es muy sencillo. Es igual que el mtodo *processRAM()* pero utilizando el mtodo *poke()* en vez de *peek()*. Una vez ya implementado el mtodo, simplemente implementamos otro mtodo en el objeto *ALEInterface* que simplemente haga una llamada a este mtodo que est en el objeto *StellaEnvironment* y ya habremos actualizado la RAM de la Atari 2600. Por ltimo solo quedara recompilar ALE. Este proceso que hemos realizado a sido con el nico objetivo de comprender mejor el significado de las posiciones de RAM de los diferentes juegos a analizar durante este trabajo. Se adjuntarn para cada juego un archivo denominado **RAM.txt** en el cual hemos analizado cada posicin de RAM de manera aislada.

```

void StellaEnvironment::processBackRAM()
{
    // Copy RAM back
    for (size_t i = 0; i < m_ram.size(); i++)
        m_osystem->console().system().poke(i + 0x80, *m_ram.byte(i));
}

```

Figure 8: Mtodo que actualiza la RAM de la Atari 2600

1.2 Bots bsicos

Se han desarrollado bots semi-deterministas empleando tcnicas bsicas de inteligencia artificial para poder extraer datos de gameplay. Los primeros volcados de datos se hicieron con operarios humanos jugando a los juegos, pero al ver que nuestras scores eran mas bien bajas, se opt por implementar IA bsica para cada uno de los juegos.

Estas implementaciones bsicas mejoraron mucho las scores obtenidas, por lo que los datos extrados de los bots eran mas afines a obtener mayores puntuaciones que los nuestros.

Adems, sobre esta IA bsica, se pueden hacer iteraciones de mejora, teniendo en cuenta ms datos o mas informacin en pantalla, como se ha comentado anteriormente en la subseccin 1.1.

Este scripting bsico ayudar mas adelante a la implementacin utilizando machine learning, ya que los datos extrados y procesados para la implementacin bsica sern utilizados por el algoritmo de machine learning.

A continuacin, describiremos cada uno de los bots bsicos y su funcionamiento al igual que algunos detalles de implementacin.

1.2.1 Plantilla comn

Todos los bots comparten una serie de utils que analizaremos a continuacin.

argv 1 es la ROM que utilizar nuestro ejecutable, argv 2 se corresponde con el contenido multimedia (video y audio), argv 3 escribir la RAM en consola y argv 4 exportar los datos de gameplay a un archivo *valores separados por comas* (CSV) si as se requiere. Se han parametrizado estas opciones porque las ejecuciones son mucho mas lentas conforme mas informacin requiramos, esto se nota sobre todo a la hora de desactivar el contenido multimedia.


```

/**
 * argv[1] : rom
 * argv[2] : media? true/>false<
 * argv[3] : print_ram? true/>false<
 * argv[4] : to csv? true/>false<
 */
const bool display_media(argc >= 3 ? atoi(argv[2])==1 : false);
const bool printRam(argc >= 4 ? atoi(argv[3])==1 : false);
toCSV = argc == 5 ? atoi(argv[4])==1 : false;

```

Figure 9: Opciones de ejecucin

```

alei.setBool("sound", display_media);
alei.setBool("display_screen", display_media);
alei.loadROM(argv[1]);

```

Figure 10: La ROM y el contenido multimedia corren a cargo de ALE.

Algunas de estas opciones corren a cargo del entorno (Figura 10), mientras que las otras han sido implementadas por nosotros.

Otra de las partes comunes a todos los bots es el bucle principal de ejecucin que podemos ver a continuacin en la figura 11. Este bucle est situado en **main()** y es el encargado de analizar y ejecutar las acciones requeridas en cada step del juego en activo.

```

for (step = 0; !alei.game_over() && step < maxSteps; ++step)
{
    // Debug mode *****
    if(printRam) printRAM();
    if(display_media) checkKeys();
    // *****

    // Total reward summation
    totalReward += manualInput ? manualMode() : agentStep();
}

```

Figure 11: Bucle principal de ejecucin.

En este bucle observamos nuestra opcin **printRam** que llama al mtodo encargado de imprimir la RAM, el cual convierte a hexadecimal cada uno de los valores de RAM obtenidos mediante el mtodo **getRAM().get(i)** del entorno ALE, adems hace un seguimiento de los valores de RAM del step anterior para comprobar si dichos valores han cambiado como hemos visto anteriormente en la figura 1.

Otra opción que encontramos en el bucle principal de ejecución es **display_media**, pero en este caso es usado para llamar al método **checkKeys()**, esto se hace porque no tiene sentido trackear el input si no existe contenido de video. Este método de input es propio, ya que el método de input de ALE "pausa" el agente, lo cual no nos interesa para extraer datos. **checkKeys()** simplemente activa o desactiva el modo manual propio con la tecla "E", reflejado en la variable **manualInput**.

La variable **manualInput** decide qué función se llama para calcular **totalReward**. **manualMode()** mapea el teclado a diferentes acciones del juego, mientras que **agentStep()** es el bot autónomo específico a cada juego.

Otra de las partes comunes a todos los agentes es la parte de volcado de datos, para ello se han implementado dos funciones de escritura que imprimen strings o dobles en el archivo CSV anteriormente comentado, como bien podemos ver en la figura .

```
void write(double d)
{
    if (toCSV)
    {
        csv << to_string(d);
    }
}
```

Figure 12: Una de las funciones de escritura en CSV.

1.2.2 Breakout

El Breakout es el juego más simple de todos los que analizaremos en esta parte de la sección, pues el número de inputs que tiene es más bien pequeño. En el Breakout contamos con un total de 5 vidas para pasarnos los 2 niveles de los que dispone. El Breakout fue el primer juego en el que vimos la necesidad de automatizar al jugador, ya que la velocidad de la pelota va incrementando en función a los ladrillos restantes que quedan, lo cual provocaba que perdiésemos siempre en este punto.

Simplemente proporcionando la posición X de la pelota y de la pala al bot y moviendo la pala hacia la pelota, el agente ya jugaba bastante bien. Una mejora que se hizo al algoritmo es hacer un control del tamaño de la pala, al igual que en vez de tener en cuenta solo la posición actual de la pelota, añadir a los datos la posición anterior. Con todas estas mejoras la puntuación se maximizó hasta el punto en el cual el agente fue capaz de pasarse el juego completo. A continuación se muestra en el algoritmo 1 el pseudocódigo de la



Figure 13: Juego Breakout.

IA del Breakout.

Algorithm 1: Breakout agent

```
if lives()  $\neq$  lastLives then
    --lastLives;
    act(FIRE);
end
wide := getRAM.get(108);
playerX := getPlayerX();
ballX := getBallX();
if BallX_LastTick < ballX then
    | ballX + = ((rand()%2) + 2);
end
if BallX_LastTick > ballX then
    | ballX - = ((rand()%2) + 2);
end
ballX_LastTick := getBallX();
if ballX < playerX + wide then
    | reward + = act(LEFT);
else
    if (ballX > playerX + wide)&&( playerX + wide < 188) then
        | reward + = act(RIGHT);
    end
end
```

Si analizamos el pseudocódigo anterior, podemos ver cuatro partes. En la primera parte, constituida por el primer if, vemos como el agente presiona la tecla **FIRE** cuando pierde una vida, esto se debe a que en Breakout cuando pierdes una vida tienes que sacar la pelota pulsando esa tecla.

En la segunda parte recogemos los datos principales, en este caso **wide**, que corresponde al ancho de la pala, además de **playerX** y **ballX**.

La tercera parte calcula la dirección de la bola, esto se saca comprobando la posición anterior de la bola con la actual en el eje X, una vez sabemos la dirección aplicamos una suma con un poco de aleatoriedad (para evitar ejecuciones deterministas) en la dirección recogida. Una vez hecho eso nos guardamos la posición de la pelota para la siguiente iteración.

En la cuarta parte aplicaremos el input en función a la posición del jugador respecto a la pelota, teniendo en cuenta el ancho de la pala recogido anteriormente.

1.2.3 Boxing

Boxing es un juego en el cual controlamos a un boxeador y tenemos que asestar mas golpes que el rival para ganar la ronda. En este juego tuvimos los mismos problemas que con el Breakout, por lo que decidimos implementar otro bot, el cual jugaba mejor que nosotros, lo que se resume en todo ventajas.



Figure 14: Juego Boxing.

En este agente hemos tomado una estrategia agresiva, si el rival lanza un puetazo, lo intentaremos bloquear con nuestros propios puos poniendo a nuestro boxeador exactamente en la misma posicin "Y" que el boxeador contrario (una de las features de Boxing es que la colisin de los puos bloquea los golpes). Adems, nuestro boxeador nunca se mover hacia la izquierda, solo se mueve hacia la derecha en posicin de ataque intentando posicionarse en la "Y" del rival, como hemos comentado antes. A su vez, basndonos en ciertas tolerancias, se atacar siempre que sea posible. Esta estrategia agresiva funciona una gran mayora de las veces, adems para evitar el determinismo se ha implementado una pequena aleatoriedad cada vez que recogemos las posiciones del jugador 1 y del jugador 2 de RAM, como bien se puede ver en el ejemplo de la figura 15.

A continuacin se muestra en el algoritmo 2, la IA implementada. Un detalle a comentar antes de entrar a analizar el algoritmo es que **player_pos**

```
int getP2_X()
{
    return alei.getRAM().get(33) + ((rand() % 2) - 1);
}
```

Figure 15: Uso de **rand()** para evitar el determinismo.

es un tipo de dato struct.

Algorithm 2: Boxing agent

```
player_pos p1(getP1_X(), getP1_Y());
player_pos p2(getP2_X(), getP2_Y());
absp1p2X := abs(p1.x - p2.x);
absp1p2Y := abs(p1.y - p2.y);
if absp1p2Y > 3 and absp1p2Y < 20 then
    | reward += act(FIRE);
else
    | if absp1p2X > 25 and absp1p2X < 40 then
        | reward += act(RIGHT);
    | else
        | reward += (p1.y > p2.y) ? act(UP) : act(DOWN);
    | end
end
```

En la primera parte de recogida de datos, encapsulamos en un struct la posicin del jugador uno y la del jugador dos, extrayndolos de RAM. Como ya hemos comentado anteriormente, estas posiciones implementan una aleatoriedad mnima para evitar ejecuciones deterministas. Adems, en la misma seccin, calcularemos la distancia entre ambos jugadores en "X" y en "Y", representadas en **absp1p2X** y **absp1p2Y** respectivamente, para luego utilizarlas mas adelante.

Una vez tenemos todos los datos procesados, observaremos si nos encontramos en un rango de tolerancia "Y" vlido para atacar, si lo estamos, atacaremos (como punto a destacar, este rango es bastante amplio para enfatizar esta estrategia ofensiva). Si no podemos atacar, nos moveremos hacia la derecha con ciertas tolerancias, o nos situamos en la "Y" del enemigo.

1.2.4 Demon Attack

El Demon Attack es un juego en el cual controlamos a una nave con un patrón de movimiento similar al Space Invaders. Tendremos que disparar a las naves rivales para pasar de fase adems de evitar todo contacto enemigo.

El jugador cuenta con una serie de vidas para pasarse los niveles, si estas vidas se acaban el juego termina. De nuevo, vimos la necesidad de crear un bot debido a las bajas puntuaciones obtenidas.



Figure 16: Juego Demon Attack.

El Demon Attack es, sin duda, uno de los juegos mas complejos con los que hemos tenido que lidiar, debido a la cantidad de entidades en pantalla, desde disparos hasta mltiples enemigos. La estrategia de IA seguida en el Demon Attack es compleja debido al factor de la gran cantidad de entidades, pues no solo leemos valores en RAM, sino que aprovechamos la caracterstica de ALE que nos permite leer los pxeles de la pantalla.

Al analizar la RAM, observamos que haban demasiadas caractersticas a tener en cuenta, a pesar de tener un anlisis detallado de la misma. Hay posiciones de RAM que representan varias cosas en diferentes situaciones, un ejemplo es el valor representado por $RAM[20]$:

"20: Coordenada "X" de las balas del enemigo. Tambin es la coordenada "X" de la mosca que se acerca para intentar matar al jugador 1 (se utiliza como una bala). Va desde 29 hasta 147 (incluso cuando dispara la mosca). El valor es todo el rato el mismo hasta que el enemigo dispara y actualiza la coordenada. Si hay mas de 2 cambios consecutivos en 2 frames significa que no es una bala y que es la mosca que sigue al jugador".

