

The background is a dark blue-grey color. It features several thin, gold-colored lines that form abstract, angular shapes. These lines radiate outwards from the central text area, creating a sense of depth and connectivity. The lines vary in length and orientation, some forming sharp points and others creating more complex, multi-sided shapes.

NEURAL NETWORKS

SIA - TP3 - 2022

TABLA DE CONTENIDOS

1. INTRODUCCIÓN
Introducción al problema y
sus soluciones.

2. PERCEPTRON SIMPLE
Ejercicio 1
Consideraciones
Conclusiones

3. PERCEPTRON SIMPLE
Ejercicio 2
Consideraciones
Conclusiones

4. PERCEPTRON MULTICAPA
Ejercicio 3
Consideraciones
Conclusiones

INTRODUCCIÓN

- Se implementaron los siguientes algoritmos:
 - Perceptrón simple con función de activación escalón:
 - ✓ Función lógica 'Y'
 - ✓ Función lógica 'O exclusivo'
 - Perceptrón simple lineal y perceptrón simple no lineal
 - Perceptrón multicapa:
 - ✓ Función lógica 'O exclusivo'
 - ✓ Discriminar si un número es par, con entradas dadas por el conjunto de números del 0 al 9 representados por imágenes de 5 x 7 píxeles.
 - ✓ Discriminar con 10 unidades de salida de modo que cada salida represente a un dígito, las entradas dadas por el conjunto de números del 0 al 9 representados por imágenes de 5 x 7 píxeles.



PERCEPTRON SIMPLE

Con función de activación escalón

Ejercicio 1

En este ejercicio se busca de separar linealmente los conjuntos generados por las funciones lógicas AND y XOR.

- AND: $x = \{-1, 1\}, \{1, -1\}, \{-1, -1\}, \{1, 1\}$, $y = \{-1, -1, -1, 1\}$
- XOR: $x = \{-1, 1\}, \{1, -1\}, \{-1, -1\}, \{1, 1\}$, $y = \{1, 1, -1, -1\}$

El algoritmo implementado se basa en:

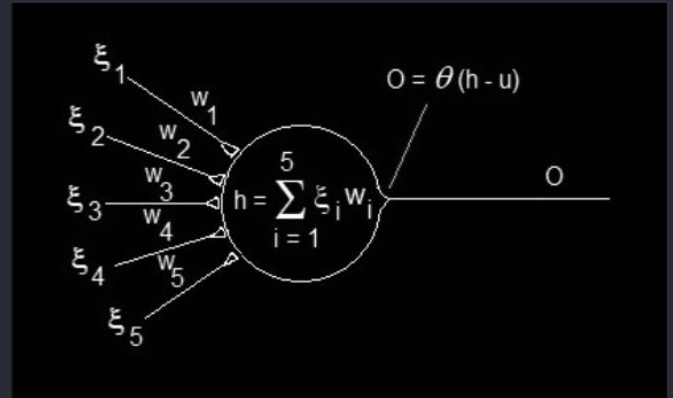
- El Modelo de neurona de McCulloch y Pitts donde la función de activación está dada por la función signo.

Se busca el hiperplano de separación dado por la ecuación:

$$\sum_{i=1}^n w_i * \varepsilon_i - w_0$$

De la cual se puede extraer la ecuación de la recta:

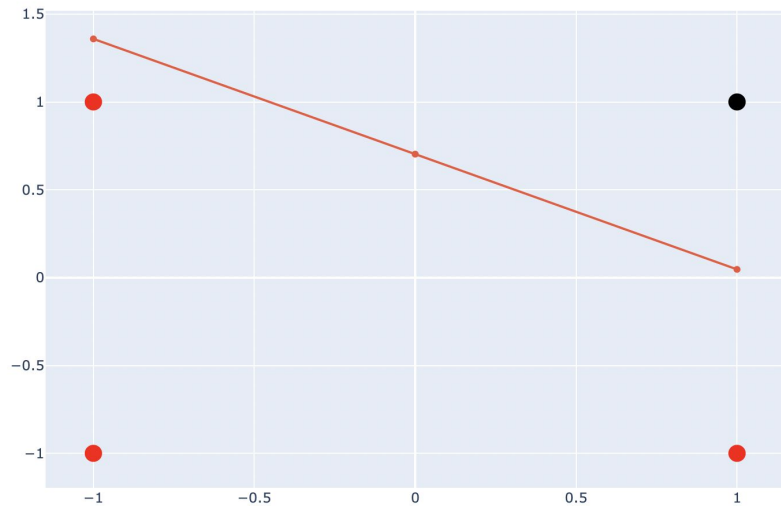
$$y = \frac{-w_0}{w_1} * x - \frac{w_2}{w_1}$$



