

Práctica 3. Redes Neuronales

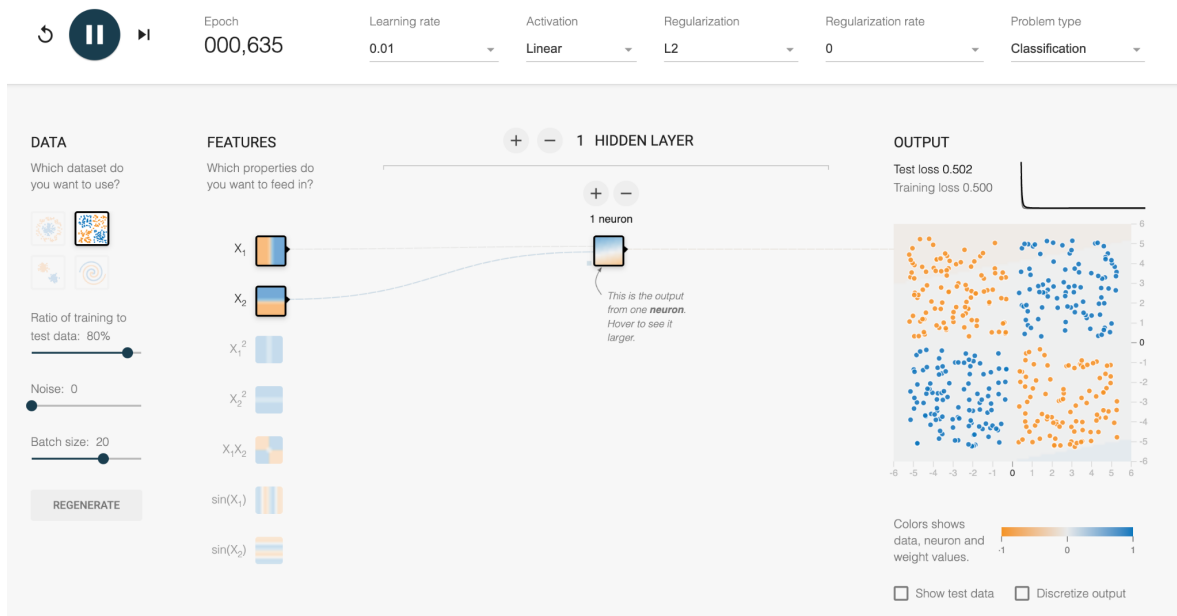
Utilizando el DEMO (<https://playground.tensorflow.org/>) que vimos en la clase pasada, realicen los siguientes puntos, contesten las preguntas, justifiquen su respuesta y agreguen evidencia de las redes neuronales (imagen).

Ejercicio 1. Lineales

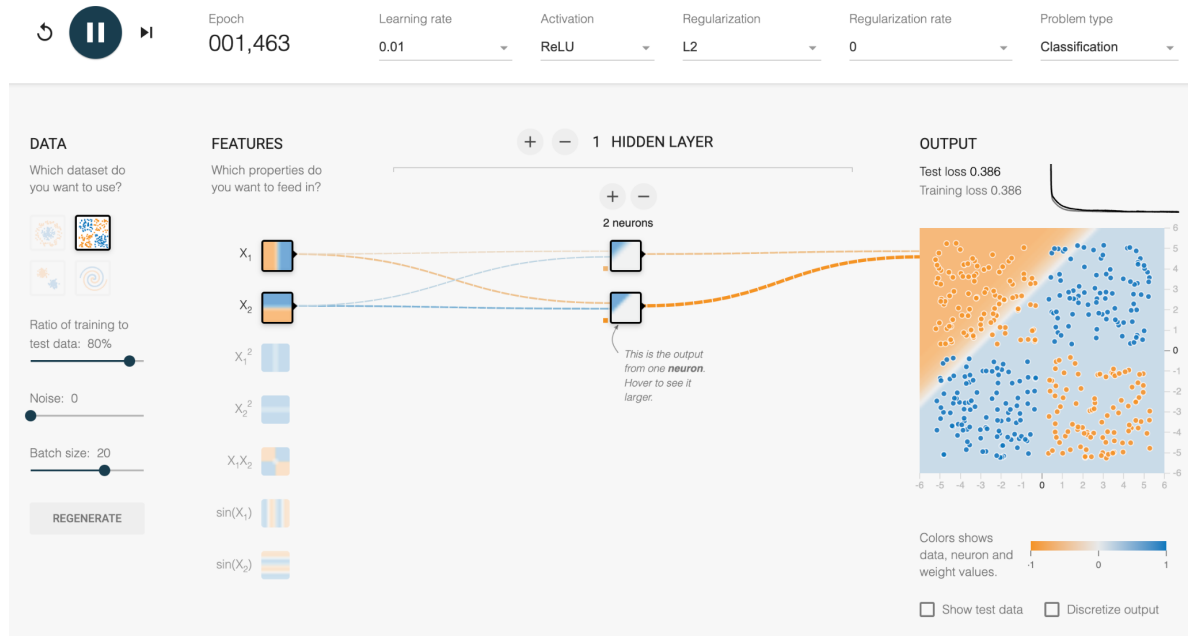
1. El modelo de la Figura 1 combina dos características de entrada en una sola neurona.