Es por ello por lo que se dedujo que para implementar un bot bsico (sin machine learning), era ms fcil hacer un anlisis de lo que estuviera pasando en la pantalla en un instante determinado, sin dejar de lado la RAM. Para esto, se dise un mtodo de visin en el cual la nave es capaz de ver y filtrar enemigos en un rea de visin y actuar concorde la situacin dependiendo de unas reglas determinadas. En la figura 17 se puede observar el rea de visin de la nave.



Figure 17: Vision de la nave.

La visin de la nave se divide principalmente en 3 partes. En la figura vemos una linea roja representada con el nmero 1, esta linea roja representa la visin de la nave en funcin a la anchura de la misma. El nmero 2 es un

rea parametrizable que permite ampliar esta visin de forma simtrica por los dos lados de la nave. Tanto el rectngulo rojo como el azul se extienden por casi toda la "Y" de la pantalla y lo nico que filtran son los disparos enemigos, como los disparos enemigos tienen el mismo color es muy simple filtrarlos. Por lo tanto 1 y 2 se encargan de detectar disparos enemigos, una vez detectados se actuar en consecuencia.

El nmero 3 es un rea reducida dentro del rea completa que se encarga de detectar todas las posibles amenazas, no solo disparos, que estn cerca de la nave. Es decir, el objetivo de la visin es la deteccin de entidades peligrosas en un rea de riesgo para posteriormente evitarlas.

El seguimiento de enemigos para atacarles y matarlos corre a cargo de la RAM, aunque no es del todo exacto debido al problema comentado anteriormente de posiciones de RAM no especificas a una nica funcionalidad.

Para analizar la parte prctica de este algoritmo es necesario separarlo en distintas partes, deteccin, esquivar y seguimiento.

La parte de deteccin funciona recogiendo la pantalla entera en grayscale, esto lo hacemos gracias a una funcionalidad de ALE, una vez tenemos la pantalla guardada, podemos ver cual es el valor de cada pxel. Gracias a esto, podemos determinar el color de los disparos y de las naves enemigas, los que guardaremos en las siguientes variables para despus su posterior filtrado:

```
// non modificable
const int LINE_WIDTH(160);           // Amount of pixels in a line
const int SHIP_WIDTH(7);             // Ship width in pixels
const int LEFT_THRESHOLD(21);        // MIN X non-pixel coordinate the ship can move 25
const int RIGHT_THRESHOLD(138);      // MAX X non-pixel coordinate the ship can move 135
const int P1_BULLETS_COLOR(174);
const int P1_SHIP_COLOR(115);
const int EN_BULLETS_COLOR(176);
```

Figure 18: Constantes fijas relativas a la visin de la nave.

En la figura 18 podemos observar diversas constantes, entre ellas el color de las balas del jugador 1, **P1_BULLETS_COLOR**, el color de la nave del jugador 1, **P1_SHIP_COLOR** y el color de las balas enemigas, **EN_BULLETS_COLOR**. Gracias a estar en grayscale, con tener un solo valor es suficiente. Adems podemos observar otras constantes como **LINE_WIDTH**, que representa el ancho en pxeles de la resolucin de la Atari, **SHIP_WIDTH**, el ancho en pxeles de la nave y **LEFT / RIGHT_THRESHOLD**, que es la coordenada "X" mnima y mxima en la que se puede mover la nave aliada.

Adems de estas constantes, Demon Attack cuenta con cuatro constantes adicionales que nos permiten parametrizar algunas caractersticas de nuestro algoritmo, como se puede ver en la figura 19 .

```
// modifiable params
const int VISION_THRESHOLD(10);
const int SECOND_VISION_LINE_THRESHOLD(168);
const bool bRandomisePlayerGPos(false);
const bool bRandomiseEnemyGPos(false);
```

Figure 19: Constantes modificables (no runtime).

La primera variable de la figura representa el rea extra de visin 2 definida en la imagen 17, la segunda variable define en que linea empieza el rea de visin 3. Las dos variables restantes es aleatoriedad opcional a la hora de recoger las posiciones, como se ha hecho antes en el *Boxing*. En este caso, debido a la gran cantidad de entidades en pantalla se ha optado por deshabilitar la aleatoriedad y hacer las batallas mas deterministas.

Otro dato a remarcar antes de empezar con el algoritmo es el Enum utilizado para las colisiones:

Algorithm 3: Enum empleado para las colisiones.

```
enum BlockingHit { EMoveRight, EMoveLeft, ENotBlocking };
```

Para entender el Enum del algoritmo 3 tenemos primero que comprender las respuestas que puede dar una colisin y como se operan.

- **EMoveRight:** La amenaza se ha detectado en la parte izquierda de nuestra rea de visin, para contrarrestarla nos moveremos a la derecha si podemos.
- **EMoveLeft:** La amenaza se ha detectado en la parte derecha de nuestra rea de visin, para contrarrestarla nos moveremos a la izquierda si podemos.
- **ENotBlocking:** No se ha detectado amenaza ninguna.

Una vez que ya sabemos qu indicadores utilizaremos para las colisiones, podemos proceder a despiezar el algoritmo.

```

DirtyState ds(false, ENotBlocking);
float agentStep()
{
    // get screen information
    alei.getScreenGrayscale(grayscale);

    // Reseting variables
    float reward = 0;
    BlockingHit eBH = ENotBlocking;

    // Iterating from bottom line to top line which defines a vision rectangle (see is BlockingHit)
    for(int line = 185; line > 60; --line) {
        eBH = isBlockingHit(line);
        if (eBH != ENotBlocking && !ds.Dirty)
        {
            /**
             * Dirtying the warning state, now we have to avoid every enemy and undesired object
             */
            ds.Dirty = true; ds.Direction = eBH;
            break;
        } else if (eBH != ENotBlocking) {
            /**
             * Theoretically at this point the direction is already set so we don't have to re-set it here
             * we just mark the dirty state to true
             */
            ds.Dirty = true;
            break;
        }
    }
}

```

Figure 20: Primera parte del algoritmo de IA para el Demon Attack.

En esta primera parte recogemos los pxeles de la pantalla con la funcionalidad de ALE `getScreenGrayscale`, para luego emplearlos dentro del bucle en la funcin `isBlockingHit(int)`, que analizaremos mas adelante. Este bucle se encarga de recorrer 125 filas de la pantalla. En funcin a la respuesta de `isBlockingHit(int)`, marcar una variable a true, la cual representa que ha habido una colisin, adems dentro de este struct **DirtyState**, disponemos de otra variable de tipo `EBlockingHit` para marcar tambien la direccin contraria a la colisin dada por `isBlockingHit`. Se ha decidido este thresold ya que el resto de lneas de pantalla contienen informacin no relevante para el problema.

Para resumir esta parte del cdigo, marcamos el estado a Dirty y recogemos la direccin conveniente, siempre que `isBlockingHit` haya detectado una colisin. Hasta que no salimos del estado Dirty, no podremos recoger nuevas direcciones, pero de ello se encargar otra parte del cdigo. Es decir, tericamente se ha implementado una mini mquina de estados finita (FSM), ya que la nave posee dos estados principales (dirty y no dirty) y la forma de transicionar entre ellos es mediante el cdigo de colisiones.

```

BlockingHit isBlockingHit(int line) {
    int const FILTER_LINE(LINE_WIDTH*line);
    int const VISION_X_AREA(SHIP_WIDTH + (VISION_THRESHOLD*2));
    for(int i = 0; i < VISION_X_AREA; ++i) {
        const int impact_pixel_val(grayscale[FILTER_LINE + (getP1_X(true)-VISION_THRESHOLD) + i]);
        // Avoid flies
        if (line > SECOND_VISION_LINE_THRESHOLD && ImpactValIsAnEnemy(impact_pixel_val)) {
            if(i < (VISION_X_AREA/2)) return EMoveRight;
            else if(i >= (VISION_X_AREA/2)) return EMoveLeft;
        }
        // Avoid enemy shooting
        if(impact_pixel_val == EN_BULLETS_COLOR) {
            // Blocking hit, now we have to determine the direction we want to based on the impact point
            if(i < (VISION_X_AREA/2)) return EMoveRight;
            else if(i >= (VISION_X_AREA/2)) return EMoveLeft;
        }
    }
    return ENotBlocking;
}

```

Figure 21: Funcin isBlockingHit.

La funcin isBlockingHit tiene que ser capaz de recorrer el array obtenido por **getScreenGrayscale** de manera eficiente. Para ello lo primero que hace es resolver la posicin del array que ocupa el primer pxel de la lnea a analizar, esto es fcil, pues simplemente multiplicando el ancho de la lnea con el nmero de lnea ya tenemos este nmero (asumiendo que las lneas empiezan en 0).

La siguiente tarea es resolver el threshold, el cual consiste en el rea que ocupa la nave, en este caso 7, mas el threshold determinado, que en nuestro caso corresponde al nmero de pxeles extras que tendremos en cuenta en cada lado, por ejemplo, si el threshold es 2, el rea total ser $7+2+2 = 11$.

La iteracin que haremos en el bucle ir desde 0 hasta el rea total que ocupa la visin de la nave, esto es, como ya hemos dicho anteriormente, el rea de la nave mas el threshold. Esto hace que solo iteremos los pxeles necesarios para nuestra situacin.

La primera lnea que encontramos en el bucle puede parecer confusa, pero tiene su explicacin:

Algorithm 4: Calculo del valor del pxel en un punto determinado de nuestra area de visin.

```

const int impact_pixel_val(grayscale[FILTER_LINE +
    (getP1_X(true)-VISION_THRESHOLD) + i]);

```

Necesitamos el valor del pxel en un punto determinado de nuestra rea de visin. Para ello, necesitamos pedirle al array grayscale esta informacin, pero antes necesitamos calcular qu posicin del array es la correcta. Al haber precalculado la posicin del array del primer pxel de la lnea lo tenemos mucho mas fcil, pues tenemos por donde empezar. A ese nmero le tenemos que

sumar la posicin del jugador uno, representado en `getP1_X(true)`, al cual le pasamos el parametro `true` para que nos devuelva la posicin en pxeles, la cual es ms precisa para este problema. Una vez tenemos la posicin del jugador solo nos queda sumarle `i` para abordar todo el rea de visin. Pero como bien podemos observar en el algoritmo 4, estamos restando a la posicin del jugador el `VISION_THRESHOLD`, esto se hace porque sino estaramos solo teniendo en cuenta el rea extra por la derecha, ya que la colisin estara desplazada a la izquierda como muestra la figura 22, por ello es necesario restar `VISION_THRESHOLD`.

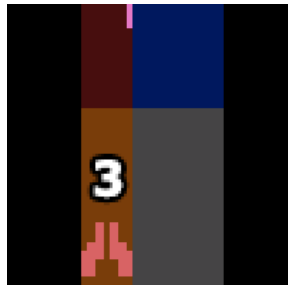


Figure 22: Area de sin restar `VISION_THRESHOLD` al jugador.

Una vez hemos resolucionado el valor del pxel, podemos hacer el filtrado del mismo, el cual se realizar en el cuerpo del bucle.

```
// Avoid flies
if (line > SECOND_VISION_LINE_THRESHOLD && ImpactValIsAnEnemy(impact_pixel_val)) {
    if(i < (VISION_X_AREA/2)) return EMoveRight;
    else if(i >= (VISION_X_AREA/2)) return EMoveLeft;
}
// Avoid enemy shooting
if(impact_pixel_val == EN_BULLETS_COLOR) {
    // Blocking hit, now we have to determine the direction we want to based on the impact
    if(i < (VISION_X_AREA/2)) return EMoveRight;
    else if(i >= (VISION_X_AREA/2)) return EMoveLeft;
}
```

Figure 23: Resolucin de la colisin

Como bien observamos en la figura 23, haremos una primera distincin para las moscas y enemigos mas cercanos a nuestra segunda rea de visin seguido de la visin normal. Priorizamos el segundo rea de visin porque est mas cerca del jugador, lo cual se traduce en ms peligro a perder una vida. En el primer if, miramos si nos encontramos en el rea de riesgo y filtramos por color principalmente si lo que hay en ese pxel es un enemigo. Si es as, es que ha habido una colisin. Si la `i` en la cual nos encontramos es menor (o est a la

izquierda) de la mitad del rea de visin, nos moveremos a la derecha (camino ms corto para evitar la colisin). Haremos lo mismo para el otro ladro. El segundo if es exactamente lo mismo solo que nos centramos nicamente en las balas enemigas.