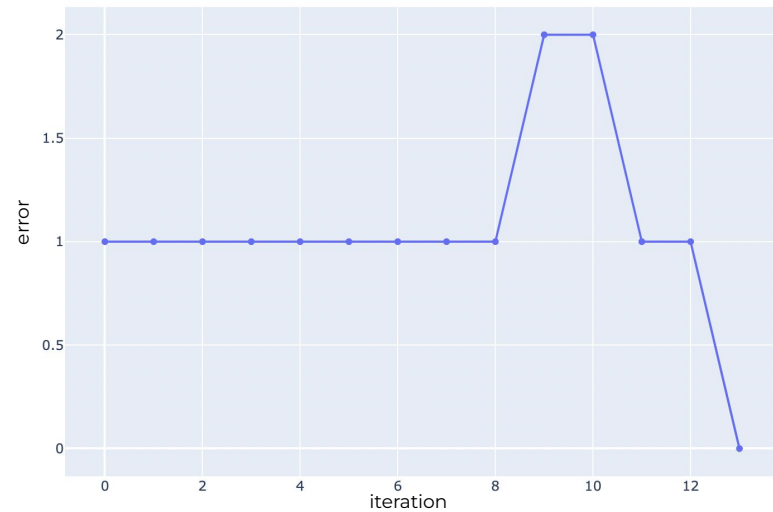
Resultados: AND

$$\eta = 0.1$$

And



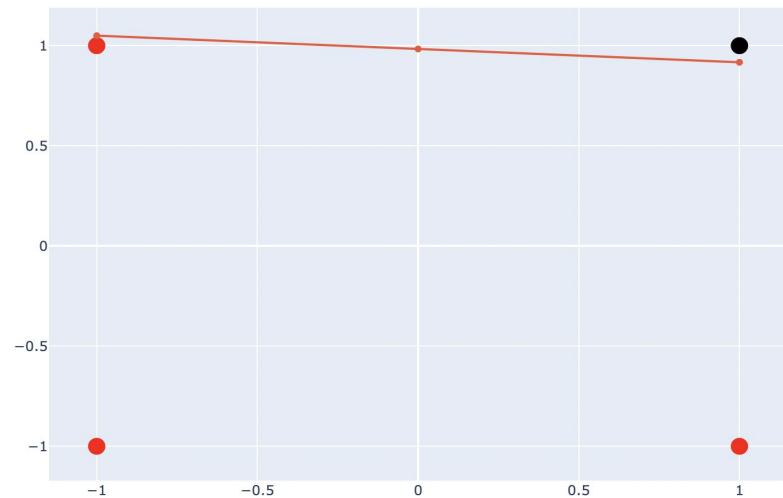
Error



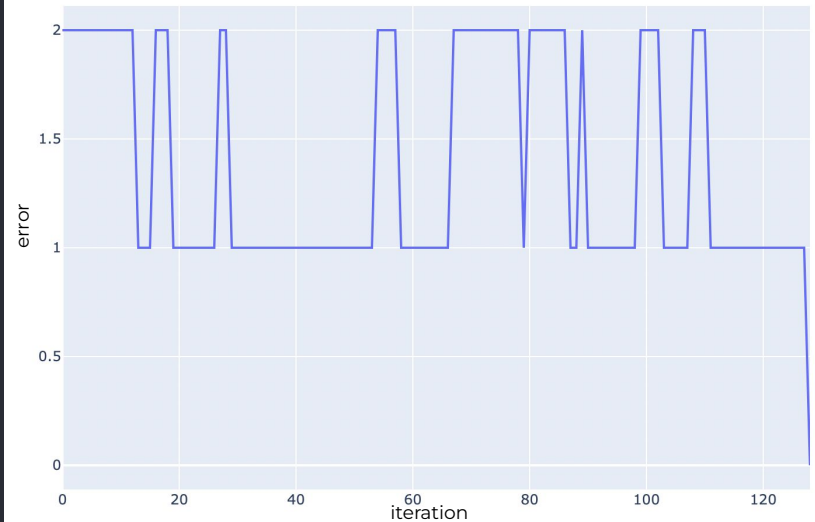
Resultados: AND

$$\eta = 0.01$$

And

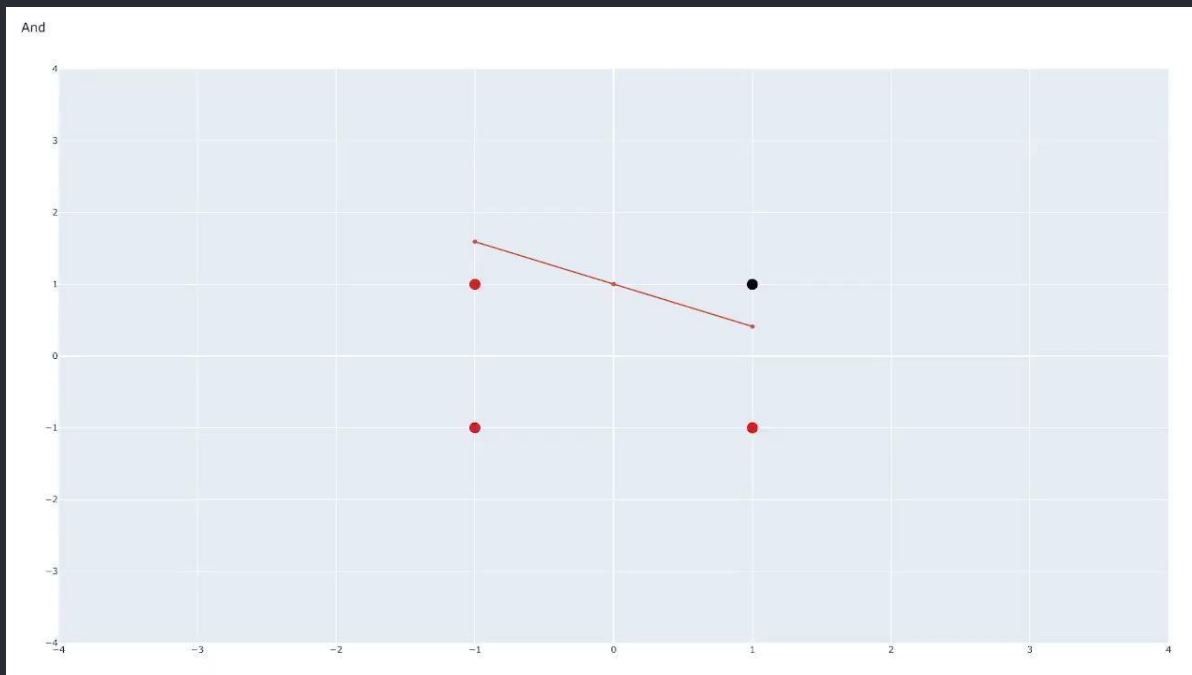


Error



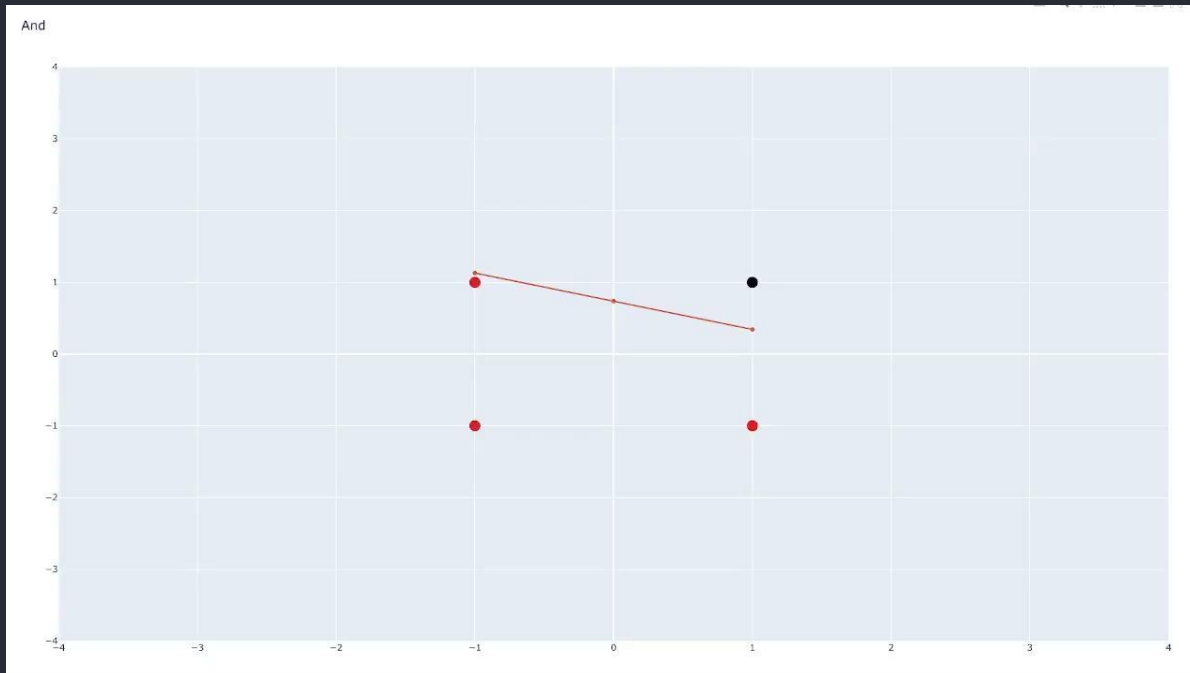
Resultados: AND

ANIMATION $\eta = 0.1$



Resultados: AND

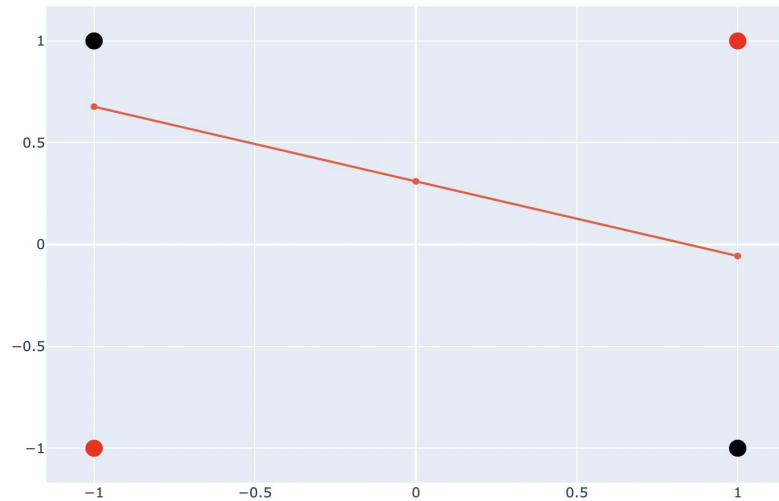
ANIMATION $\eta = 0.01$



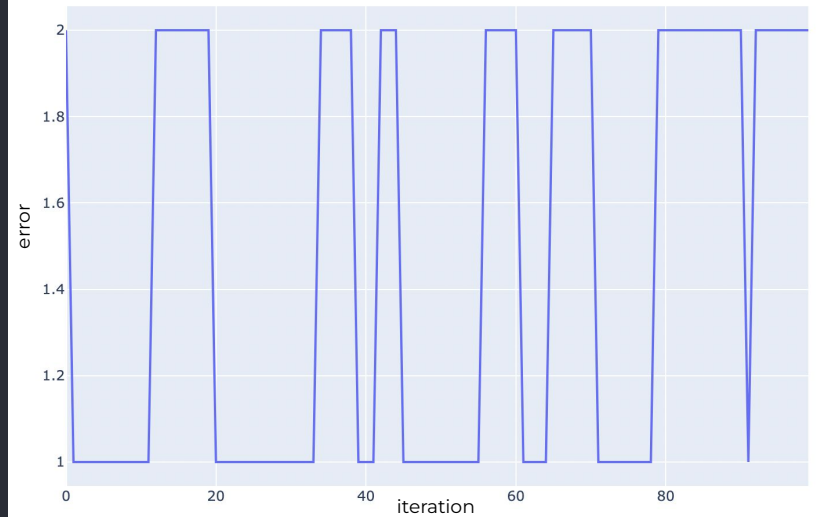
Resultados: XOR

$\eta = 0.1$

Xor



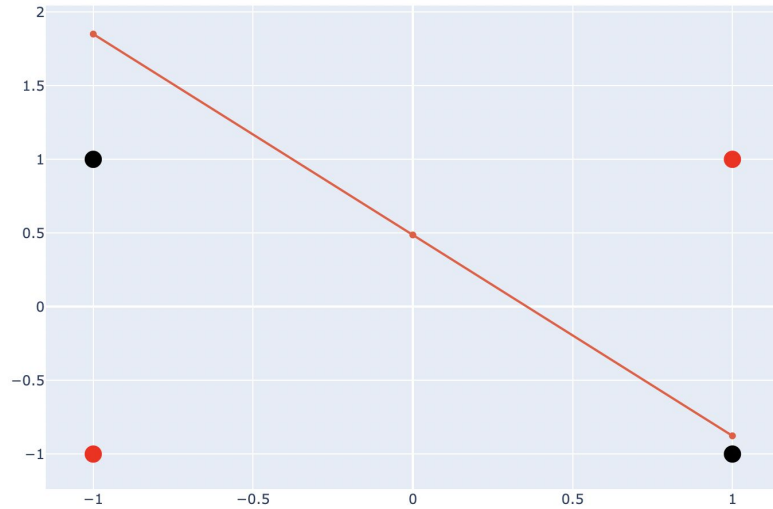
Error



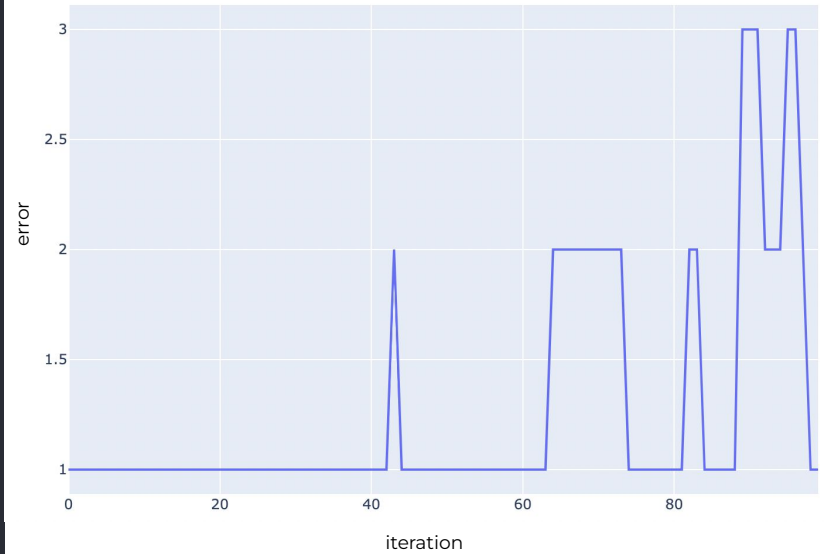
Resultados: XOR

$\eta = 0.01$

Xor



Error



CONCLUSIONES

AND:

- En el caso del AND al ser un problema linealmente separable, el perceptrón simple escalón logró encontrar un hiperplano de separación que divide los datos del problema correctamente.
- A mayor valor de $\eta = 0.01$ se tarda más en converger a la respuesta.

XOR:

- En el caso del XOR, al tratarse de un problema donde los conjuntos no son linealmente separables, el perceptrón intenta encontrar la solución que mejor se ajuste.



PERCEPTRON SIMPLE

Lineal y No Lineal



Ejercicio 2

- En el perceptrón lineal, en vez de utilizar una función de activación escalonada como en el Ejercicio 1, se utiliza una función lineal, en este caso la función identidad. De esta forma, la salida puede pertenecer a los reales.
- En el perceptrón no lineal, se utilizan funciones sigmoideas, en este caso, la función tangente hiperbólica. A la hora de actualizar los pesos, se utiliza la derivada de la función de activación.

$$\Delta w_i = \eta \sum_{\mu=1}^p (\zeta^\mu - g(h^\mu)) g'(h^\mu) \xi_i^\mu$$

- El conjunto de entrenamiento son 200 tuplas de tres valores reales cuyo valor de salida es un número real.
- Se busca encontrar los valores de w que mejor ajusten datos de entrada con los de salida.

Ejercicio 2

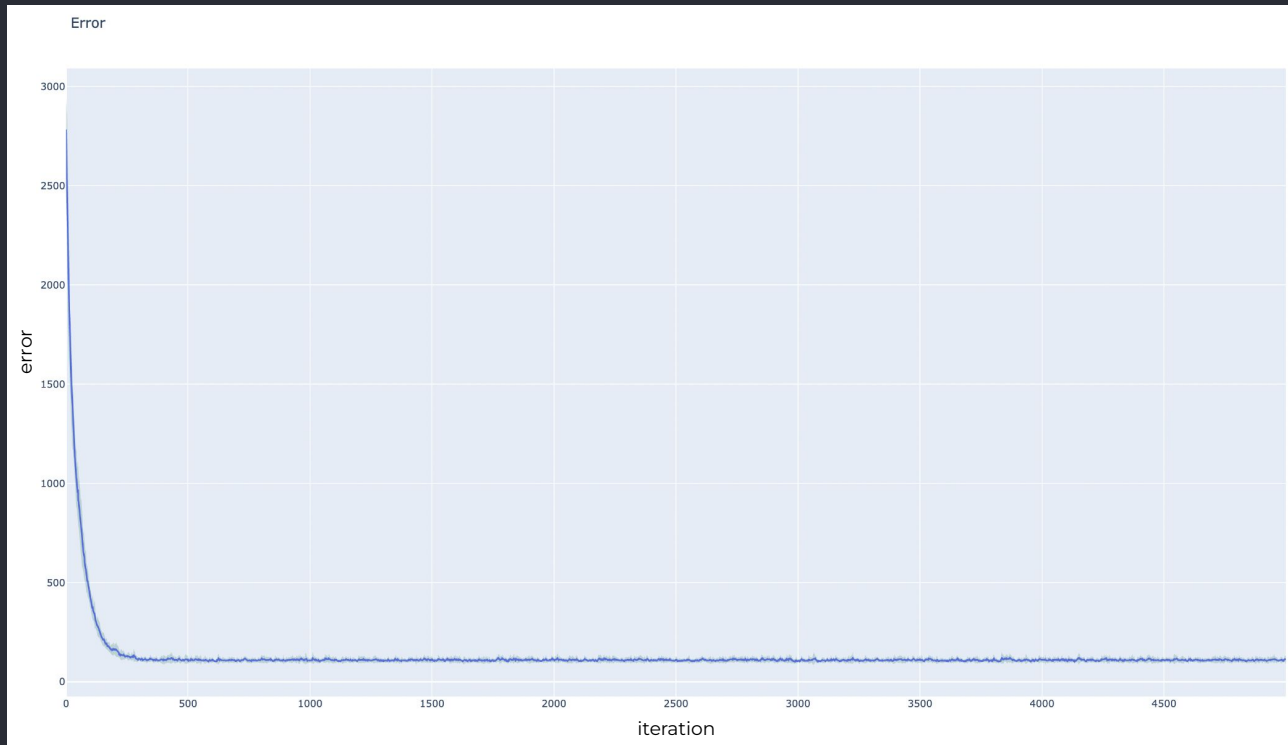
Consideraciones:

- Dado que los valores de salida del problema están por fuera del intervalo $[-1, 1]$, se decidió normalizar dichos valores para el caso del perceptrón no lineal.
- Para normalizar los valores se utilizó la función:

$$2 * (Y - \min(Y)) / (\max(Y) - \min(Y)) - 1$$

Perceptron Lineal

Error



Perceptron Lineal

Error - Conclusiones

- El perceptrón lineal no consigue ajustar de manera correcta los valores del problema, lo que permite concluir que no se trata de una transformación lineal.