- **¿Aprenderá este modelo alguna no linealidad?** Ejecútenlo para confirmar su suposición.

R: No, ya que al tener una función de activación lineal, esta no aprenderá alguna no linealidad.



2. Prueben aumentar el número de neuronas en la capa oculta de 1 a 2, y prueben también cambiar de una activación lineal a una activación no lineal como ReLU.



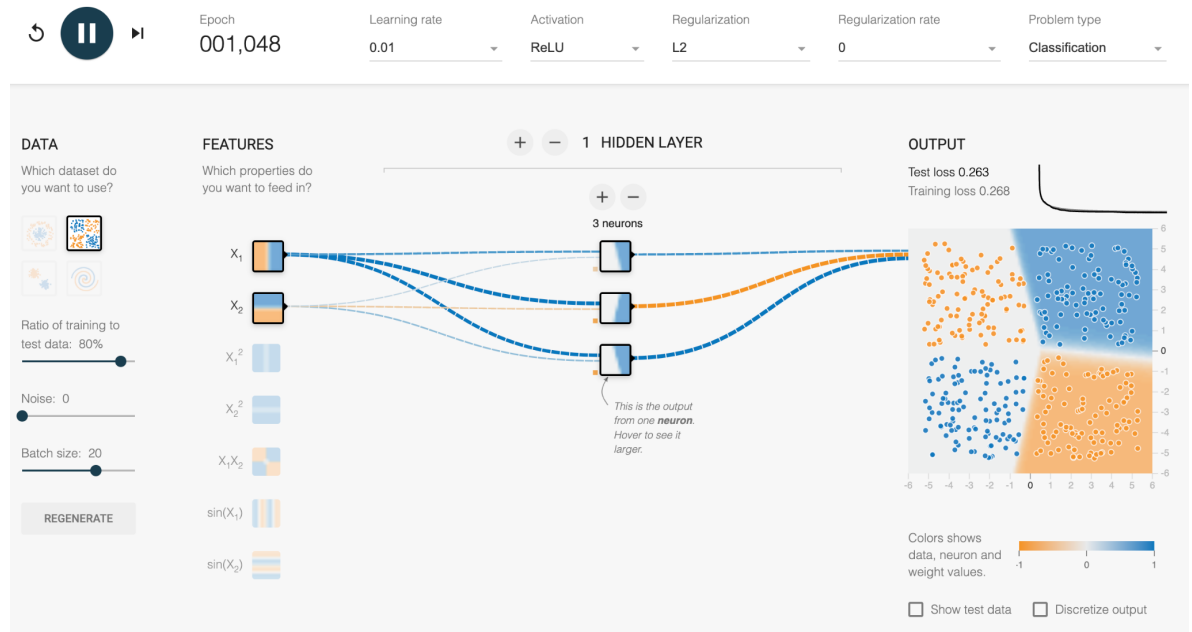
- ***¿Se puede crear un modelo que aprenda no linealidades?***

R: Si, uno de los requisitos de esto, es que la función de activación sea no lineal, como en el punto número 2 que la función de activación es ReLu.