Finalmente, si no se ha encontrado colisin ninguna, como hemos visto en la figura 21, devolveremos *ENotBlocking*.

```

/*****
 * Blocking hit detected at this point
 *****/
if(eBH != ENotBlocking)
{
    // Left thresold
    if(getP1_X() == LEFT_THRESOLD) {
        // If we cannot move more to the left we change direction
        ds.Direction = EMoveRight;
    }
    // Right thresold
    else if(getP1_X() == RIGHT_THRESOLD) {
        // If we cannot move more to the right we change direction
        ds.Direction = EMoveLeft;
    }
    switch(ds.Direction){
        case EMoveLeft:
            reward+= alei.act(PAYER_A_LEFT);
            break;
        case EMoveRight:
            reward+= alei.act(PAYER_A_RIGHT);
            break;
        default:
            reward+= reward+=alei.act(PAYER_A_FIRE);
    }
} else {
    // Here we are safe, don't move but keep shooting
    reward+=alei.act(PAYER_A_FIRE);
    ds.Dirty = false;
    const int mypos = getP1_X();
    const int en_pos = EnemyHandler();
    if(mypos < en_pos) {
        reward+= alei.act(PAYER_A_RIGHT);
    } else if (mypos > en_pos) {
        reward+= alei.act(PAYER_A_LEFT);
    } else {
        reward+=alei.act(PAYER_A_FIRE);
    }
}
return (reward + alei.act(PAYER_A_NOOP));

```

Figure 24: Aplicando movimiento a la nave

La parte que extiende a la figura 20, es la figura 24. En sta figura

podemos observar la parte del algoritmo que finalmente aplica movimiento a la nave. En este punto ya hemos solucionado la colisin y solo nos queda movernos en la direccin conveniente.

Esta parte del cdigo se divide en dos dependiendo respectivamente si hemos encontrado una colisin o no. En la primera parte ya tenemos en el **DirtyState** apuntado la direccin en la cual movernos, sin embargo si nos encontramos en los bordes de la pantalla, no nos podremos mover la la direccin del borde, es por esto por lo que cambiamos de direccin cuando llegamos a un borde. Si no estamos en ningn borde, simplemente aplicamos la direccin anotada por **ds.Direction**.

```
struct DirtyState {
    bool Dirty;
    BlockingHit Direction;
    DirtyState(bool bDirty, BlockingHit bDirection)
        : Dirty(bDirty), Direction(bDirection) { }
};
```

Figure 25: Struct Dirty State

En la segunda parte, podemos considerar que el jugador uno est a salvo de posibles ataques enemigos, esto viene determinado, como ya hemos comentado anteriormente, por el resultado de **isBlockingHit**. Por lo tanto, lo que haremos en este modo es disparar siempre que podamos a la vez que intentar traquear a los enemigos en RAM para dispararles dentro de su colisin. **EnemyHandler** se encargará de devolvernos el enemigo mas relevante en un instante determinado (usualmente el mas cercano al jugador). Esta funcin no funciona el 100% de los casos, ya que como hemos comentado anteriormente, la RAM no es especifica a casos concretos sino que va cambiando, por lo tanto habr, veces que el jugador 1 se quede disparando a la nada porque no tiene nada que trackear. Esto se resuelve parcialmente gracias a la aleatoriedad y al movimiento de las entidades enemigas, que acabaremos matando si se acercan a nuestro lser. La solucin completa y real sera llevar un seguimiento mas especifico de todas las unidades en RAM, lo cual puede llevarnos a implementar demasiados casos especificos para un escenario mas simple de solucionar mediante un algoritmo genrico (mediante una buena funcin de fitness) o un neuroevolutivo en su defecto.

1.3 Perceptrn

El perceptrn es una de las tcnicas de *Machine Learning* ms sencillas de comprender, implementar y tiles. En su sencillez radica los buenos resultados que suele dar, siempre que el problema a resolver no sea muy complejo, ya que pasado cierto nivel de complejidad, el perceptrn seguramente no sea la mejor opcin. El perceptrn lo que intenta hacer es separar los datos que le demos, asegurndonos que obtendr la mejor solucin posible si los datos de los cuales partimos son linealmente separables, y en el caso de no serlo, deberemos dar la mejor solucin posible (aquella con menor error).

El perceptrn es una tcnica de aprendizaje supervisado en la que le damos datos etiquetados y, a partir del error obtenido al intentar etiquetar los datos, se hacen ajustes en la configuracin para acercarnos ms a los resultados correctos. Esta tcnica puede utilizarse para clasificacin o para regresin. El resultado que al final obtendremos ser, dependiendo del nmero de entradas que tengamos, una lnea, un plano o un hiperplano. Los elementos del perceptrn son las entradas, los pesos y la salida.

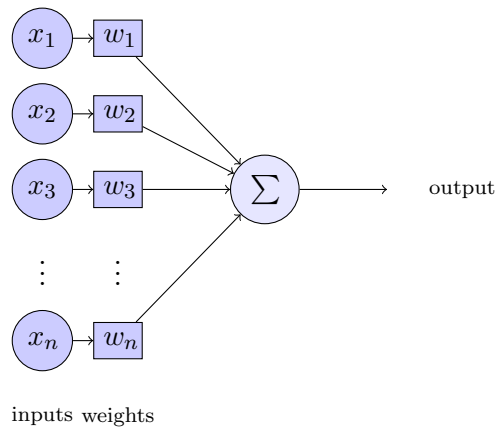


Figure 26: Estructura bsica del perceptrn

La salida que tiene el perceptrn, en el caso del clasificador, es siempre de +1 o -1. Para obtener esta salida, la frmula que se aplica es la siguiente cuando tenemos N entradas:

$$y = \text{signo}\left(\sum_{i=1}^N x_i * w_i\right)$$

Ahora bien, como hemos dicho, lo que el perceptron hace es separar

datos, es decir, encontrar la línea, plano o hiperplano que los separa, pero si nos fijamos en la fórmula que hemos descrito encontramos un problema, y es que no hay término independiente, todos los pesos están asociados con alguna entrada, por lo que la línea, plano o hiperplano siempre pasar por el origen de coordenadas, cosa que seguramente no sea así para la mayoría de los problemas. La solución es, como hemos dicho, añadir un término independiente, al cual se le suele llamar *bias*. Una manera de hacerlo es añadir un peso w_0 y una entrada x_0 artificial que siempre tenga el valor 1, con lo que al final nos queda la siguiente fórmula:

$$x_0 = 1$$

$$y = \text{signo}\left(\sum_{i=0}^N x_i * w_i\right)$$

Finalmente, la figura 27 muestra la estructura final del perceptrón.

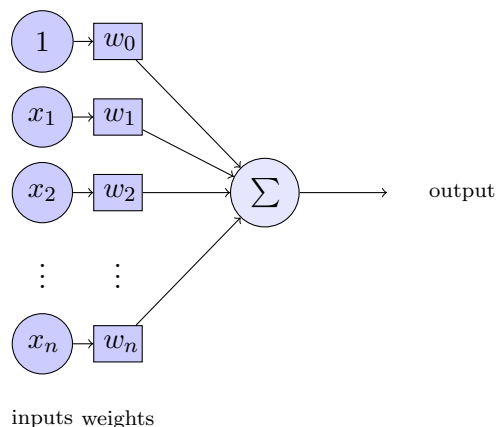


Figure 27: Estructura del perceptrón

En nuestro trabajo hemos implementado el algoritmo de aprendizaje del perceptrón (PLA), el cual nos permite ir ajustando la configuración del perceptrón, lo que quiere decir que vamos ajustando los pesos con el fin de que obtengamos el resultado que buscamos para las entradas. El algoritmo de aprendizaje del perceptrón se muestra en el algoritmo 5.

Algorithm 5: PLA

```
predictions = getPredictions(inputs);
mPoint = getRandomMisclassifiedPoint(predictions, targets);
if mPoint.empty() then
|   return;
else
|   for  $i \leftarrow 0$  to  $N$  do
|   |    $\text{weights}[i] += mPoint.target * mPoint.prediction$ 
|   end
end
```

El código del PLA deberá ejecutarse tantas veces como fuera necesario, y eventualmente llegará a una solución, siempre que los datos fueran linealmente separables. Ya que no siempre van a ser linealmente separables, podemos poner un bucle y hacer tantas iteraciones como fuera necesario (a cada iteración de este bucle se le conoce con el concepto de **poca**).

Algorithm 6: PLA con pocas

```
for  $epoch \leftarrow 1$  to  $epochs$  do
|   predictions = getPredictions(inputs);
|   mPoint = getRandomMisclassifiedPoint(predictions, targets);
|   if mPoint.empty() then
|   |   return;
|   else
|   |   for  $i \leftarrow 0$  to  $N$  do
|   |   |    $\text{weights}[i] += mPoint.target * mPoint.prediction$ 
|   |   end
|   end
end
```

Por último, vemos el problema de que si no tenemos datos linealmente separables, las pocas no solucionan el problema totalmente (solucionan el problema de no quedarnos en un bucle infinito), ya que puede que antes de llegar a la última poca, ya hayamos encontrado una solución bastante buena, pero que la hayamos perdido por el camino al actualizar los pesos, por lo que a cada poca que se está ejecutando el PLA, deberemos de estar guardando la mejor solución hasta el momento. A esto se le conoce con el nombre del algoritmo *Pocket*.

Algorithm 7: PLA con pocas y Pocket

```
pocket = []
for epoch  $\leftarrow$  1 to epochs do
    predictions = getPredictions(inputs);
    mPoint = getRandomMisclassifiedPoint(predictions, targets);
    if numberOfMisclassifiedPoints(pocket, targets)  $\neq$ 
        numberOfMisclassifiedPoints(weights, targets) then
        | pocket = weights
    end
    if mPoint.empty() then
        | return;
    else
        for i  $\leftarrow$  0 to N do
            | weights[i] += mPoint.target*mPoint.prediction
        end
    end
end
if !pocket.empty() then
    | weights = pocket
end
```

Nuestro código es muy similar al descrito, ya que la parte más importante es la que hemos explicado, la del PLA. La estructura que hemos empleado es la de una clase llamada **Perceptron** que contiene los métodos necesarios para obtener una predicción, la cual se obtiene haciendo uso del método *double getPrediction(const vector<double> &inputs)* y el PLA se ejecuta haciendo uso del método *void trainPerceptron(unsigned epochs, const vector<vector<double>> &inputs, const vector<double> &targets)*.

1.4 Redes neuronales

1.4.1 Algoritmo Genético (GANN)

Los algoritmos genéticos pertenecen a la categoría de los algoritmos evolutivos, los cuales intentan optimizar algún problema basándose en la biología, basándose en cómo la naturaleza ha evolucionado a lo largo de miles de años e intentando utilizar la misma técnica. Concretamente, los algoritmos genéticos intentan imitar la estructura genética de los individuos y cómo esta estructura va evolucionando de una generación de individuos a la siguiente, esa es la idea principal de dichos algoritmos. Antes de pasar a explicar la combinación de los algoritmos genéticos con las redes neuronales vamos a explicar las ideas principales de los algoritmos genéticos.