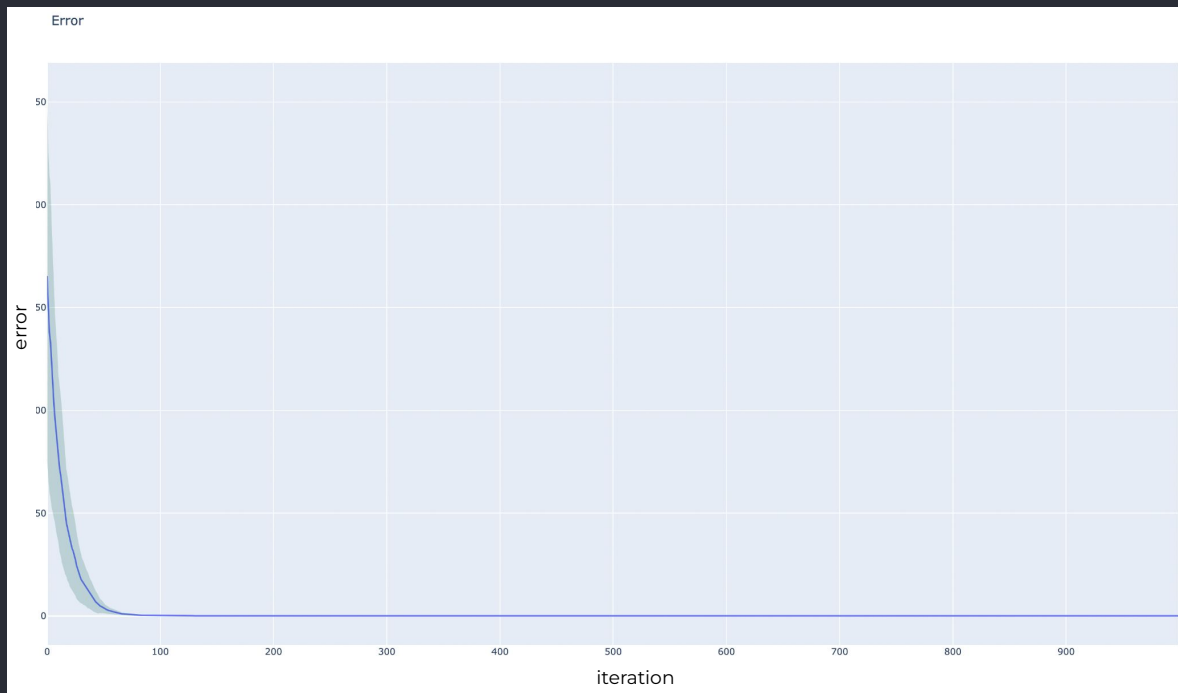
Perceptron Lineal

Transformación Lineal

- Además de los datos brindados por la cátedra, se generó un archivo de entrada y otro de salida a partir de la función $Y(x) = 2 \cdot x$.
- Veamos que esta función corresponde a una transformación lineal:
 - ✓ Sean u, v pertenecientes a los reales $F(u+v) = 2 \cdot (u+v) = 2 \cdot u + 2 \cdot v = F(u) + F(v)$.
 - ✓ Sean u, k pertenecientes a los reales $F(k \cdot u) = 2 \cdot k \cdot u = k \cdot 2 \cdot u = k \cdot F(u)$.
 - ✓ $F(0) = 2 \cdot 0 = 0$.
- Veamos ahora cómo ajusta estos valores el perceptrón lineal.

Perceptron Lineal

Transformación Lineal - Error



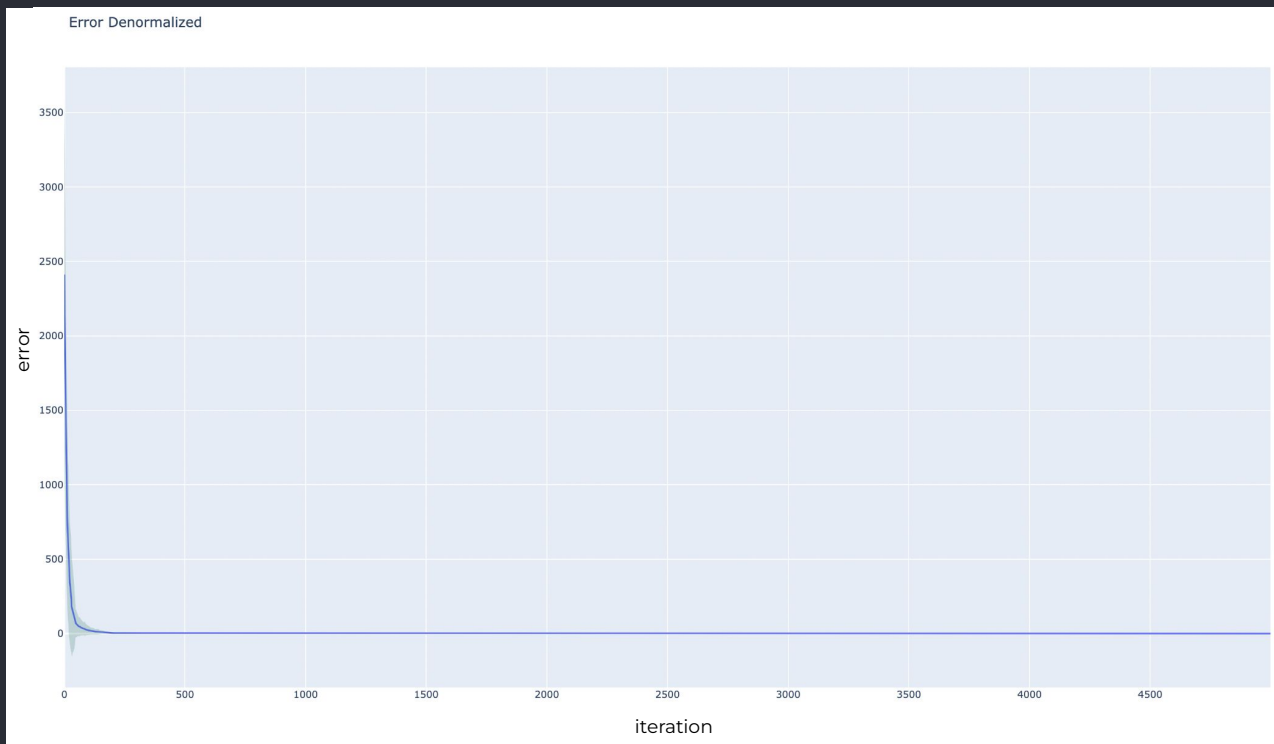
Perceptron Lineal

Transformación Lineal - Error
Conclusiones

- Se puede concluir que el perceptrón lineal es capaz de aproximar correctamente los valores para una transformación lineal.