- ***¿Puede modelar los datos con eficacia?***

R: Si, ya que la mayoría de datos existentes en el mundo son no lineales, lo cual hace que un modelo de este tipo sea de los mejores para poder trabajar.

3. Intenten aumentar el número de neuronas de la capa oculta de 2 a 3, utilizando una activación no lineal como ReLU.



- ¿Puede modelar los datos eficazmente?

R: Sí, debido a que los modelos, suelen ser no lineales, por lo que ésta aprende de la mejor forma utilizando ReLU, ayudándonos justamente a que modele de la mejor forma posible.

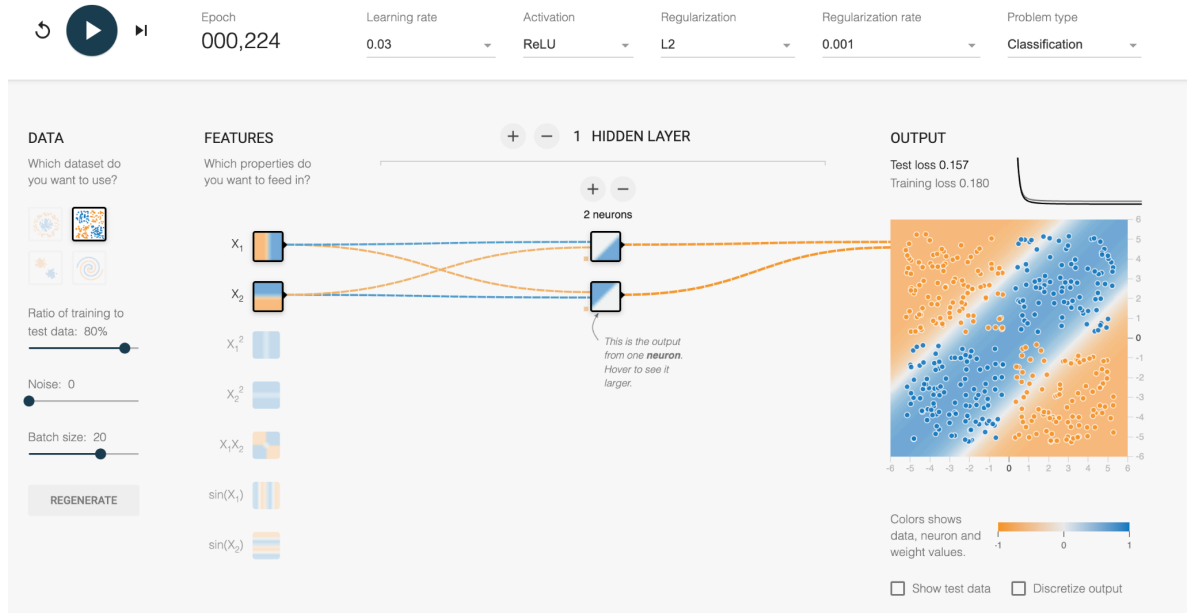
- ¿Cómo varía la calidad del modelo de una ejecución a otra?

R: Parece que entre más neuronas, aprende de mejor forma, pero hay que saber nivelarlo para la cantidad de datos que se tienen, el sesgo aumenta la flexibilidad del modelo.