Como hemos dicho, los algoritmos genéticos se basan en la evolución de la estructura genética de los individuos y cómo se transmite esta herencia genética de una generación a la siguiente. Hablando sobre esta estructura, todo **individuo** vivo tiene una serie de **clulas** y cada una de estas clulas contiene el mismo conjunto de **cadenas de ADN**, lo que se conoce como **cromosomas**. El ADN que contiene un cromosoma es una doble cadena en forma de hebra, y las hebras están conectadas entre sí en una **doble espiral**. Finalmente, esta doble espiral está formada por pequeños bloques básicos, llamados **genes**, que a su vez están formados por sustancias llamadas **nucleótidos**. Solo hay 4 tipos de nucleótidos: timina, adenina, guanina y citosina. En la figura 28 se puede apreciar lo que hemos comentado de una manera más gráfica. Esta es la estructura básica de la información genética de los seres vivos.

Ahora sabemos un poco más cuál es la estructura básica, pero aún no hemos explicado qué es lo que contienen estas partes, lo cual va a ser lo importante para poder comprender correctamente el funcionamiento biológico y así poder hacer una buena aproximación con los algoritmos genéticos a este enfoque de la evolución respecto a la herencia genética. Lo importante es comprender que la información que hay dentro de una clula (conjunto de cadenas de ADN, los cromosomas) es lo que contiene toda la información que define a un individuo y, por lo tanto, tiene toda la información que podrá pasarse al individuo de la siguiente generación (normalmente de forma parcial, no total). Esta colección de cromosomas (información genética dentro de la clula) es lo que se conoce como **genoma**, que no es más que la colección de genes que definen las características del individuo (e.g. ojos marrones, pelo liso, color del pelo, etc.). Volviendo a los **genes**, la configuración que cada gen puede tener es lo que da rasgo a alguna característica concreta, y a esto se le llama **alelo** (característica concreta del individuo), y la posición física donde está el gen a lo largo de la cadena del cromosoma se le llama **locus**.

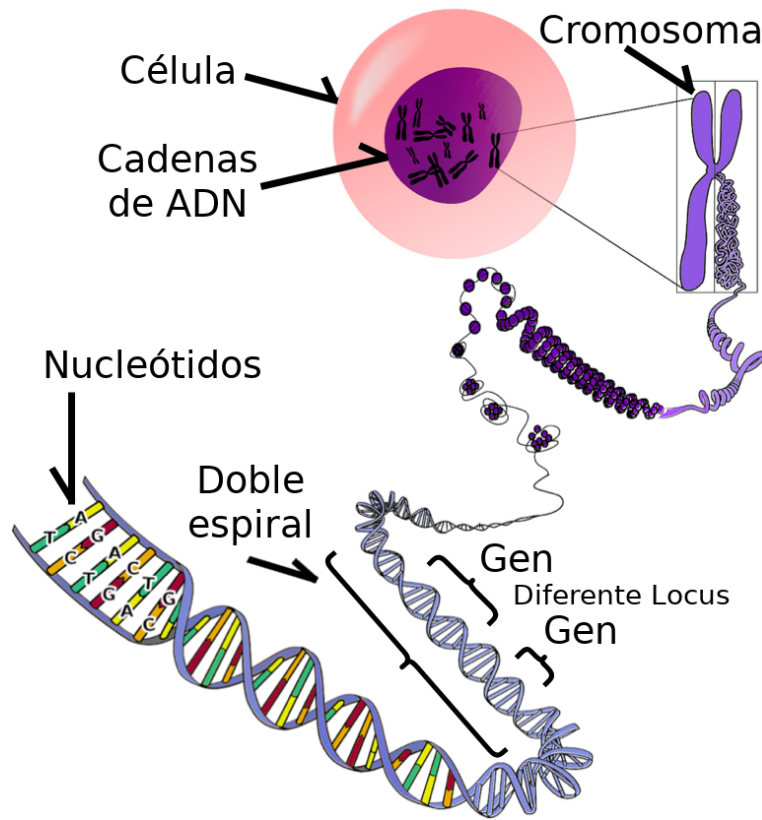


Figure 28: Estructura bsica del ADN

Ahora bien, el estado de los alelos en un genoma particular se conoce como **genotipo** (informacin gentica que da lugar a las caractersticas del individuo). El genotipo es una informacin que una vez est definida no cambia. Al organismo en s, al individuo, a la expresin de ese genotipo en el mundo real, se le conoce como **fenotipo**. Por poner un ejemplo, el ADN que tiene cada ser humano sera el genotipo propio de cada uno, mientras que la persona en s, el como se expresa en el mundo fsico, sera su fenotipo, el cual, adems, est influido por el ambiente. En la figura 29 se ve esta diferencia claramente, que es lo que se conoce como genotipo vs fenotipo.

Observando la figura 29, podemos darnos cuenta de algo, y es que aunque tengamos 2 copias exactas a nivel gentico de un individuo (i.e. clones), lo que significara que tienen el mismo genotipo, podran no tener el mismo fenotipo, ya que el ambiente importa, aunque respecto a los algoritmos genticos esto se obvia normalmente, al no ser que de alguna forma el individuo

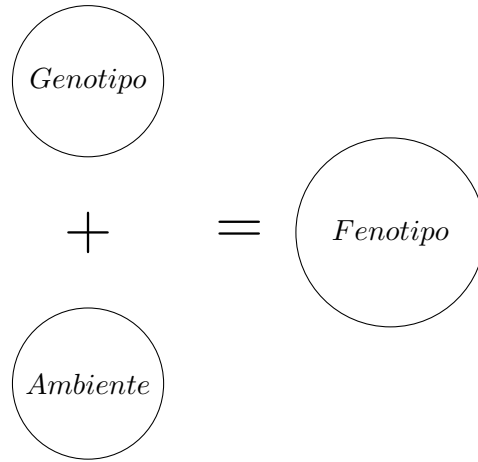


Figure 29: Genotipo vs Fenotipo

evolucione mientras est en vida (nos referimos a algoritmos, no a seres vivos). Si tuviramos un algoritmo gentico que combinado con otra tcnica, el individuo evolucionase a la vez que se ejecuta entonces podramos hablar de que el entorno importa (e.g. algoritmo gentico que resuelve un camino en un laberinto y aprende a la vez que lo recorre, pero adems el laberinto vara, por lo que un individuo podra aprender muy bien a resolverlo, pero otro individuo con el mismo genotipo tiene un laberinto complicado, por lo que no aprende bien, as que el fenotipo vara).

Una vez hablado de todo lo anterior, lo que es la parte biolgica, necesario para comprender correctamente el funcionamiento de un algoritmo gentico, podemos pasar a explicar como funciona el algoritmo en s. Los algoritmos genticos lo que intentan es aproximar el enfoque de la evolucin desde el punto de vista de la herencia gentica: de algn modo codifican el genotipo del individuo, de algn modo se evala lo bien que hace su tarea para sobrevivir, para tener ms posibilidades de pasar parte de su informacin gentica a la siguiente generacin, y se hace evaluando su fenotipo. En definitiva, los algoritmos genticos se basan en los principios de Charles Darwin de su teora de la seleccin natural. Resumidamente es la supervivencia del ms apto.

Las fases principales de los algoritmos genticos son:

1. Crear una **poblacin inicial** de individuos. Esta poblacin suele generarse de forma aleatoria y, aunque muchos individuos harn mal su tarea, otros lo harn medio bien o incluso bien. Tambin se pueden inicializar siguiendo una distribucin de probabilidad que se sabe que sigue el

patrón de los individuos que hacen bien su tarea.

2. Calcular su **fitness**. Este valor se calcula de la función de *fitness*, la cual nos indica cómo de bien un individuo hace la tarea para la que se supone que ha sido concebido. El objetivo suele ser maximizar esta función.
3. **Selección**. Seleccionar los individuos más aptos para pasar su información a la siguiente generación. Esto es una cuestión de probabilidad, no de directamente hacer que los mejores individuos sean los que pasen su información a la siguiente generación. Los mejores individuos tendrán más posibilidades, pero los peores también tendrán una pequeña posibilidad de pasar su información a la siguiente generación.
4. **Crossover (emparejamiento)**. Unir la información genética de los individuos seleccionados para pasar a la siguiente generación.
5. **Mutación**. Las mutaciones son algo que sucede cuando hablamos de la genética. A veces son buenas y a veces son malas, pero está claro que son necesarias. La mutación puede aportarnos algo que puede que ningún individuo de la población inicial tuviera en su información genética, y esa mutación podrá hacer que un nuevo gen surjera y predominara por encima de todos.

La estructura principal de los algoritmos genéticos se muestra en el algoritmo 8.

Algorithm 8: Estructura general de los algoritmos genéticos

```
population = createNewPopulation();
while true do
  fitness = evaluatePopulation(population);
  newPopulation = [];
  for individual ← 1 to population do
    parents = doSelection(population, fitness);
    newIndividual = doCrossover(parents);
    newIndividual = doMutation(newIndividual);
    newPopulation[individual] = newIndividual;
  end
  population = newPopulation;
end
```

Conforme se vaya pasando de generación en generación, los nuevos individuos irán mejorando, ya que han ido obteniendo los mejores genes de sus

antepasados o al menos esa es la idea principal de estos algoritmos. Dentro de estos algoritmos luego hay muchos aspectos que controlar para que estos funcione como es debido, aunque lo bsico ya est expuesto. Ahora pasaremos a hablar por qu pueden ser tiles a la hora de combinarlos con las redes neuronales.

Las redes neuronales forman parte de los algoritmos de aprendizaje supervisado, lo que quiere decir que necesita de datos que previamente estn etiquetados. Con estos datos utiliza el algoritmo de aprendizaje *back propagation* (propagacin hacia atrs), el cual es el mejor algoritmo de aprendizaje hasta la fecha para las redes neuronales para el aprendizaje supervisado, pero no el nico. La idea de juntar los algoritmos genticos con las redes neuronales (GANN) es la de no utilizar el algoritmo de aprendizaje *back propagation*, sino que en su lugar utilizar el propio algoritmo gentico como algoritmo de aprendizaje para la red neuronal. Como hemos dicho, el algoritmo de aprendizaje *back propagation* es el mejor que se conoce a da de hoy para el aprendizaje supervisado en las redes neuronales, as que cul es el objetivo de utilizar un algoritmo gentico como algoritmo de aprendizaje? Hay varias razones, y una de las principales razones es que al utilizar un algoritmo gentico como algoritmo de aprendizaje para las redes neuronales dejamos de estar utilizando aprendizaje supervisado y pasamos a estar utilizando **aprendizaje por refuerzo**.

El aprendizaje por refuerzo no necesita de datos etiquetados como s lo necesita el aprendizaje supervisado. Este en su lugar necesita una medida de como de bien est comportndose el algoritmo, lo cual encaja a la perfeccin con la funcin de *fitness*, que es justamente lo que hace. La idea es codificar de alguna manera las redes neuronales de manera que un algoritmo gentico pueda tratarlas, modificarlas, etc. Tenemos que conseguir el genotipo de la red neuronal, y la forma de hacerlo es con la codificacin de la red. Hay muchos tipos de codificaciones, siendo la ms tpica la **codificacin binaria**, la cual coge N bits para representar un peso de la red neuronal y entonces ya ejecuta todas las fases que hemos comentado. Esta codificacin es la ms tpica, pero tambin es muy restrictiva y nos aporta poca flexibilidad. En nuestro caso, la codificacin que hemos hecho de la red neuronal es una codificacin de **nmeros reales**, en la cual cada peso de la red se trata como lo que es, un nmero real. Esta codificacin tambin trae problemas, siendo uno de ellos el lmite superior e inferior que se le da a los pesos, ya que en funcin de esos lmites podremos generar unos pesos en un rango o en otro.

Por la forma en la que tenemos nuestra implementacin de la red neuronal, en la cual hemos implementado el perceptrn multicapa, la manera ms sencilla para representar la red, para codificarla, era coger nuestras matrices de pesos

y aplanarlas para obtener un vector. Respecto a los *biases* hicimos lo mismo, coger la matriz que contiene todos los *biases*, aplanarla, y aadir todos los pesos al final del vector que contiene los pesos. Como luego es necesario decodificar y volver a obtener nuestras matrices de pesos y *biases*, con el vector de topologa, que es el vector que nos indica cual es la topologa de esa codificacin en concreto, se puede volver a desaplanar sin ningn problema. El aplanamiento y unin de matrices de pesos y *biases* es sencillo, pero el desaplanar puede no serlo tanto, as que se puede observar el cdigo en el algoritmo 9.