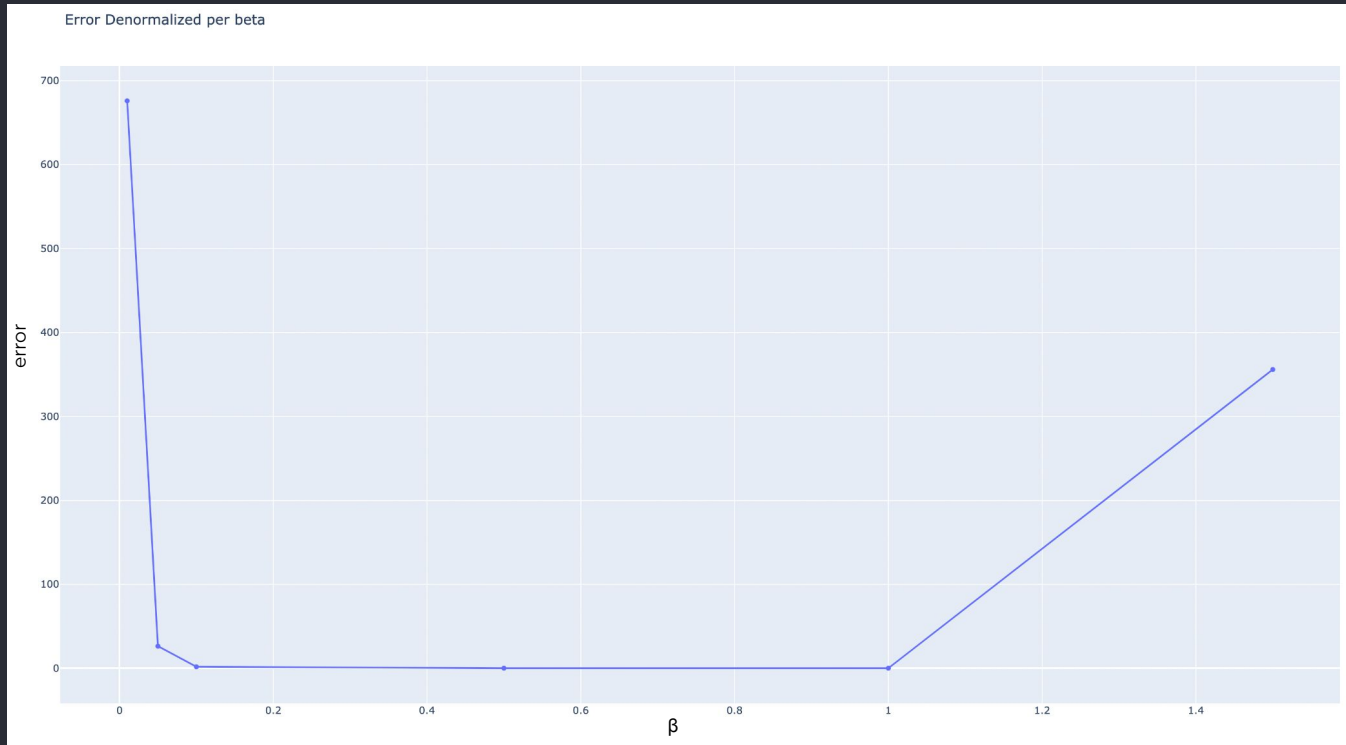
Perceptron No Lineal

Error



Perceptron No Lineal

Error valor de β - $\eta=0.1$



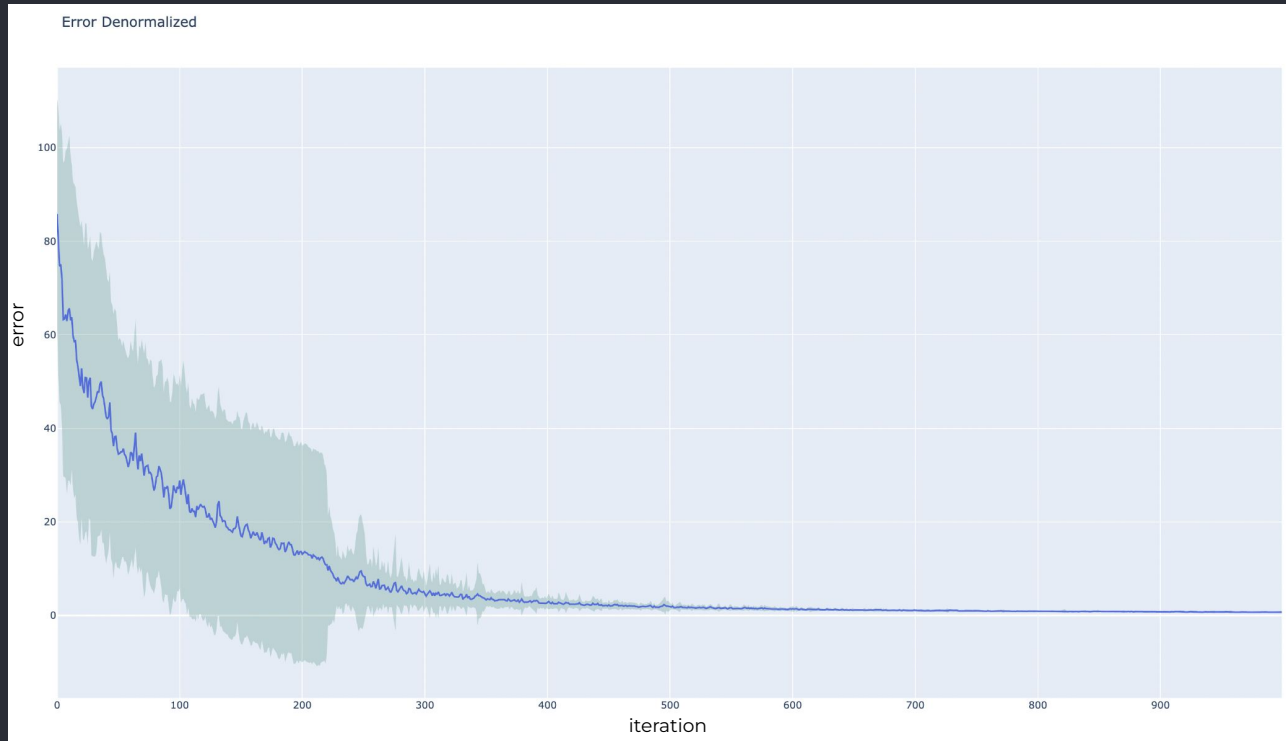
Perceptron No Lineal

Error - Conclusiones

- El perceptrón no lineal consigue ajustar de manera correcta los valores del problema, lo que permite concluir que se trata de una transformación no lineal.
- Los mejores valores de beta para la función sigmoide en este problema, con una tasa de aprendizaje fija de 0.1, están en el rango $[0.1, 1]$

Perceptron No Lineal

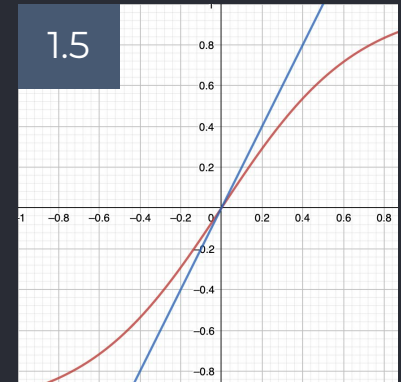
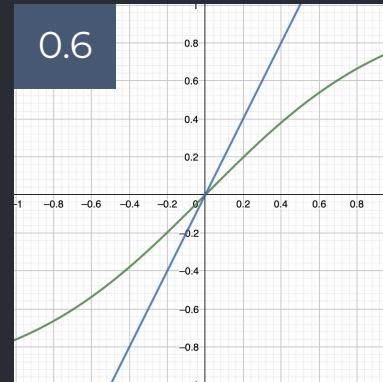
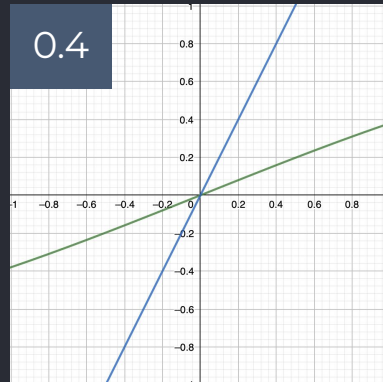
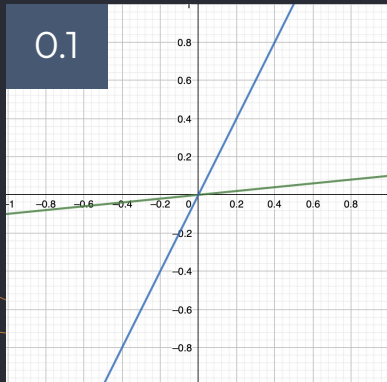
Transformación Lineal - Error



Perceptron No Lineal

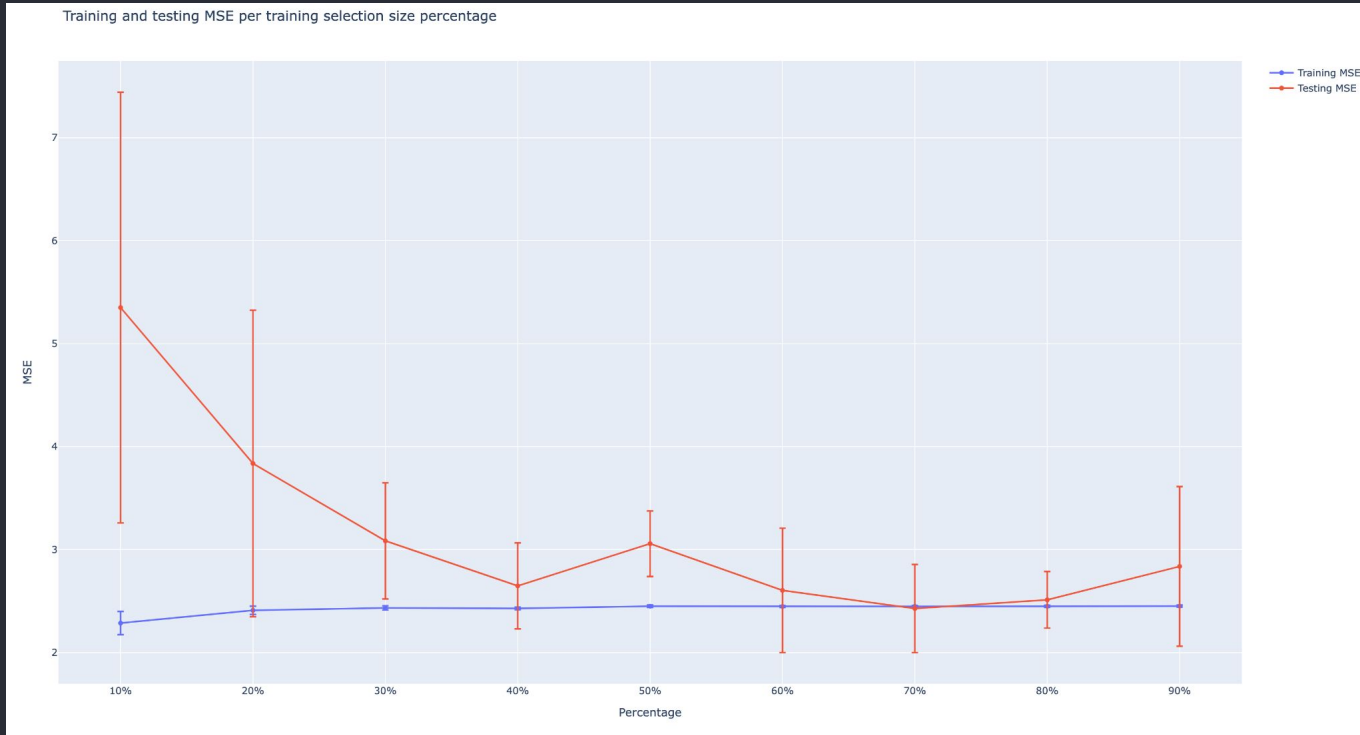
Transformación Lineal - Error Conclusiones

- Se puede concluir que el perceptrón no lineal es capaz de aproximar correctamente los valores para una transformación lineal, pero no mejor que un perceptrón lineal.
- Esto lo logra mediante valores de beta adecuados que permitan que la función sigmoide logre el mejor ajuste posible respecto a la función original.



Perceptron No Lineal

Porcentajes de entrenamiento y prueba - 10 corridas



Perceptron No Lineal

Porcentajes de entrenamiento y prueba
CONCLUSIONES

- Los mejores porcentajes para entrenar y testear son 80% y 20% respectivamente.
- Con porcentajes muy bajos en el conjunto de entrenamiento se produce sobreajuste ya que se cuenta con pocos registros.
- Con porcentajes muy altos en el conjunto de entrenamiento (90%) también se produce un pequeño sobreajuste, el cual puede deberse a sobreentrenar la red con muchos datos para los cuales se conoce el resultado deseado.

Perceptron No Lineal

Capacidad de generalización
métrica error
(validación cruzada k=10)

| Group | Training MSE | Training Values STD | Testing MSE | Testing Input Values STD |
|-------|--------------|---------------------|-------------|--------------------------|
| 1 | 0.24591 | 2.15895 | 0.23869 | 2.15473 |
| 2 | 0.24276 | 2.18622 | 0.22768 | 1.87865 |
| 3 | 0.24556 | 2.16820 | 0.23503 | 2.04103 |
| 4 | 0.24238 | 2.12127 | 0.35451 | 2.44638 |
| 5 | 0.24607 | 2.15863 | 0.24418 | 2.04963 |
| 6 | 0.24534 | 2.16313 | 0.24186 | 2.10047 |
| 7 | 0.24523 | 2.17407 | 0.41958 | 1.82110 |
| 8 | 0.24413 | 2.17404 | 0.22385 | 2.00367 |
| 9 | 0.24572 | 2.14352 | 0.28848 | 2.24238 |
| 10 | 0.24611 | 2.13987 | 0.25092 | 2.22928 |

- Considerando como métrica el error se puede concluir que el perceptrón no lineal tiene una buena capacidad de generalización. Es decir, permite predecir de manera correcta valores fuera del conjunto de entrenamiento.
- Para este caso, se puede decir que el mejor conjunto de entrenamiento es el 8, ya que es el que mejor predice los valores de testeo.
- Además, puede observarse que el conjunto 8 es de los más representativo, ya que tiene una mayor desviación estándar respecto al resto de los conjuntos de entrenamiento.

PERCEPTRON MULTICAPA



Ejercicio 3

- El perceptrón multicapa consiste de un conjunto de perceptrones como los utilizados en los ejercicios anteriores pero, que en este caso, trabajan en conjunto, divididos en distintas capas.
- Las capas intermedias de la red son llamadas capas ocultas.
- Se utiliza la técnica de retro propagar la diferencia entre la predicción y el resultado de cada unidad de la capa de salida hacia las capas inferiores.
- Para actualizar los pesos, hay dos opciones, de manera incremental o por lotes.

Consideraciones:

- Se actualizan los pesos de manera incremental.