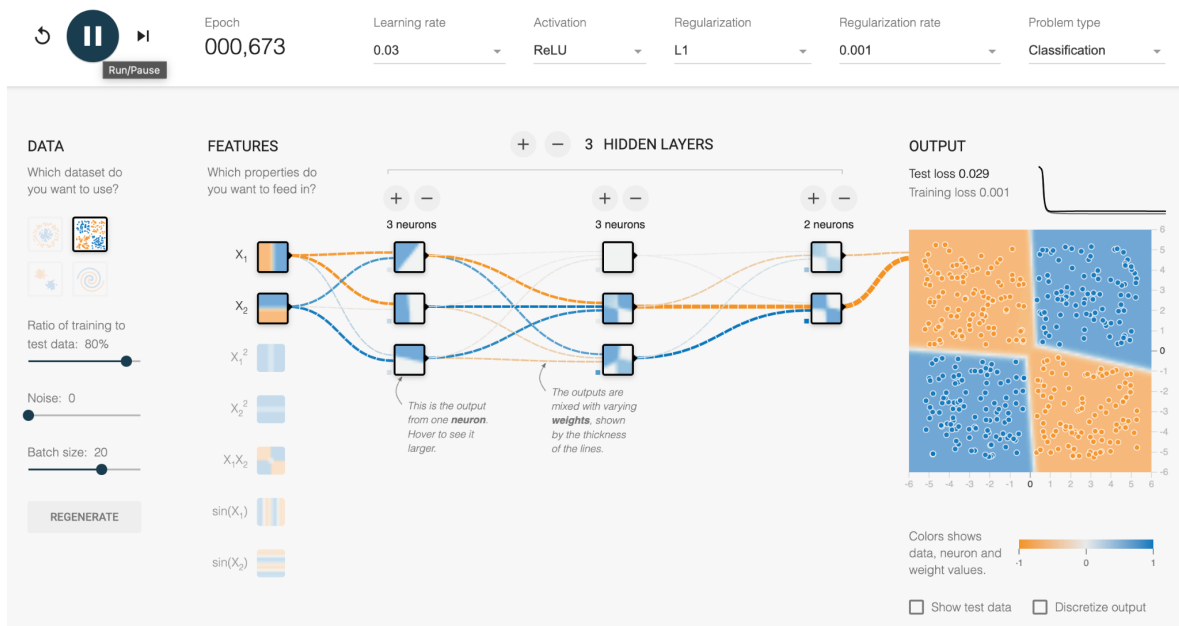
- Continúen experimentando, añadiendo o eliminando capas ocultas y neuronas por capa. También pueden cambiar los ritmos de aprendizaje, la regularización y otros parámetros de aprendizaje.

- ¿Cuál es el menor número de neuronas y capas que puedes utilizar para obtener una pérdida en la prueba de 0,177 o inferior?

R: 1 capa y 2 neuronas



- ¿El aumento del tamaño del modelo mejora el ajuste o la rapidez de convergencia? **R: No**
- ¿Cambia la frecuencia con la que converge a un buen modelo? **R: Sí**
Por ejemplo, prueben la siguiente arquitectura:
 - Primera capa oculta con 3 neuronas.
 - Segunda capa oculta con 3 neuronas.
 - Tercera capa oculta con 2 neuronas.