Algorithm 9: Desaplanamiento de las matrices de pesos y *biases*

```

Data: const vector<int>&topology, const Mat &flattened
Result: vector<vector<Mat>>weightsAndBiases
vector<vector<Mat>>res;
vector<Mat>weights;
vector<Mat>bias;
int flattenedCol = 0;
for (size_t i = 0; i < topology.size() - 1; i++) do
    Mat matWeights(topology[i], topology[i + 1]);
    for (int row = 0; row < matWeights.rows(); row++) do
        for (int col = 0; col < matWeights.cols(); col++) do
            matWeights.set(row, col, flattened.get(0, flattenedCol));
            flattenedCol++;
        end
    end
    weights.push_back(matWeights);
end
for (size_t i = 0; i < topology.size() - 1; i++) do
    Mat matBias(1, topology[i + 1]);
    for (int col = 0; col < matBias.cols(); col++) do
        matBias.set(0, col, flattened.get(0, flattenedCol));
        flattenedCol++;
    end
    bias.push_back(matBias);
end
res.push_back(weights); res.push_back(bias);
return res;

```

Una vez ya desaplanados los pesos y *biases* ya solo nos quedaba crear una nueva red neuronal y utilizar *setWeights()* y *setBiases()*.

Ahora vamos a pasar a la parte ms importante de nuestra implementacin

de la GANN, que son los operadores genéticos que hemos empleado. Los operadores genéticos son las operaciones que realizamos para obtener la selección, el emparejamiento y la mutación de los individuos.

Para la selección de los individuos hay varios operadores genéticos, siendo los más famosos los siguientes:

1. Selección proporcional del *fitness* (Fitness proportionate selection) más conocido con el nombre de Roulette Wheel Selection. Este operador de selección consiste en asignarle a cada individuo una probabilidad basada en su *fitness* respecto a la suma total del *fitness* de toda la población. De esta manera, luego se simula la rotación de una ruleta, donde los que mayor probabilidad tienen son los que mayor sector tienen en la ruleta. Es el operador de selección más utilizado y es el que hemos utilizado en nuestra implementación. La fórmula que utiliza es la siguiente:

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j}$$

2. Selección por torneo. Este método intenta simular un torneo, donde la probabilidad de selección varía a cada iteración. Cuando se han ejecutado los torneos que se haya configurado, es cuando se escoge a los mejores individuos (donde K es el número de individuos que participan en el torneo). La fórmula que utiliza es la siguiente:

$$p_0 = p$$

$$p_{iteration} = p * (1 - p)^{iteration}$$

3. Selección basada en recompensas. Este método recompensa a los individuos que han hecho algo de forma correcta. Hay muchas fórmulas que pueden ser utilizadas como recompensa, una básica es la siguiente:

$$individual^{(generation+1)}_{selected} \Rightarrow reward^{(generation)} = 1$$

Respecto al emparejamiento de individuos, hay muchas maneras de hacerlo. Nosotros utilizamos 2 enfoques diferentes, y en ambos era para el emparejamiento de 2 individuos, es decir, de 2 redes neuronales. En el primero de ellos lo que hacamos era alternar entre coger un peso de una red y luego otro de la otra, con lo que acabamos con una nueva red con cerca del 50% del genotipo de las 2 redes que se utilizaban como padres. El segundo enfoque lo que hacamos era coger los 2 pesos de la misma posición de los padres

y hacer una media ponderada, lo cual nos daba ms seguridad de que los resultados obtenidos en la funcin de *fitness* de la nueva red neuronal iban a ser ms prximos a los de las redes padre, ya que si ambas redes neuronales padre obtenan buenos valores de *fitness* la probabilidad de que sus pesos se parecieran es muy alta, como veremos que sucede en los experimentos realizados. Esta ltimo enfoque es el que se ha quedado en la ltima versin del GANN. Podemos ver ambos enfoques en el algoritmo 10.

Algorithm 10: Operadores genticos de emparejamiento empleados

```

Data: const DNA &a, double ownFitness, double aFitness
Result: DNA newIndividual
DNA res(a.genes.rows(), a.genes.cols());
double random = UtilG :: getRandomDouble(0.0, 1.0);
if random <= this->crossoverRate then
    res.setMutationRate(this->mutationRate);
    res.setCrossoverRate(this->crossoverRate);
    if a.genes.rows() == this->genes.rows() && a.genes.cols() ==
        this->genes.cols() then
        for (int row = 0; row < a.genes.rows(); row++) do
            for (int col = 0; col < a.genes.cols(); col++) do
                // First approach
                if (row + col)%2 == 0 then
                    res.genes.set(row, col, this->
                        genes.get(row, col));
                else
                    res.genes.set(row, col, a.genes.get(row, col));
                end
                // Second and definitive approach. Average
                res.genes.set(row, col, (a.genes.get(row, col) + this->
                    genes.get(row, col))/2.0);
            end
        end
    end
else
    if ownFitness > aFitness then
        | res = *this;
    else
        | res = a;
    end
end
return res;

```

Por ltimo, al igual que con los operadores genticos de emparejamiento, hay muchas posibilidades para los operadores genticos de mutacin. En nuestra implementacin aplicamos 2 tipos de mutaciones. La primera mutacin que aplicamos es cambiar un nmero real por otro, por defecto en el rango $[-1.0, 1.0]$, aunque es posible multiplicar dicho rango por factor, dando lugar a $[-1.0 \cdot \text{factor}, 1.0 \cdot \text{factor}]$. El otro operador genético de mutación que hemos empleado ha sido la permutación. La implementación de ambos operadores se muestra en el algoritmo 11.

Algorithm 11: Operadores genéticos de mutación empleados

```
// First mutation. Replace a real - number
for (int row = 0; row < this->genes.rows(); row++) do
    for (int col = 0; col < this->genes.cols(); col++) do
        double random = UtilG::getRandomDouble(0.0, 1.0);
        if random <= this->mutationRate then
            this->genes.set(row, col, UtilG::
                getRandomDouble((double)factor *
                    -1.0, (double)factor * 1.0));
        end
    end
end

// Second mutation. Permute 2 weights/biases
for (int row = 0; row < this->genes.rows(); row++) do
    for (int col = 0; col < this->genes.cols(); col++) do
        double random = UtilG::getRandomDouble(0.0, 1.0);
        if random <= this->mutationRate then
            int randomRow1 = rand()%this->genes.rows(); int
            randomCol1 = rand()%this->genes.cols(); int
            randomRow2 = rand()%this->genes.rows(); int
            randomCol2 = rand()%this->genes.cols(); double
            value1 = this->genes.get(randomRow1, randomCol1);
            double
            value2 = this->genes.get(randomRow2, randomCol2);
            this->genes.set(randomRow1, randomCol1, value2);
            this->genes.set(randomRow2, randomCol2, value1);
        end
    end
end
```

Una vez ya hemos visto los operadores genéticos que hemos empleado solo nos queda por ver dos cosas más: la función de *fitness* y otras técnicas que hemos implementado para evitar la rápida convergencia o para evitar la pérdida de los individuos más aptos.

Vamos a empezar por las funciones de *fitness* que hemos empleado, las cuales van ligadas al problema que queremos resolver, y en este caso el problema que queremos resolver es enseñar a un algoritmo a jugar lo mejor posible a los 4 videojuegos para la videoconsola Atari 2600 que comentamos al inicio (i.e. breakout, boxing, demon attack y stargunner) de esta memoria. Las funciones de *fitness* que hemos empleado han ido cambiando mucho, ya que cuando utilizábamos una u otra, veíamos que los resultados variaban, y mucho. A continuación, mostramos las funciones de *fitness* que hemos empleado para cada uno de los videojuegos.

1. Breakout:

- (a) $fitness = score$; // Ha funcionado a la perfección desde el principio.

2. Boxing:

- (a) $fitness = score$; // Se quedaba quieto todo el rato y aprendía a dar golpes de vez en cuando.
- (b) $fitness = score + (double)scoreP1 * ((double)scoreP1 / (double)(scoreP2 + 1))$; // Intentábamos recomensar a P1 cuando la diferencia de resultados no fuera exagerada a favor de P2 y que, cuando mejor lo hiciera P1, más se le recompensaba. Obtuvo mejores resultados que el anterior, llegando a obtener puntuaciones igualadas, pero sin muy buenos resultados.
- (c) Algoritmo 12. Se mueve hacia la derecha y acorrala todo el rato al oponente, consiguiendo doble de puntos. Los resultados suelen ser de 100 a 10 ganando P1. A pesar de obtener unas puntuaciones muy buenas no parece que haya aprendido muy bien a jugar: solo avanza y golpea.

3. Demon Attack:

- (a) $fitness = score$; // Obtenía unas puntuaciones de 3000, las mejores que hemos obtenido, pero la jugabilidad por parte del bot era pésima: se quedaba en el centro y disparaba.

- (b) $fitness = score * ((moveLeft/100.0) * (moveRight/100.0)); //$
Se le recompensaba al bot cuando ms se moviera. El problema era que aprendia a matar a algn enemigo muy esporadicamente a moverse todo el rato.
- (c) $fitness = ((double)(score^2)/(500.0^2)) * ((moveLeft/1000.0) * (moveRight/1000.0)); //$ Intentamos arreglar el problema anterior penalizando un poco el moverse y aumentando ms la recompensa por disparar y matar a los enemigos. Factores de multiplicacin muy desbalanceados. Muy malos resultados.
- (d) $fitness = ((double)(score^2)/(500.0^2)) * ((moveLeft/1000) * (moveRight/1000)) * ((score/1000) + 1); //$ Intentamos que la puntuacin no fuera un factor extremadamente determinante como en la anterior funcin. Resultados aceptables en cuanto a jugabilidad.
- (e) $fitness = (score/10) * ((moveLeft/1000) * (moveRight/1000)); //$ Intento de simplificar todo lo anterior. Sin xito.
- (f) $fitness = (score/10) * ((moveLeft/(int)((steps*500.0)/15000.0)) * (moveRight/(int)((steps * 500.0)/15000.0)) * (shooting/(int)((steps * 1000.0)/15000.0))); //$ Intento de volver a la compleja funcin de antes pero esta vez recompensando el que bot disparara, sin importar si mataba. El bot aprenda a disparar y moverse todo el rato en el punto ciego del juego (lago izquierdo normalmente). Los movimiento que hacia eran izquierda - derecha constantemente, con lo que maximizaba todos los valores y de vez en cuando mataba a algn enemigo. Se terminaba el tiempo y obtena un *fitness* muy elevado y una puntuacin muy baja. Adems, intentamos que el *fitness* obtenido fuera en funcin de lo que haba tardado en jugar el bot.
- (g) $fitness = (score/100)^2 * ((moveLeft + moveRight)/1000) * (shooting/2000) * (min(moveLeft, moveRight)/100); //$ Intento de obtener un mnimo en los disparos y en el movimiento y a la vez simplificar la funcin. Malos resultados.
- (h) $fitness = (score/10) * (shooting/500) * (min(moveLeft, moveRight)/100); //$ Intento de mejorar la funcin anterior restndole importancia a la puntuacin e intentando obtener un valor mnimo en el movimiento y en los disparos. Resultados decentes: cerca de los 1000 puntos.
- (i) $fitness = (score/10)^2 * (min(moveLeft, moveRight)/100); //$ Simplificacin de la frmula. Al elevar al cuadrado la puntuacin,

forzamos a que el bot disparara sin tener que poner explícitamente los disparos en la función. Mejores resultados que la fórmula anterior.

- (j) Algoritmo 13. Con esta función de *fitness* le pasamos un nuevo valor que calculamos gracias a las coordenadas X de los enemigos y hacemos que cuando siga al enemigo que se le indica, se le recompense. Muy buenos resultados: sigue al enemigo más cerca vivo en todo momento (si ya no aparece el más cercano, el siguiente objetivo es el de en medio, y luego el final. Cuando empieza la fase de las moscas, va buscando cuál es la mosca más cercana viva, teniendo en cuenta la mosca que se acerca al jugador como si fuera una bala), intentando eliminar al enemigo más cercano, ya que es el más fácil de dispararle y acertar. El problema es que no aprende muy bien a evitar los disparos del enemigo más cercano.