Ejercicio 3.1

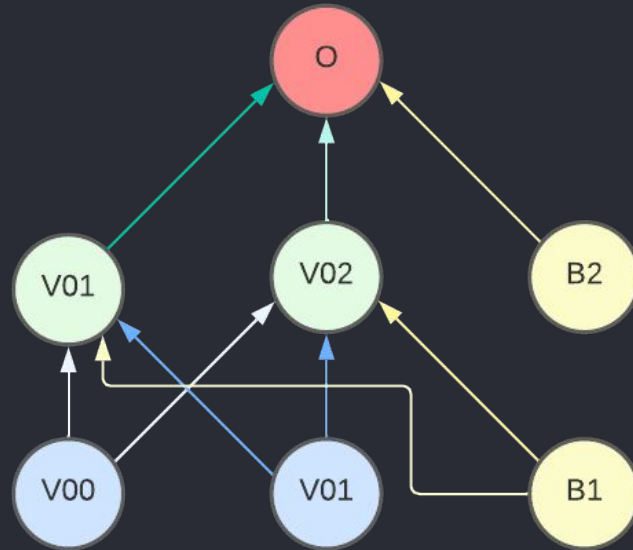
Función lógica 'O exclusivo'

$$x = \{\{-1, 1\}, \{1, -1\}, \{-1, -1\}, \{1, 1\}\}$$

$$y = \{1, 1, -1, -1\}$$

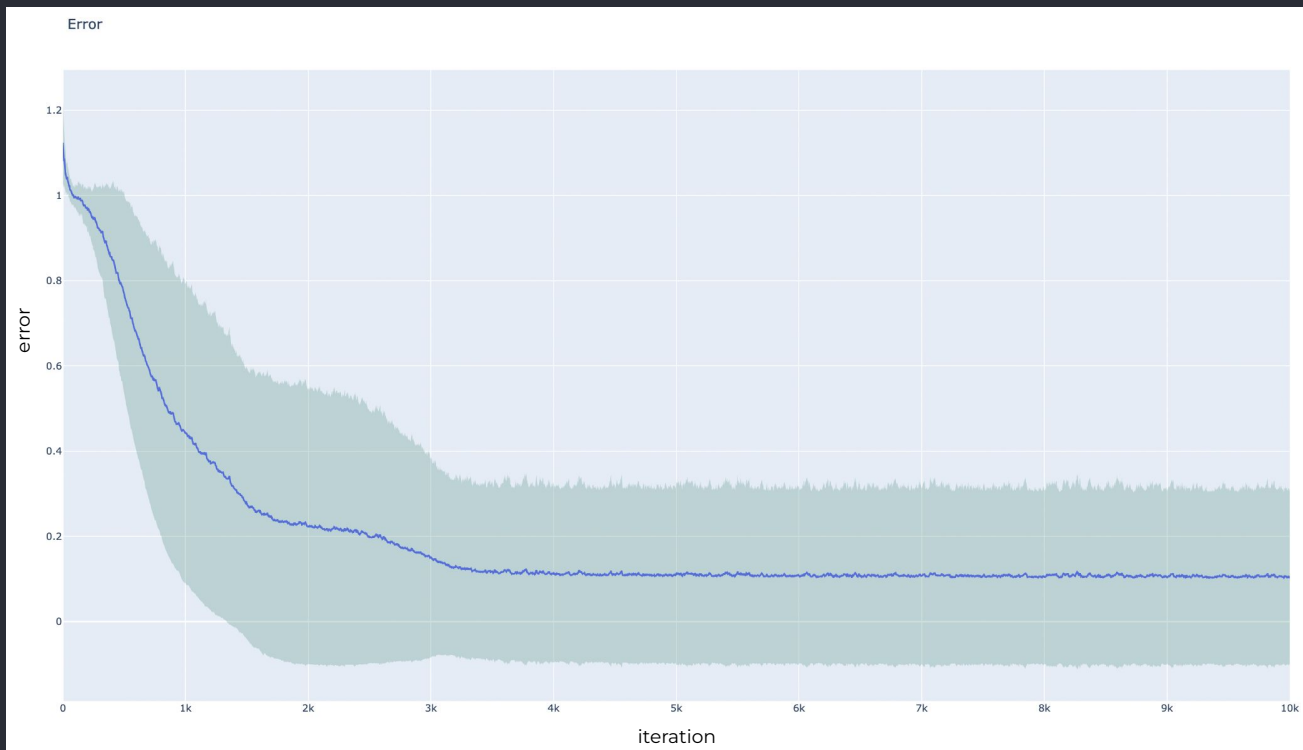
Ejercicio 3.1

Configuración



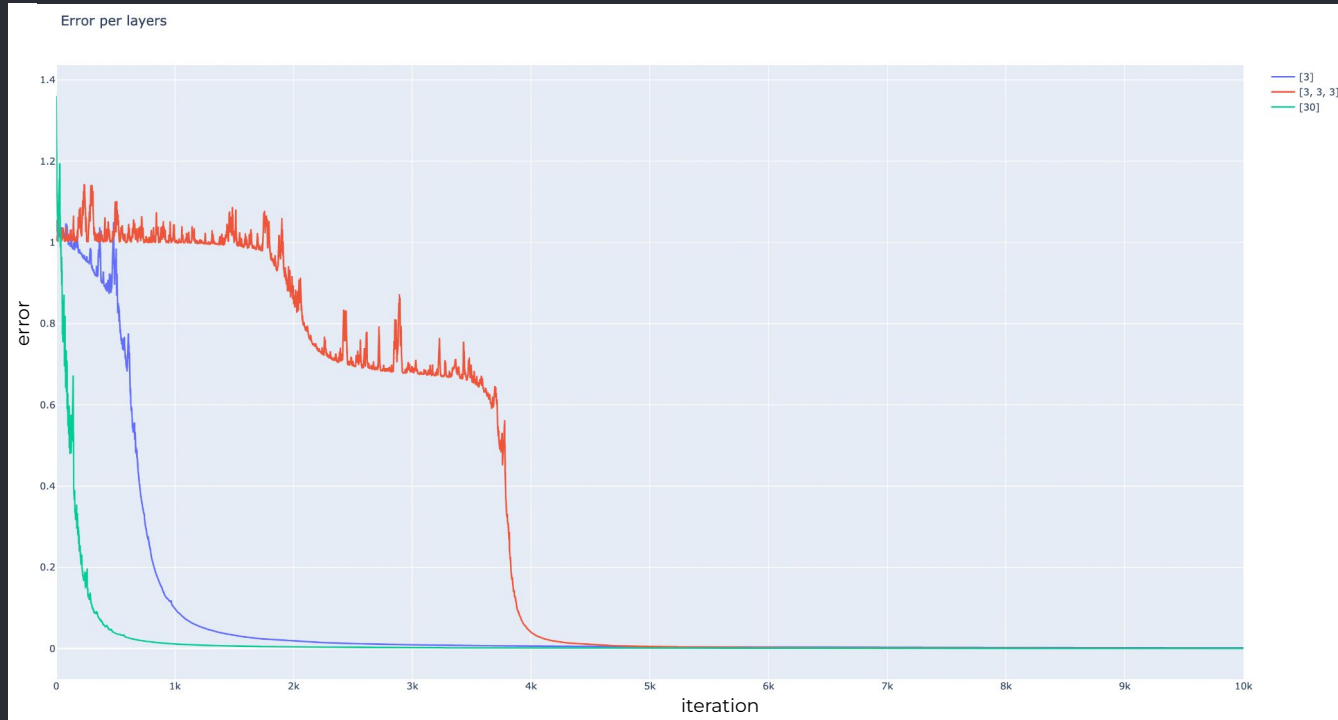
Ejercicio 3.1

Error



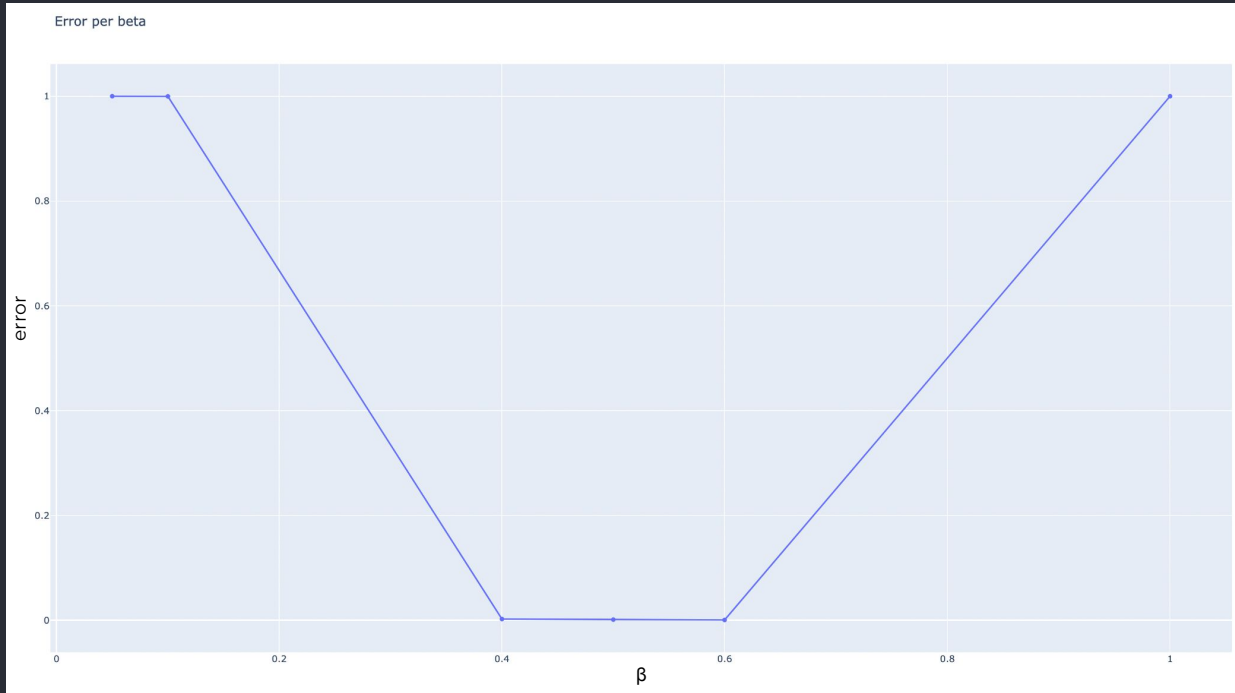
Ejercicio 3.1

Error - Cantidad de capas y unidades



Ejercicio 3.1

Error valor de β - $\eta=0.1$ - capas [30]