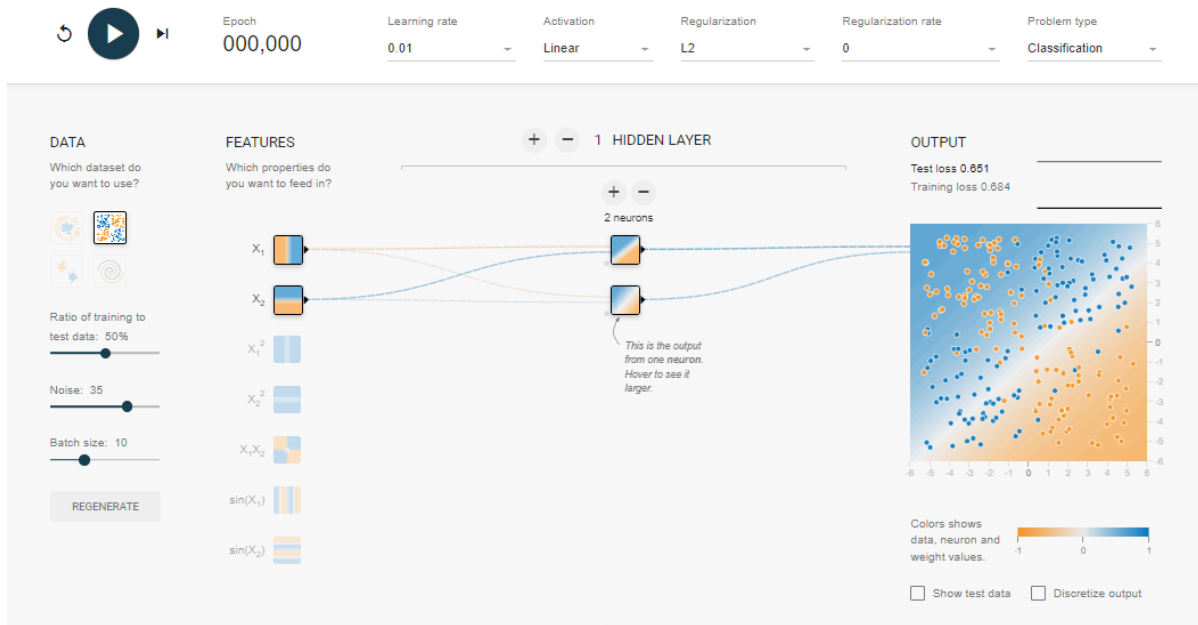


Figura 1.

Ejercicio 2. Inicialización

Este ejercicio utiliza de nuevo los datos XOR, pero examina la repetibilidad del entrenamiento de Redes Neuronales y la importancia de la inicialización.

1. Ejecuten el modelo cuatro o cinco veces como se muestra en la Figura 2. Antes de cada prueba, pulsen el botón *Reiniciar* la red para obtener una nueva inicialización aleatoria. (El botón Reiniciar la red es la flecha circular de reinicio que está justo a la izquierda del botón Reproducir). Dejen que cada prueba se ejecute durante al menos *500 pasos* para asegurar la convergencia.
 - ¿A qué forma converge la salida de cada modelo?
Se crea una forma, la mayoría de las veces, que se parece mucho a un reloj de arena, las otras veces puede ser una línea naranja entre dos segmentos azules o varia de otras formas.
 - ¿Qué nos dice esto sobre el papel de la inicialización en la optimización no convexa?
La inicialización es el factor que deriva si vamos a caer en un mínimo local o no.

- Intenten hacer el modelo ligeramente más complejo añadiendo una capa y un par de nodos extra. Repitan las pruebas del punto 1.

- ¿Añade esto alguna estabilidad adicional a los resultados?

Logra ser más consistente en clasificar, pero no nos asegura que no caigamos en un mínimos local