4. StarGunner:

- (a) $fitness = score$; // No muy buenos resultados, ya que la puntuación al ser de 100 en 100 al eliminar un objetivo, cuando un individuo eliminaba a un objetivo más que otro, su probabilidad de ser seleccionado era muy elevada aunque el bot fuera mucho más malo.
- (b) $fitness = score/100$; // Intento de mejorar el problema anterior. El bot juega bastante bien: en algunas ocasiones se queda quieto pero en otras empieza a seguir a los objetivos, consiguiendo unas puntuaciones de 3000.

Algorithm 12: Mejor funcin de *fitness* para boxing encontrada

```
if score < 0 then
    fitness = score * (1 -
        (double)moveLeft/(double)steps) * (1 -
        (double)moveUp/(double)steps) * (1 -
        (double)moveRight/(double)steps) * (1 -
        (double)moveDown/(double)steps) + (double)scoreP1 *
        ((double)scoreP1/(double)(scoreP2 + 1));
else
    fitness = score * (
        (double)moveLeft/(double)steps) *
        ((double)moveUp/(double)steps) *
        ((double)moveRight/(double)steps) *
        ((double)moveDown/(double)steps) + (double)scoreP1 *
        ((double)scoreP1/(double)(scoreP2 + 1));
end
```

Algorithm 13: Funcin de *fitness* del Demon Attack que sigue al enemigo

```
if min(moveLeft, moveRight)! = 0 then
    double punch = (double)inRangeOfEnemy/(double)steps;
    double dscore = (double)score;
    punch* = punch;
    dscore/ = 10;
    fitness = dscore * punch;
else
    | fitness = 0.0;
end
```

Una vez ya comentadas las funciones de *fitness* que hemos probado nos falta comentar las otras ténicas que hemos implementado. Otras ténicas que hemos implementado en pro del GANN han sido las siguientes:

- Elitismo. El elitismo es una ténica muy sencilla la cual nos permite no perder los mejores individuos de una determinada generacin. Consiste en coger el mejor individuo de una generacin y pasarlo directamente a la siguiente generacin N veces. Un elitismo de 0 significa que no hay elitismo.
- Inicializacin de Elitismo. El elitismo se aplica a partir de la segunda generacin, pero no en la primera. Un intento de una rpida convergencia, cosa que casi nunca es buena en los gnticos, ha sido nuestra implementacin de utilizar elitismo en la creacin de la poblacin. Por s solo no funciona muy bien, ya que para funcionar bien debera de generarse individuos con los que pudiera hacer un emparejamiento lo suficientemente bueno y para ello normalmente los individuos deben parecerse. Ahora bien, cuando se utiliza junto con una distribucin de probabilidad obtiene buenos resultados, pero solo debera de aplicarse a problemas sencillos, ya que los complejos necesitan de una convergencia ms lenta, donde haya ms diversidad.
- Ténicas de nicho. Estas ténicas intentan que la diversidad de la poblacin aumente y lo hace castigando a aquellos individuos que se parezcan mucho y premiando a aquellos individuos que aporten novedad a la poblacin. La frmula que utiliza para ello es la siguiente:

$$fitness_i = \frac{fitness_i}{\sum_{j=1}^N individual_i \simeq individual_j ? 1 : 0}$$

- Creacin de poblacin por rango de valores (distribucin de probabilidad). Inicialmente se generan pesos y *biases* aleatorios entre un rango, pero si supiramos la distribucin de probabilidad que sigue cierto problema, una buena inicializacin podra hacer que obtuviramos mejores resultados antes. El problema es que hay que ir obteniendo buenos resultados poco a poco y, entonces, ir utilizando esa distribucin para obtener buenos resultados. En los experimentos vemos un ejemplo de esto.

Tanto el elitismo, la inicializacin con elitismo y la inicializacin con rangos son sencillos de entender, por lo que no vamos a mostrar la implementacin aqu en la documentacin. En cambio, vamos a explicar qu criterio hemos utilizado para decir que dos individuos son similares. El criterio que hemos utilizado ha sido el de asignar a cada individuo un id de nicho, que es lo que se suele hacer con estas tcnicas. El id del nicho que se asigna a cada individuo depende de la puntuacin y de los pasos realizados al jugar alguno de los videojuegos. Dependiendo de esos dos valores, y del valor mximo de puntuacin de toda la poblacin de una generacin concreta y el valor mximo de pasos de toda la poblacin de la misma generacin y ponemos los puntos obtenidos en el plano. Entonces dividimos el eje X en 10 regiones, y hacemos lo mismo con el eje Y. Cada regin de las 100 obtenidas es un id de nicho diferente. En la figura 30 se muestra lo que estamos explicando poniendo como ejemplo el videojuego Breakout (suponiendo que los pasos son 7500 y estn en el eje Y y la puntuacin mxima obtenida por alg n individuo ha sido la mxima, la cual es 864).

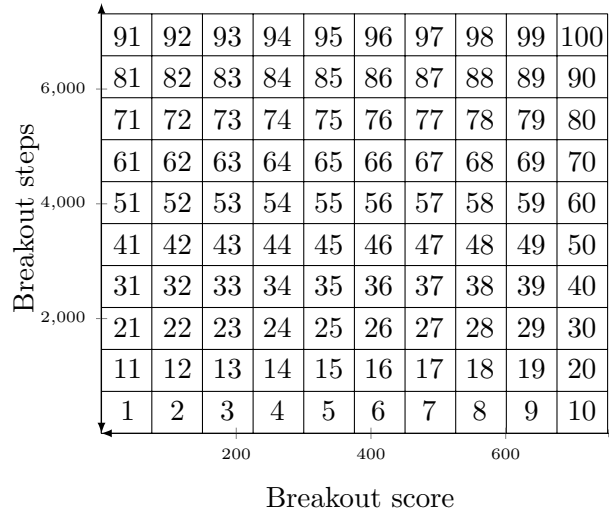


Figure 30: ID de nicho en Breakout

2 Manual de utilizacin

Antes de empezar con el manual de utilizacin para las tecnologas concretas desarrolladas hay que tener en cuenta dos cuestiones fundamentales:

1. La tecnologa que hemos desarrollado se implementado bajo entorno linux, concretamente todo est probado en la distribucin *Manjaro Linux*. No se garantiza el correcto funcionamiento de la tecnologa en otros sistemas.
2. La tecnologa que hemos desarrollado utiliza el entorno ALE (*Arcade Learning Environment*), por lo que lo que es necesario para el funcionamiento de los experimentos y de la tecnologa en s.

Respecto a ALE, es necesario saber dnde est instalado, por lo que hemos utilizado la variable de entorno del sistema **ALEPath** para localizar el entorno y poder compilar y utilizar la tecnologa sin problema. La manera de indicar donde est ALE, si por ejemplo estuviera en la ruta `/home/UDLearn/Escritorio/ALE`, sera con el siguiente comando desde la terminal: `export ALEPath=/home/UDLearn/Escritorio/ALE`. De esta manera ya estara todo listo para poder compilar la tecnologa desarrollada y utilizarla.

Para poder utilizar la tecnologa solo es necesario compilar, lo cual es sencillo ya que solo habr que ejecutar `make` en la terminal, ya que cada una de las tecnologas desarrolladas viene con su correspondiente fichero `makefile`.

Una vez ya se ha indicado donde est ALE y compilada la tecnologa que se quiera utilizar, en los siguientes apartados se indica el uso concreto de cada una de ellas.

2.1 Bots y bots naïve

2.2 Red Neuronal

2.3 Red Neuronal y Gentico: *GANN*

La utilizacin del GANN es muy sencilla. Lo primero que habra que hacer es compilar, lo cual haciendo *make* desde la terminal ya lo tendremos (consultar seccin 2 para ms detalles). A continuacin, se detallan los modos de uso.

Con el GANN tenemos 2 modos de funcionamiento:

1. Modo de **entrenamiento**: en este modo lo que se hace es utilizar el GANN para obtener resultados entrenando a cualquiera de los 4 juegos disponibles. Mientras se ejecuta se va mostrando la informacin del progreso y se va guardando un registro con la informacin ms importante. Cuando termina de ejecutarse, guarda los pesos de la mejor red encontrada.
2. Modo de **ejecucin**: en este modo lo nico que se hace es ejecutar la mejor red que hemos encontrado conforme hemos ido haciendo pruebas. Los pesos estn directamente escritos en el cdigo, as que lo que se ver ser directamente el juego elegido y el mejor bot que hemos encontrado jugando.

Para seleccionar el juego, la codificacin que hemos escogido es un nmero para cada juego. La codificacin es la siguiente:

- Breakout = 1
- Boxing = 2
- Demon Attack = 3
- StarGunner = 4

La ejecucin de ambos modos es muy sencilla, y se describe a continuacin:

1. Modo de Entrenamiento:
 - (a) De modo general:
 - i. *./main juego generaciones poblacin nombreFicheroRegistros*
 - (b) Ejemplo: Breakout con 5 generacines, 100 individuos en la poblacin y archivo llamado "breakoutRecords.txt" (registro) y archivo "breakoutRecords.weights" (pesos).
 - i. *./main 1 5 100 breakoutRecords*

Modo de ejecucin:

- (a) De modo general:
 - i. *./main juego*
- (b) Ejemplo: ver el bot del Breakout.
 - i. *./main 1*

3 Experimentos realizados y resultados obtenidos

3.1 Experimento con el Perceptrn

El experimento que hemos realizado con el perceptrn ha sido sencillo, con el objetivo de comprobar que el funcionamiento era el esperado y, sobre todo, ver grficamente como el PLA iba evolucionando los pesos y cual iba siendo la evolucin. La funcin que hemos intentado aproximar con el perceptrn ha sido la funcin AND, la cual se describe en el algoritmo 14.

Algorithm 14: Funcin AND

```
if  $a * b == 1$  then  
  | return 1;  
else  
  | return 0;  
end
```

La funcin AND es linealmente separable, por lo que el perceptrn debera de poder separar los puntos eventualmente. En la figura 31 se muestran los puntos que queremos separar (como subndice tienen el valor que se espera por el perceptrn).

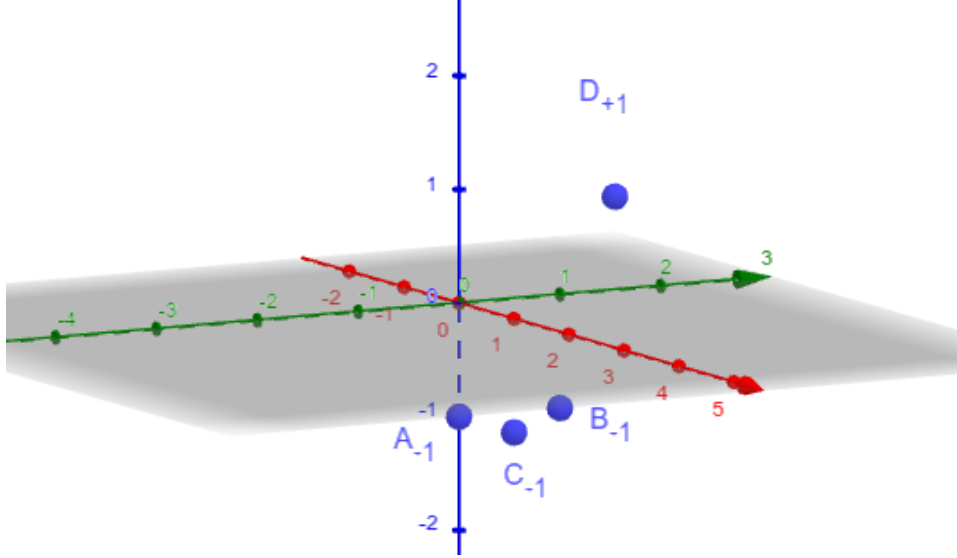
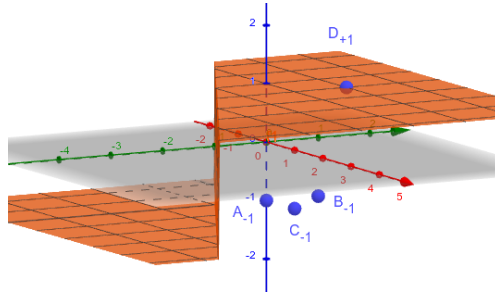
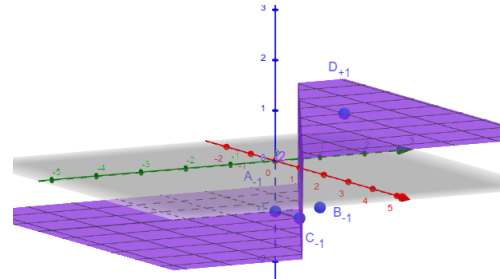


Figure 31: Puntos de la funcin AND

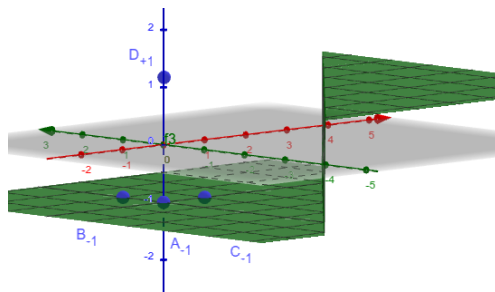
En la figura 32 podemos ver como el PLA ha ido ajustando los pesos para buscar el plano que separa correctamente estos puntos.