Ejercicio 3.1

Error - Conclusiones

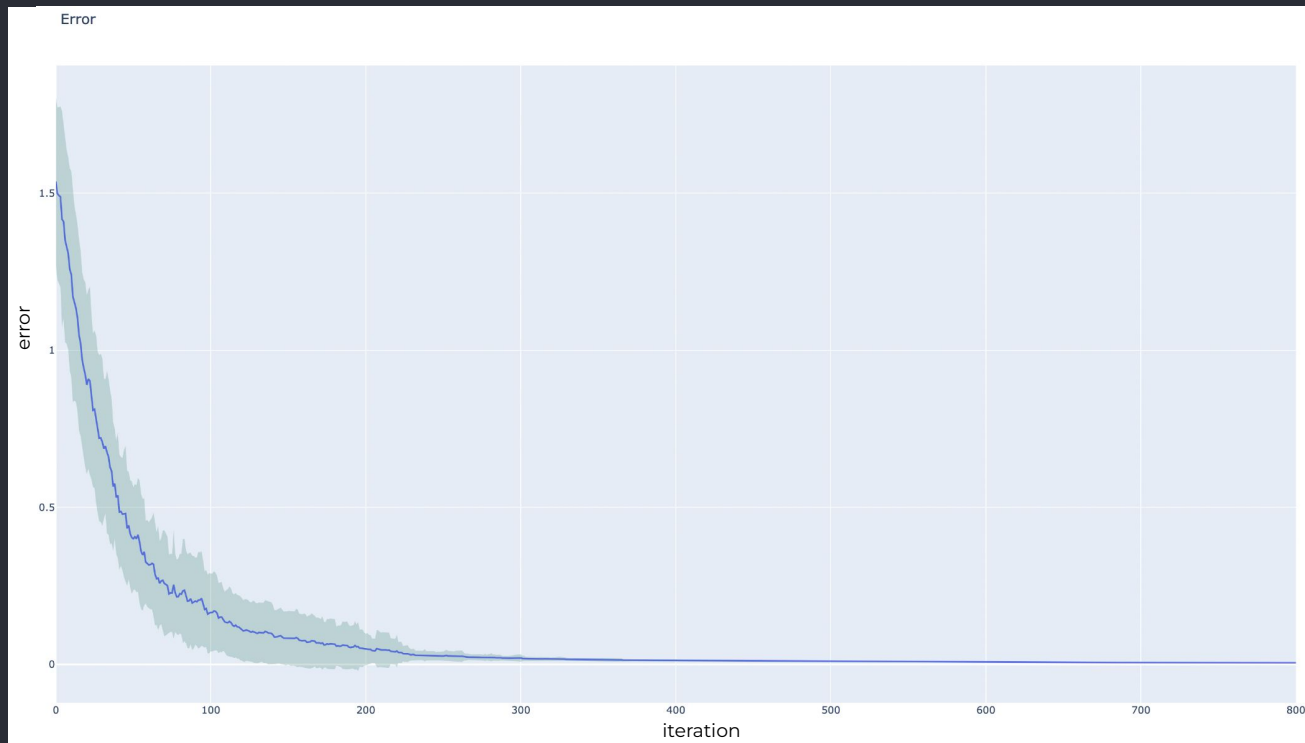
- El perceptrón multicapa es capaz de resolver el problema del XOR que el perceptrón simple escalón no consigue resolver.
- La capa de entrada es capaz de transformar los puntos de forma tal que se consiga una transformación que el perceptrón simple de la capa oculta pueda resolver (lineal/no lineal).
- Para este problema es mejor aumentar la cantidad de unidades dentro de una capa a simplemente aumentar la cantidad de capas manteniendo la cantidad de unidades.
- Los mejores valores de beta para la función sigmoide en este problema, con una tasa de aprendizaje fija de 0.1, están en el rango $[0.4, 0.6]$.

Ejercicio 3.2

Discriminar si un número es par, con entradas dadas por el conjunto de números del 0 al 9 representados por imágenes de 5 x 7 píxeles.

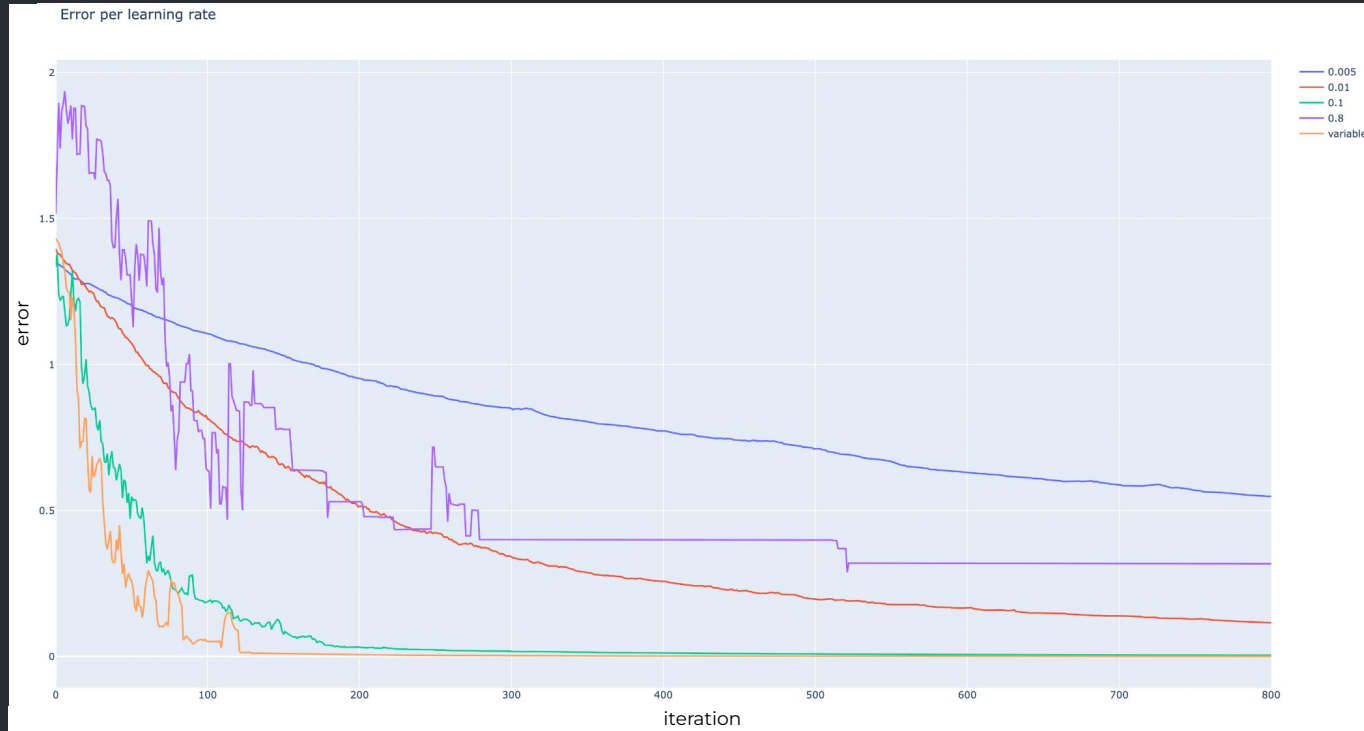
Ejercicio 3.2

Error



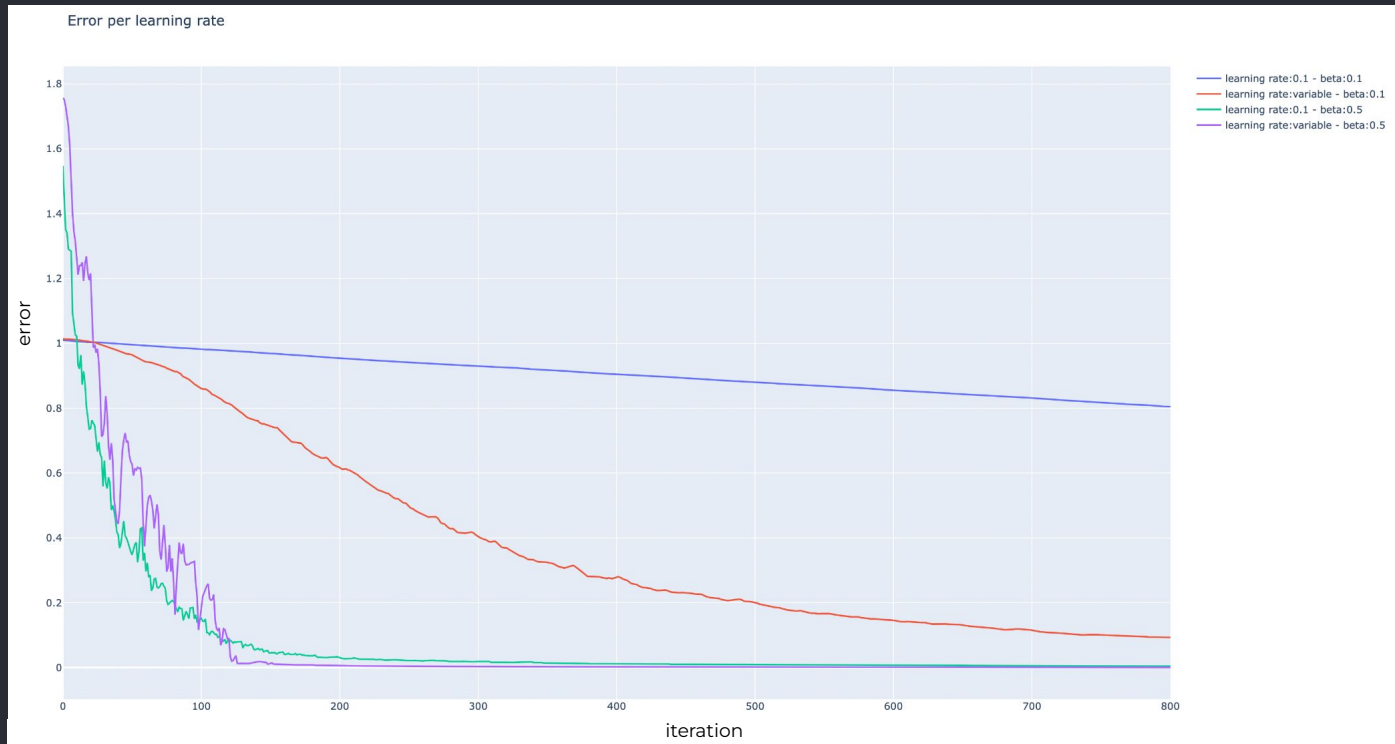
Ejercicio 3.2

Error - Tasa de aprendizaje



Ejercicio 3.2

Error - valor de β



Ejercicio 3.2

Error - Conclusiones

- El perceptrón multicapa es capaz de clasificar correctamente los elementos del conjunto de entrenamiento en par e impar.
- Para tasas de aprendizajes muy altas o muy bajas no se llega a la solución en una cantidad fija de iteraciones.
- El método variable permite llegar a la solución de manera mas rapida.
- Para valores de beta pequeños, sólo el método variable consigue llegar a la solución. Esto se debe a que utilizando momentum y una tasa de aprendizaje variable, consigue acercarse a la solución rápidamente.
- Para valores de beta grande, ambos consiguen llegar a la solución gracias a que la derivada de la sigmoide es mayor, aumentando delta w en cada iteración.

Ejercicio 3.2

Capacidad de generalización
métrica error
(validación cruzada $k=10$)

| Group | Training MSE | Testing MSE | Testing Guess |
|-------|--------------|-------------|---------------|
| 1 | 0.00133 | 0.00042 | 1.00000 |
| 2 | 0.23230 | 0.04728 | 1.00000 |
| 3 | 0.00486 | 3.76863 | 0.00000 |
| 4 | 0.00221 | 3.87293 | 0.00000 |
| 5 | 0.00210 | 0.00700 | 1.00000 |
| 6 | 0.00301 | 1.61793 | 0.00000 |
| 7 | 0.00321 | 0.00479 | 1.00000 |
| 8 | 0.00281 | 0.00173 | 1.00000 |
| 9 | 0.00173 | 0.00029 | 1.00000 |
| 10 | 0.00199 | 3.58648 | 0.00000 |
| STD | 0.06892 | 1.67406 | 0.48990 |
| MEAN | 0.02950 | 1.32559 | 0.58999 |