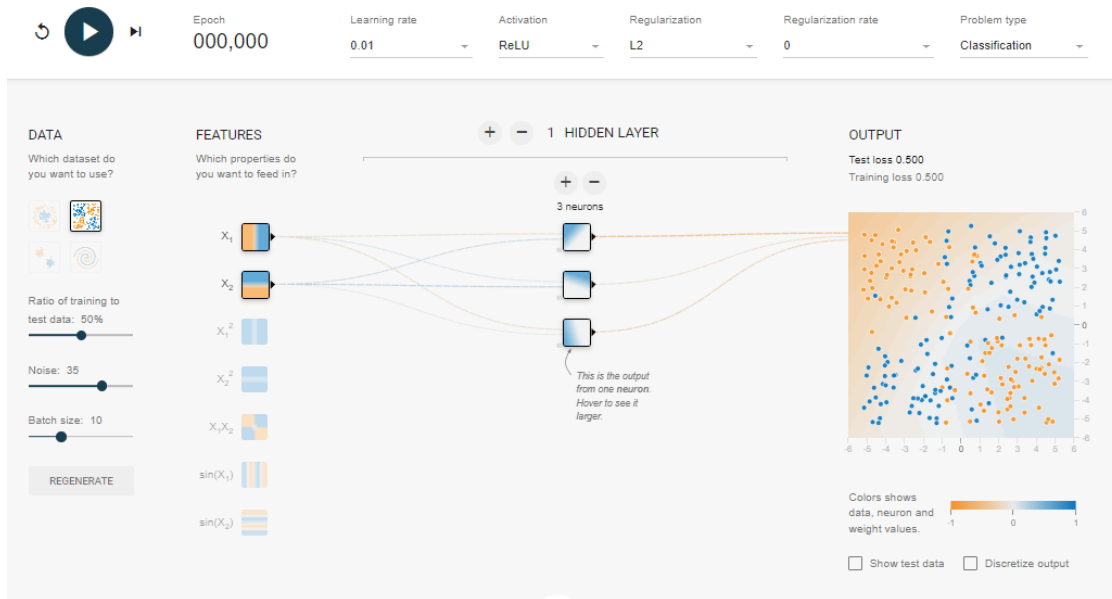


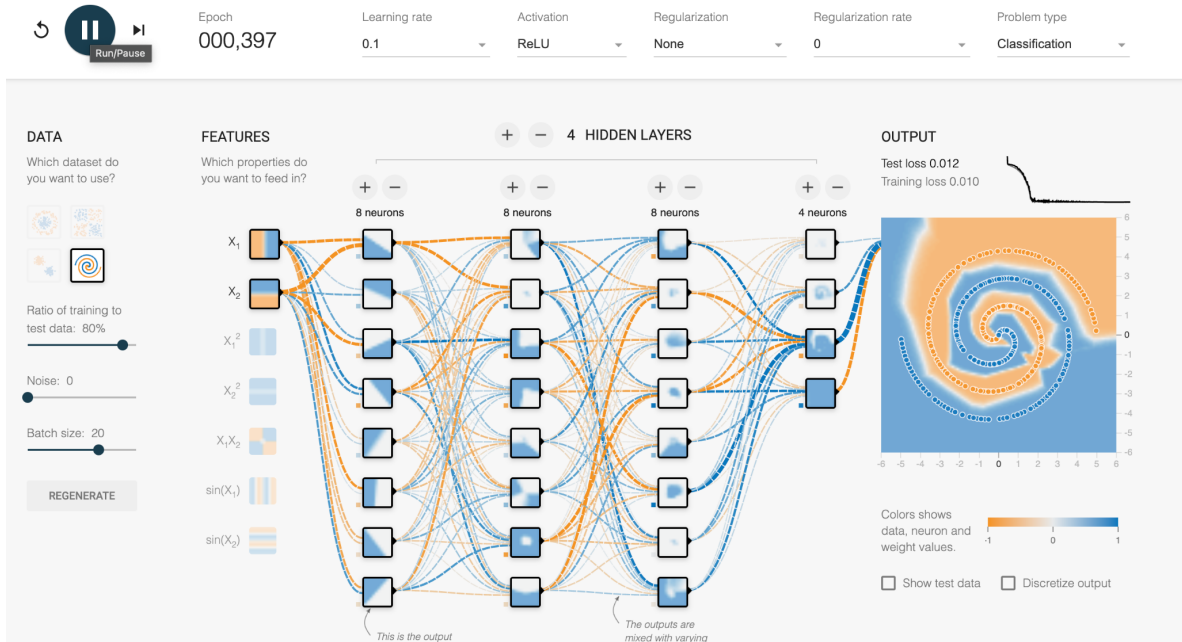
Figura 2.

Ejercicio 3. Red neuronal en Espiral

- Entrenen el mejor modelo que puedan, utilizando sólo X_1 y X_2 . Siéntanse libres de añadir o eliminar capas y neuronas, cambiar los parámetros de aprendizaje como la tasa de aprendizaje, la tasa de regularización y el tamaño del conjunto.

¿Cuál es la mejor pérdida de prueba que puede obtener?

R: La mejor pérdida que podemos obtener es 0.012



4. Incluso con las redes neuronales, a menudo es necesaria cierta ingeniería de características para lograr el mejor rendimiento. Prueben añadir funciones adicionales de producto cruzado u otras transformaciones como $\sin(X_1)$ y $\sin(X_2)$

¿Obtiene un modelo mejor?

R: Si, si agregamos el $\sin(X_1)$ y $\sin(X_2)$, obtenemos un modelo un poco mejor al que ya teníamos. Pero realmente no hay una gran mejora.