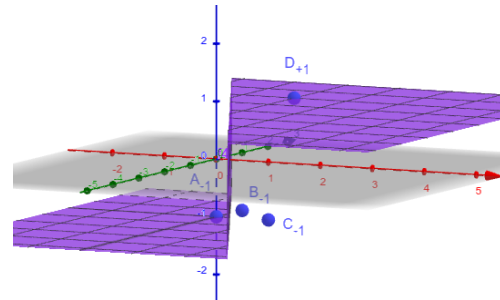
1a iteracin PLA (25% acierto)



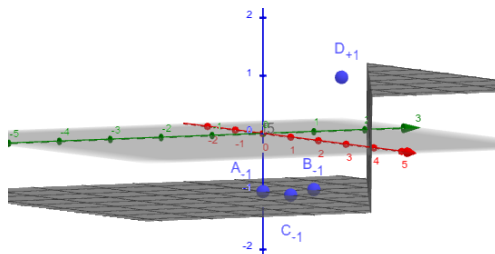
2a iteracin PLA (75% acierto)



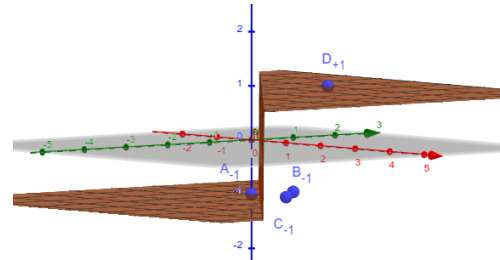
3a iteracin PLA (75% acierto)



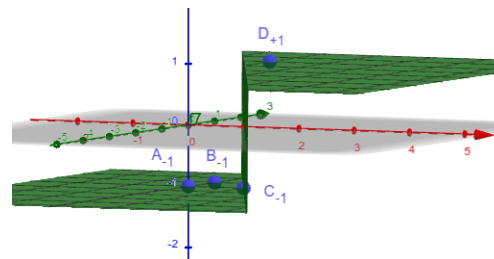
4a iteracin PLA (50% acierto)



5a iteracin PLA (75% acierto)



6a iteracin PLA (50% acierto)



7a iteracin PLA (100% acierto)

Figure 32: Iteraciones del PLA para la funcin AND

Algo interesante que podemos observar con este experimento es que el porcentaje de acierto no siempre aumenta (la razón ya se explicó en la sección 1.3), por lo que si en este experimento hubiéramos configurado un número de pocas igual a 5 habríamos obtenido como resultado final un 50% de aciertos, no un 75%, en el caso de que no hubiéramos implementado el algoritmo del *Pocket*.

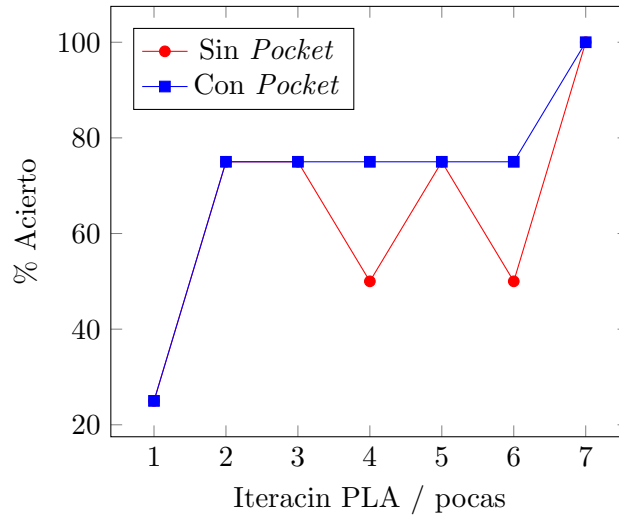


Figure 33: Resultados del PLA con y sin *Pocket*

Como ya hemos comentado, el algoritmo del *Pocket* también podemos ver la utilidad en este sencillo experimento. En la figura 33 se ve cómo a cada iteración que se ejecuta del PLA (también se puede ver como una poca) dependiendo de si tenemos el algoritmo del *Pocket* o no, al final obtendremos un resultado u otro. Esto es esencial cuando nos enfrentamos a un problema que no es linealmente separable y la mayoría de los problemas a los que se les puede aplicar el perceptrón y obtener unos resultados aceptables no son linealmente separables.

3.2 Experimentos con la Red Neuronal

3.3 Experimentos con la Red Neuronal: *Dropout*

3.4 Experimentos con la Red Neuronal: *Cross Validation*

10

3.5 Experimentos con la Red Neuronal y Gentico: *GANN*

Los experimentos realizados con el GANN han consistido en la variación de diferentes parámetros como pueden ser la activación del elitismo, cambiar la probabilidad de mutación, etc. y ver cómo se convergen antes, o cómo se alcanzan mejores puntuaciones, etc.

El primer experimento vamos a ver los mejores resultados obtenidos en el videojuego Breakout cuando tenemos elitismo y cuando no lo tenemos. Como podemos observar en la figura 34, el elitismo hace que no perdamos el mejor individuo de una generación, como sí que sucede cuando no lo utilizamos. El elitismo es parecido al algoritmo del *Pocket* que vimos en la sección 1.3 con la teoría y en la sección 3.1 con los experimentos.

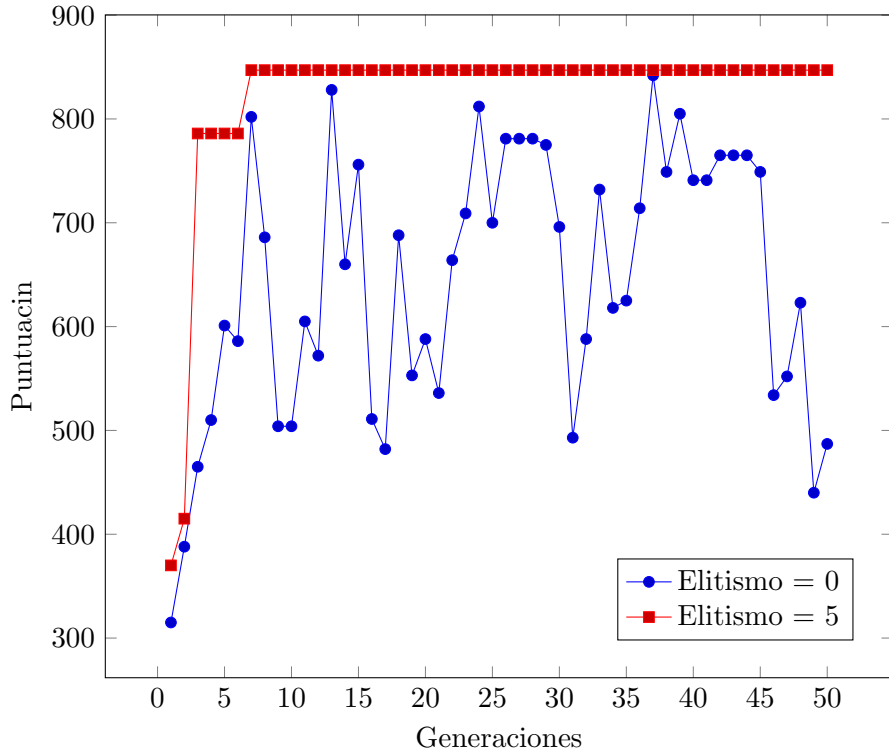


Figure 34: Breakout con y sin Elitismo

Otro de los experimentos que realizamos fue inicializar los pesos en el rango $[-2.0, 2.0]$ en vez de $[-1.0, 1.0]$. Este cambio depende mucho del problema, ya que al aumentar el rango al doble de posibles valores deberá de ser más difícil de obtener una mejor inicialización, pero eventualmente, como

ahora tenemos pesos que no estn solo limitados al rango $[-1.0, 1.0]$ sino que ahora pueden tener un rango mayor de valores, puede que ahora tengan la posibilidad de obtener mejores resultados que antes. Esto es justo lo que sucede en la figura 35. Hay que destacar que en ambos hay un elitismo de 5 y el resto de parmetros son iguales para poder comparar los resultados. Lo nico que cambia es el factor por el que se multiplica el rango de los nmeros reales. El resultado al final es que converge a una puntuacin mayor a pesar de tener que buscar unos pesos y *biases* en un rango de valores ms amplios, por lo que quiz en tener ese rango de valores le ha permitido aprender mejor a como jugar.

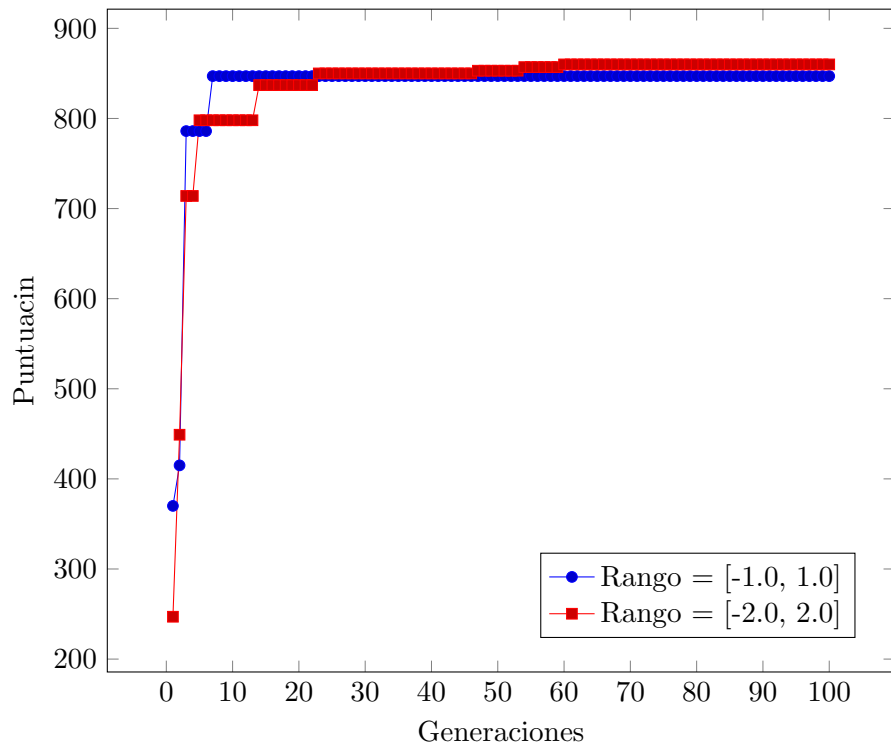


Figure 35: Breakout con diferente rango de valores reales

Una vez tenemos muchos individuos que han jugado bien al videojuego podemos plantearnos lo que ya se coment en la seccin 1.4.1, en la que se habl de la inicializacin con una distribucin de probabilidad. El modo en el que nosotros hemos realizado este experimento es cogiendo los 10 pesos de los 7 mejores individuos que hemos encontrado, y hemos unido los puntos para comprobar si se seguia un patrón en los pesos, algo que pudiramos utilizar

como distribucion de probabilidad. En la figura 36 podemos ver que los 10 pesos de los mejores individuos claramente siguen un patrón: la mayoría están próximos y en casi todos los casos se cumple que yendo de un punto a otro, o todos tienen pendiente ascendente o todos tienen pendiente descendente.