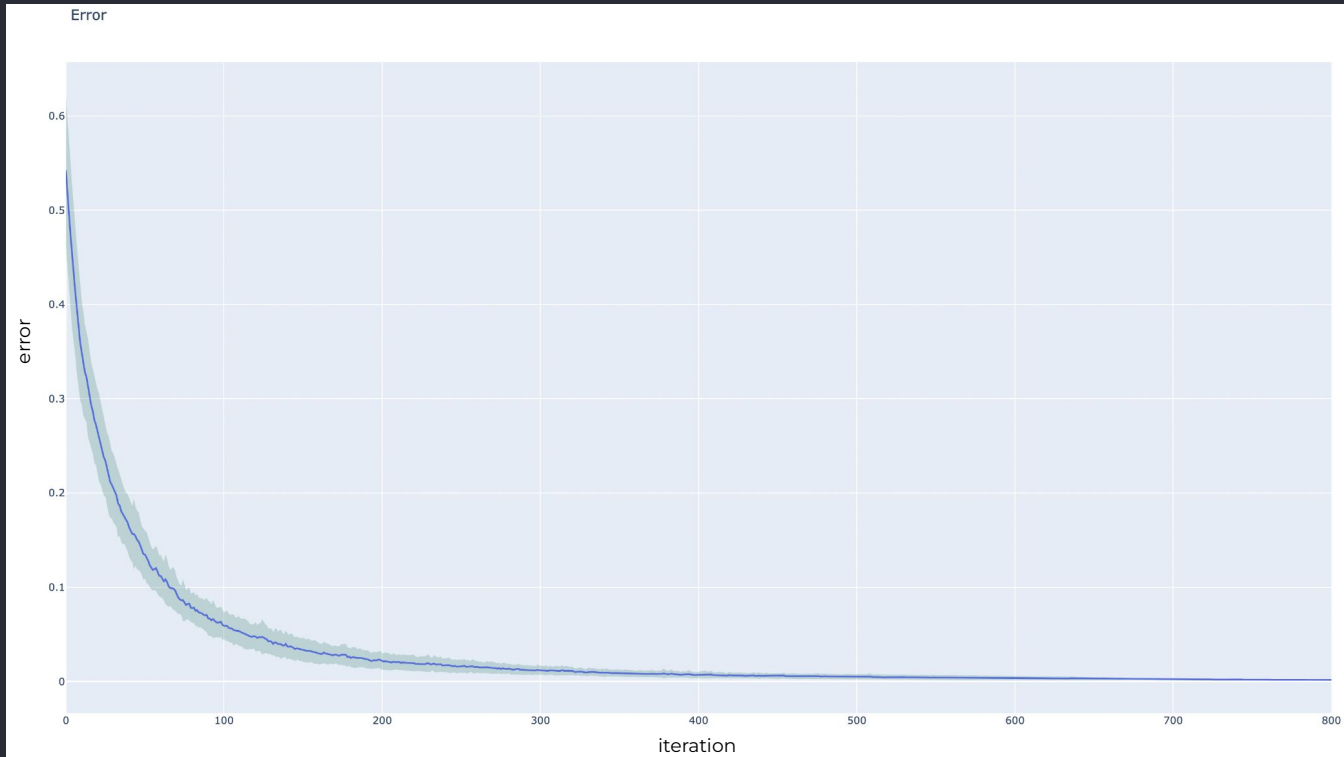
- Para este caso puede observarse que el mejor conjunto de entrenamiento corresponde al 8, ya que es el que menor error brinda en el conjunto de testeo.
- No obstante, a partir de este ejemplo y tomando el promedio de otras 15 corridas, el valor medio correspondiente a acertar el resultado del valor predicho se acerca a 0.5. Esto nos quiere decir que tiene una mala capacidad de generalización respecto al error, ya que predice el resultado de manera correcta con una probabilidad de 0.5. Esto tiene sentido, ya que la probabilidad de acertar si un valor es par o impar ciegamente es la misma que la de tirar una moneda.

Ejercicio 3.3

Discriminar con 10 unidades de salida de modo que cada salida represente a un dígito, las entradas dadas por el conjunto de números del 0 al 9 representados por imágenes de 5 x 7 píxeles.

Ejercicio 3.3

Error



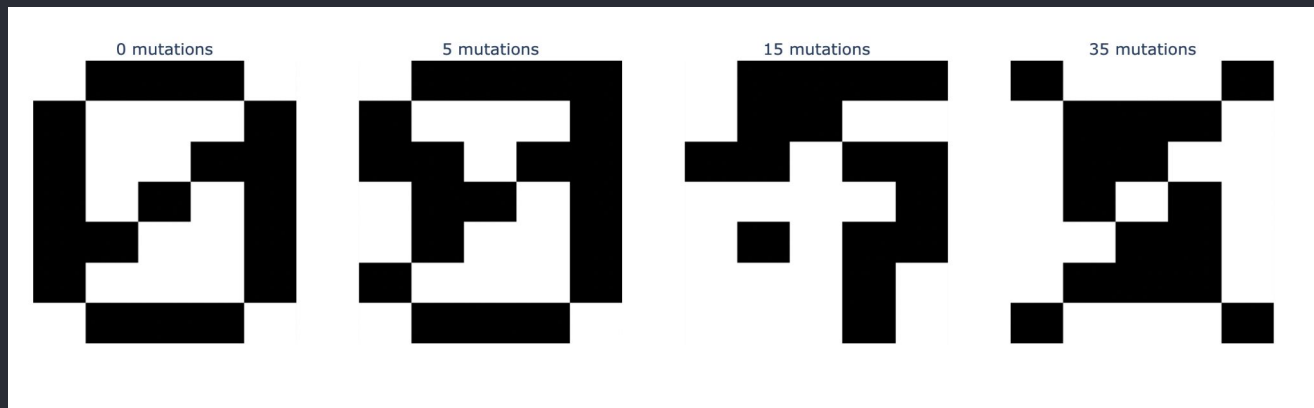
Ejercicio 3.3

Error - Conclusiones

- El perceptrón multicapa es capaz de clasificar correctamente los elementos del conjunto de entrenamiento en clases como lo pide el enunciado.

Ejercicio 3.3

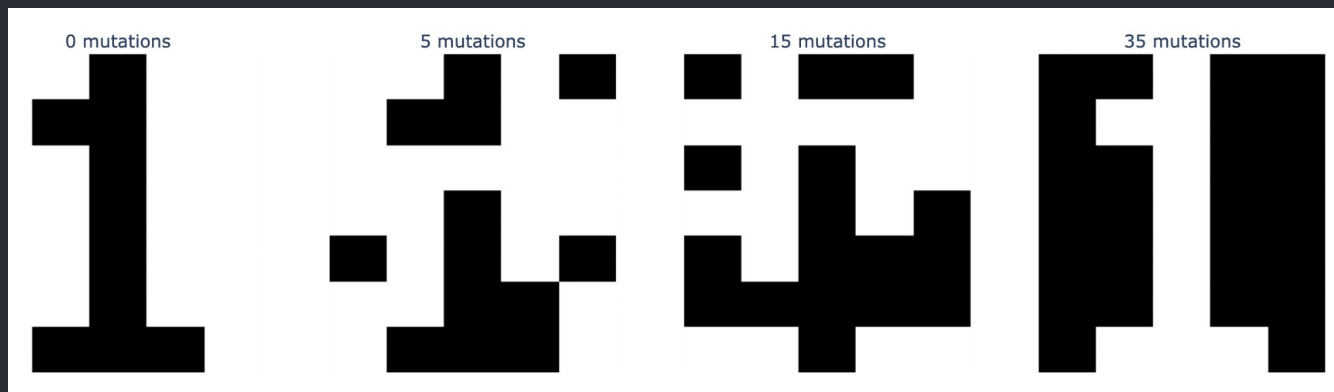
Mutación de bits



| Mutations | Prediction Error |
|-----------|------------------|
| 0 | 0.000245 |
| 5 | 0.133533 |
| 15 | 0.391851 |
| 35 | 0.480777 |

Ejercicio 3.3

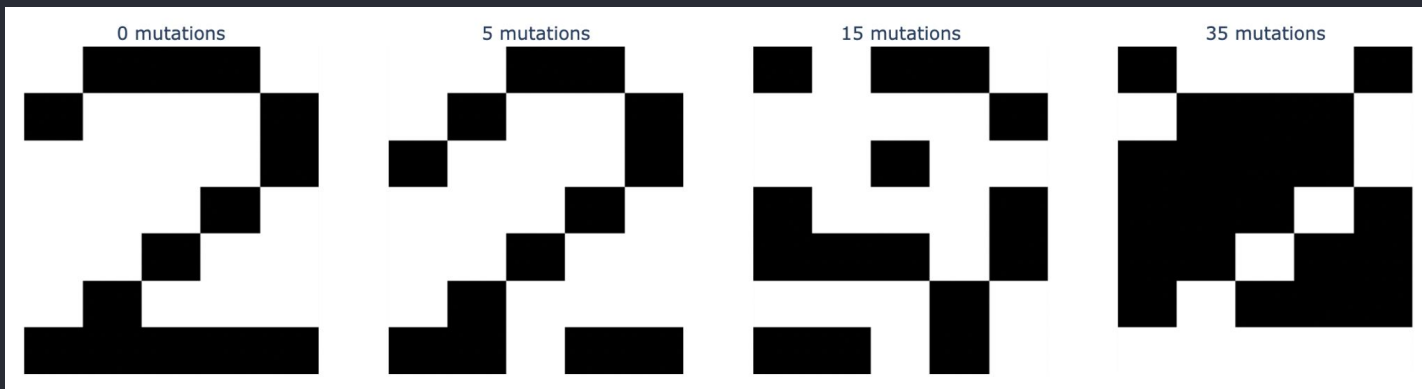
Mutación de bits



| Mutations | Prediction Error |
|-----------|------------------|
| 0 | 0.000259 |
| 5 | 0.245242 |
| 15 | 0.164090 |
| 35 | 0.611463 |

Ejercicio 3.3

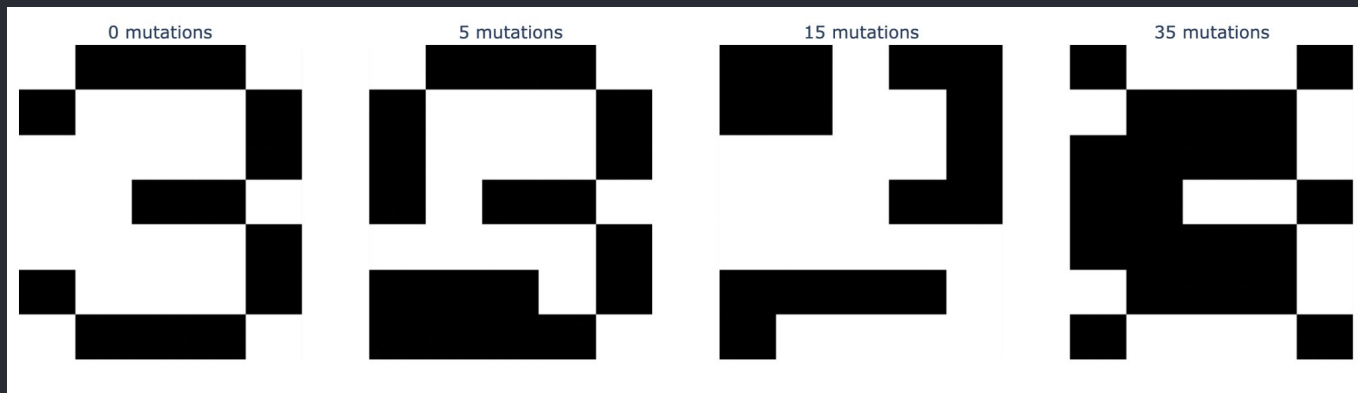
Mutación de bits



| Mutations | Prediction Error |
|-----------|------------------|
| 0 | 0.000531 |
| 5 | 0.210746 |
| 15 | 0.182518 |
| 35 | 0.838397 |

Ejercicio 3.3

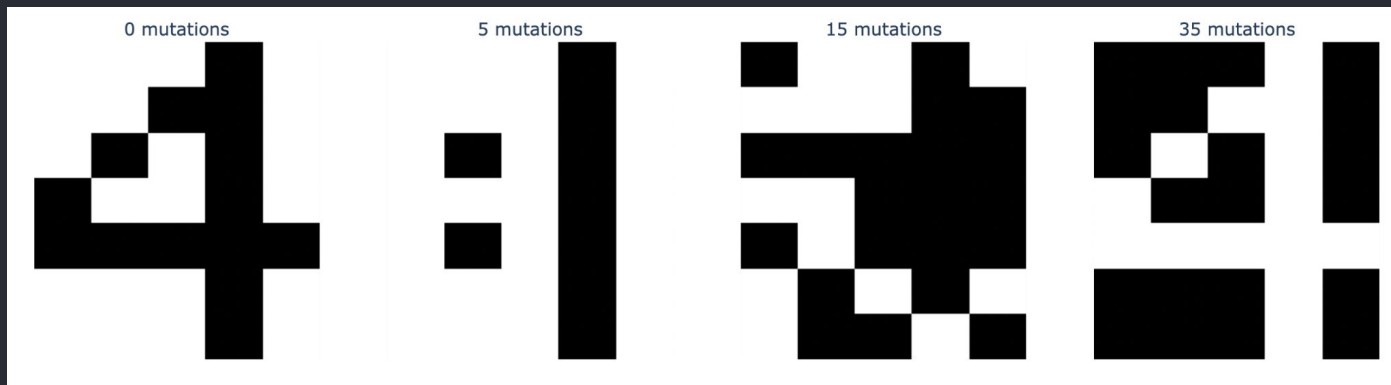
Mutación de bits



| Mutations | Prediction Error |
|-----------|------------------|
| 0 | 0.000746 |
| 5 | 0.409739 |
| 15 | 0.343855 |
| 35 | 0.730953 |

Ejercicio 3.3

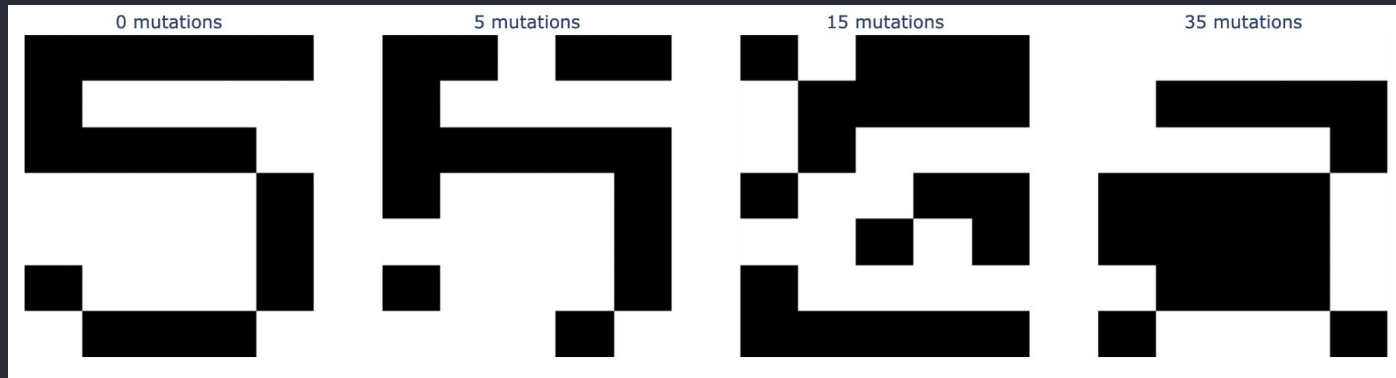
Mutación de bits



| Mutations | Prediction Error |
|-----------|------------------|
| 0 | 0.000181 |
| 5 | 0.070398 |
| 15 | 0.311944 |
| 35 | 0.308951 |

Ejercicio 3.3

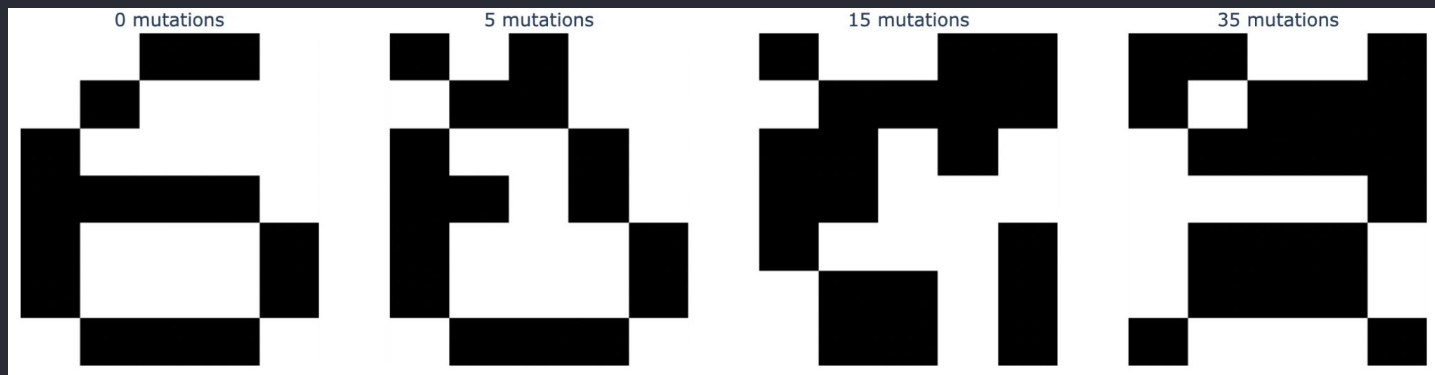
Mutación de bits



| Mutations | Prediction Error |
|-----------|------------------|
| 0 | 0.000254 |
| 5 | 0.379880 |
| 15 | 0.219964 |
| 35 | 0.610292 |

Ejercicio 3.3

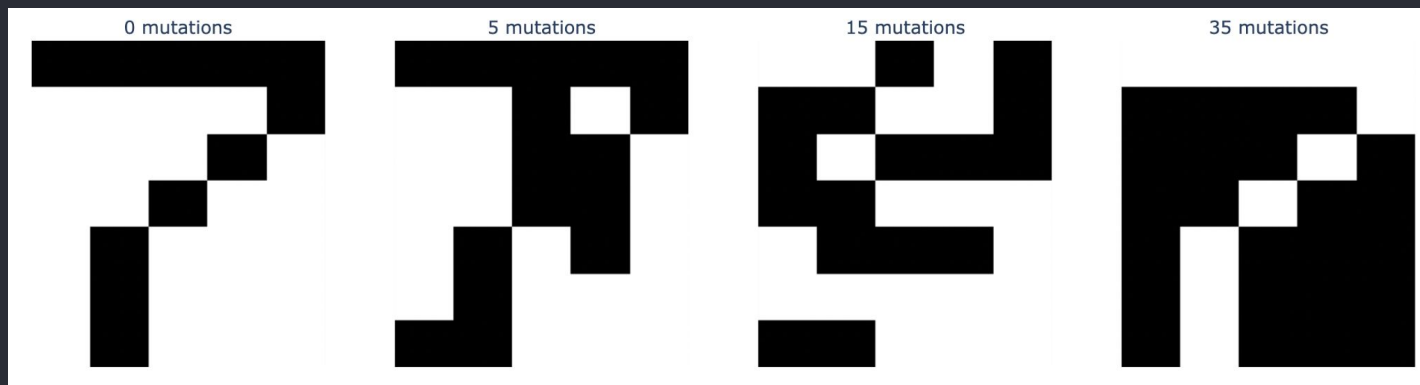
Mutación de bits



| Mutations | Prediction Error |
|-----------|------------------|
| 0 | 0.000567 |
| 5 | 0.117891 |
| 15 | 0.416034 |
| 35 | 0.733318 |

Ejercicio 3.3

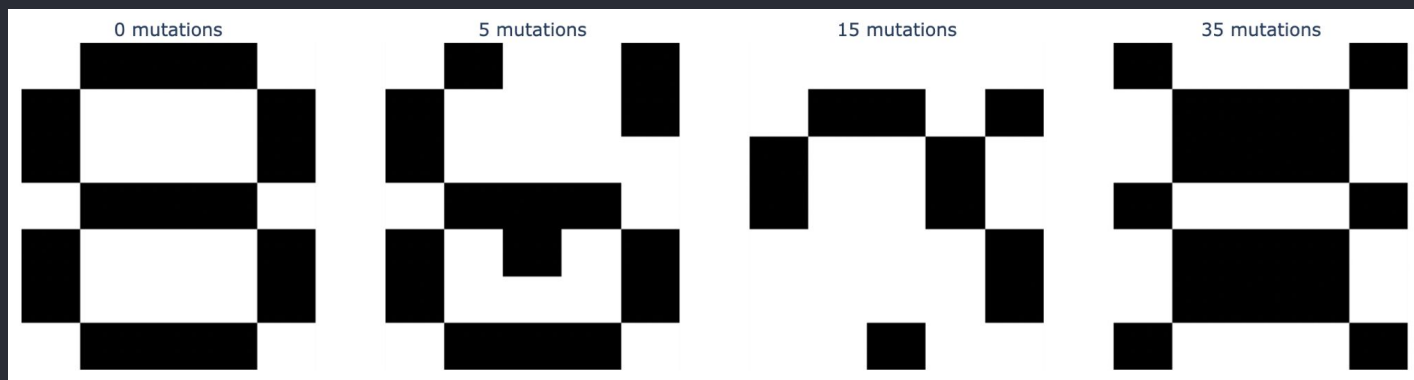
Mutación de bits



| Mutations | Prediction Error |
|-----------|------------------|
| 0 | 0.000221 |
| 5 | 0.142408 |
| 15 | 0.351682 |
| 35 | 0.657000 |

Ejercicio 3.3

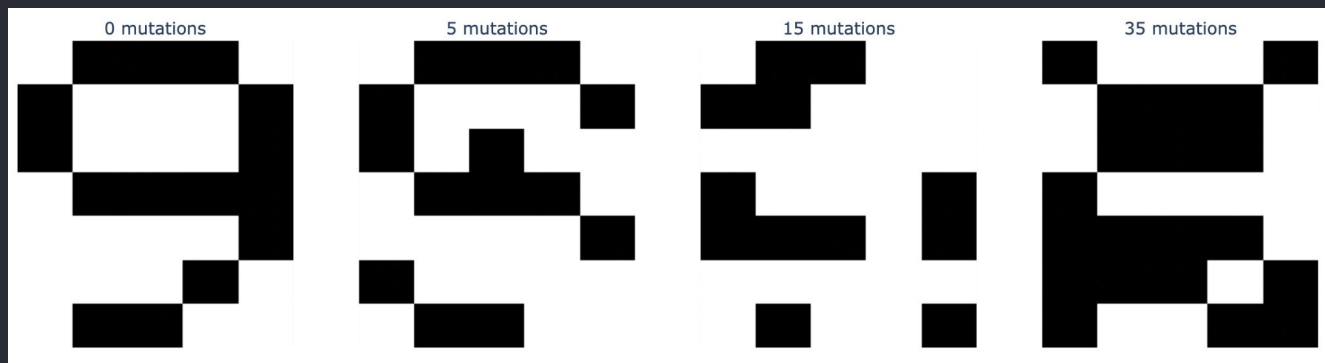
Mutación de bits



| Mutations | Prediction Error |
|-----------|------------------|
| 0 | 0.001802 |
| 5 | 0.146214 |
| 15 | 0.372302 |
| 35 | 0.384244 |

Ejercicio 3.3

Mutación de bits



| Mutations | Prediction Error |
|-----------|------------------|
| 0 | 0.000180 |
| 5 | 0.481885 |
| 15 | 0.179824 |
| 35 | 0.742323 |

Ejercicio 3.3

Mutación de bits - Conclusiones

- A medida que aumenta la cantidad de mutaciones en la imagen, mayor es el error de predicción.
- Para una cantidad de mutación baja, puede observarse que el error sigue siendo aceptable.
- Esto puede relacionarse con los QR en papel, que se busca que puedan seguir siendo escaneados a pesar de estar levemente dañados.



MUCHAS GRACIAS
¿Preguntas?