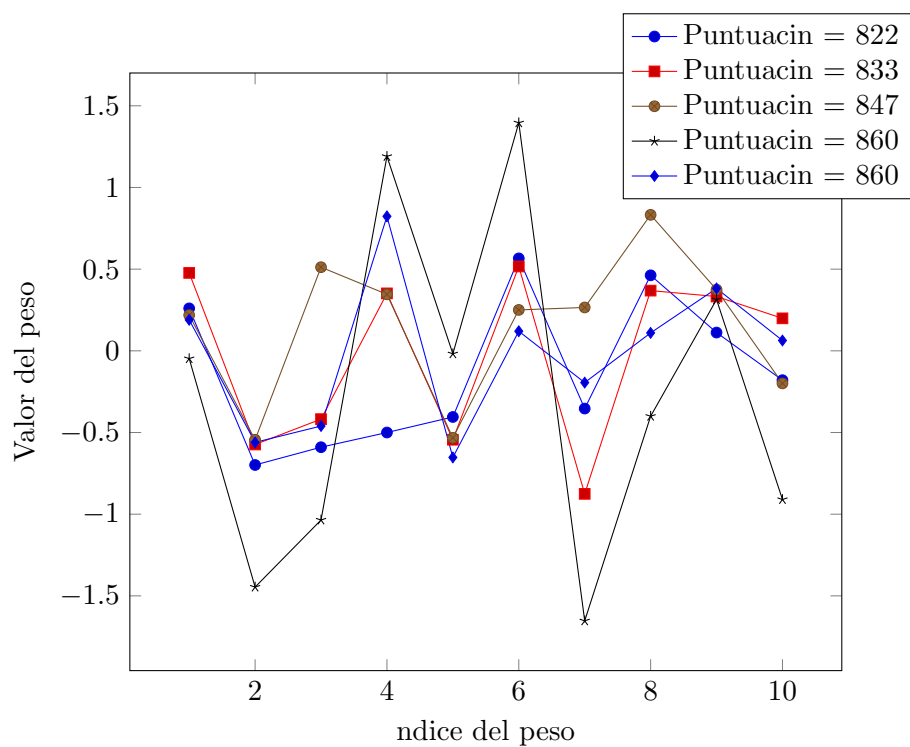


Figure 36: Distribución de probabilidad del Breakout

Si nos fijamos en la figura 37 todavía es más claro el patrón, ya que al interpolar los puntos, queda más claro.

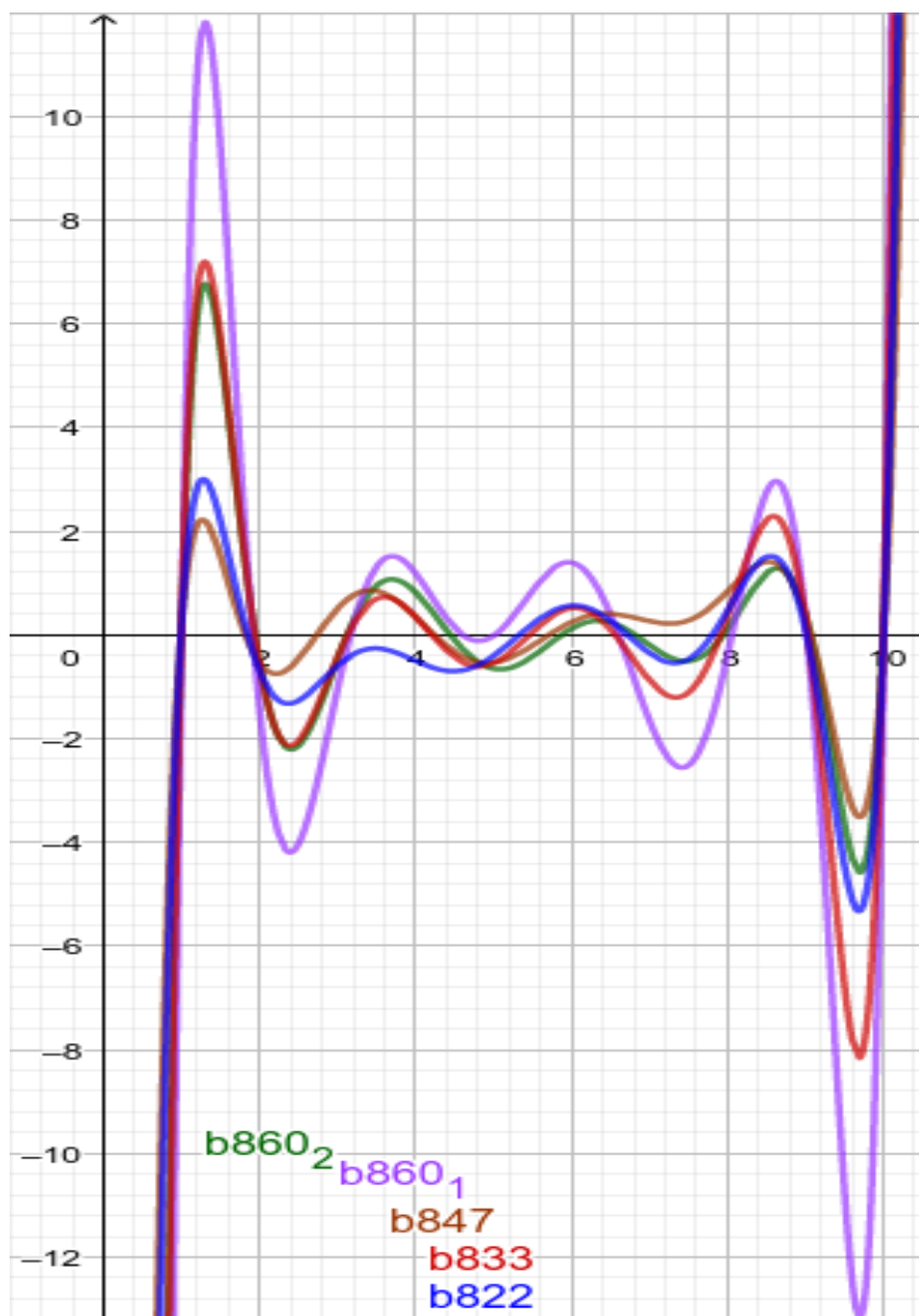


Figure 37: Distribucion de probabilidad del Breakout interpolada

Finalmente, para obtener la distribucin de probabilidad que vamos a emplear cogemos los valores mximos y mnimos de cada uno de los pesos y ya tenemos los valores que podemos darle al GANN para utilizar con la poblacin inicial. En la figura 38 observamos el resultado.

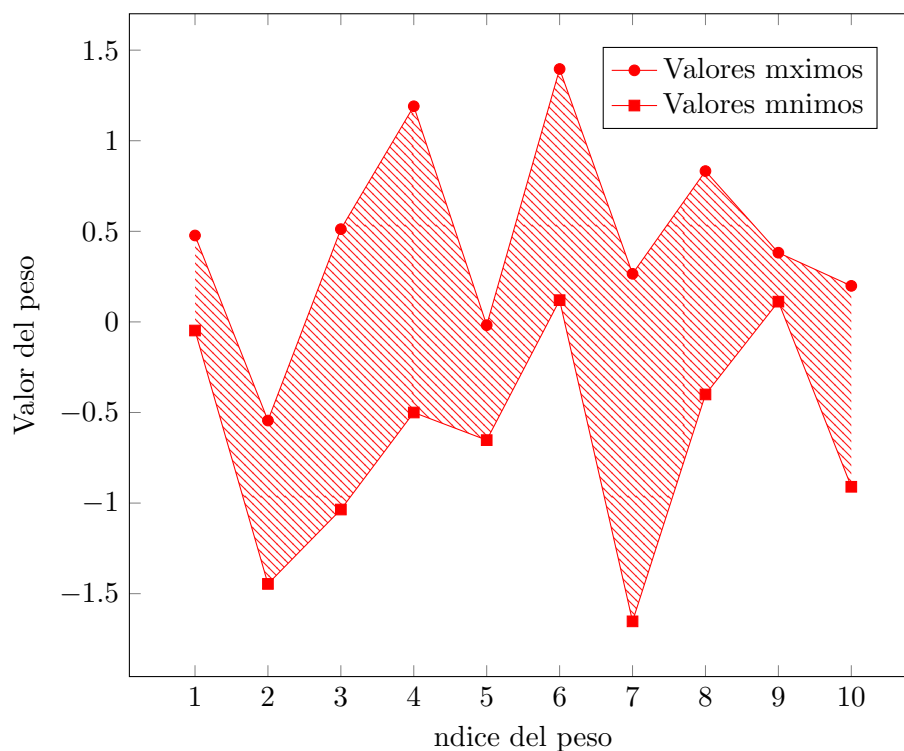


Figure 38: Distribucin de probabilidad del Breakout (final)

Una vez ya tenemos estos valores, podemos aplicar la distribucin de probabilidad que se observa en la figura 38 y as poder valorar si se obtienen mejores resultados o no. Como se puede apreciar en la figura 39 s que se obtienen buenos resultados, por lo que obtendremos una convergencia ms rpida, lo cual no siempre es bueno, pero para problemas sencillos como el Breakout ahorra mucho tiempo.

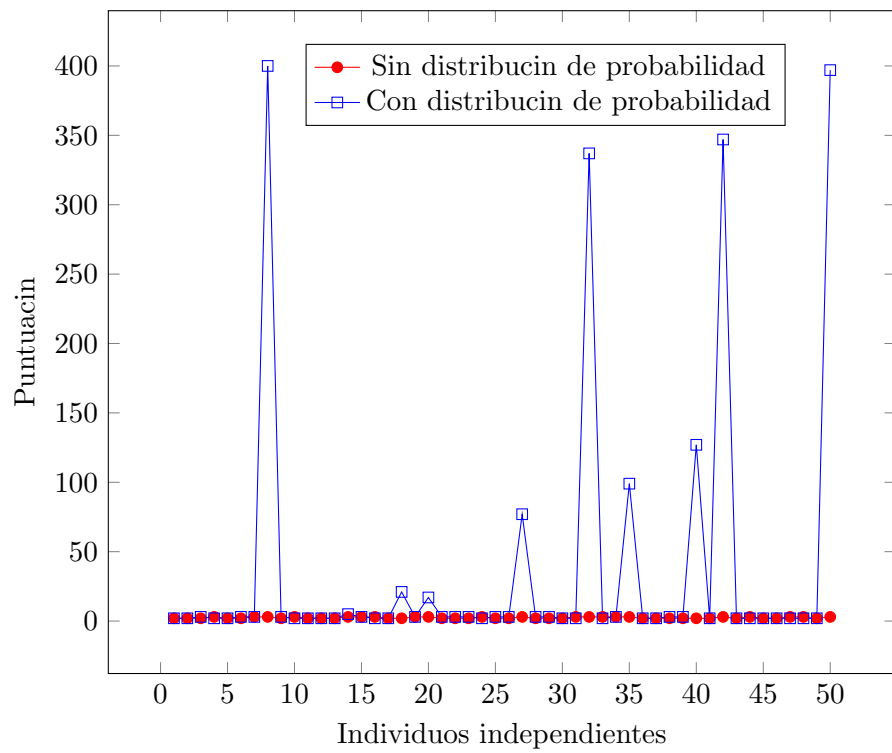


Figure 39: Experimento: Distribucion de probabilidad del Breakout

4 Conclusiones

- 4.1 Valoración personal de las prácticas realizadas
- 4.2 Indicar qué ha echado de menos el alumno en la formación recibida en la Universidad que considera le hubiera ayudado
- 4.3 Posibles sugerencias para mejorar las prácticas de empresa