

Day 65

深度學習理論與實作

# 深度學習體驗：啟動函數與正規化



出題教練

陳宇春 / 陳明佑



# 知識地圖 深度學習簡介

## 深度學習體驗 - 啟動函數與正規化

### 深度神經網路 Supervised Learning Deep Neural Network (DNN)

簡介 Introduction

套件介紹 Tools: Keras

組成概念 Concept

訓練技巧 Training Skill

應用案例 Application

### 卷積神經網路 Convolutional Neural Network (CNN)

簡介 introduction

套件練習 Practice with Keras

訓練技巧 Training Skill

電腦視覺 Computer Vision

### 深度學習簡介 Introduction of DNN

神經網路  
歷史

深度學習  
概念

深度學習  
體驗



# 本日知識點目標

- 理解批次大小 (Batch size) 與學習速率 (Learnig Rate) 對學習結果的影響
- 經由實驗，體驗不同啟動函數的差異性
- 體驗正規化 (Regularization) 對學習結果的影響

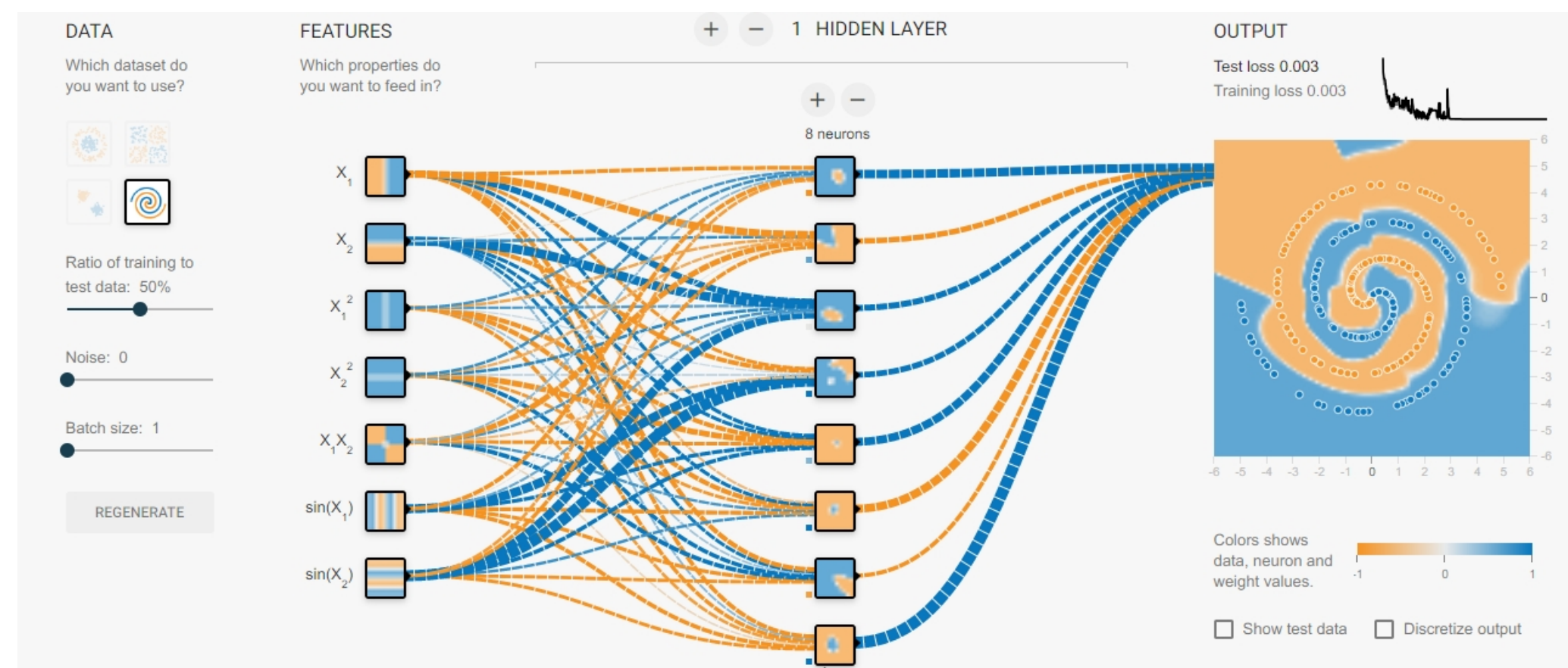
# 練習 5：切換批次大小

## 練習操作

- 資料集切換：分類資料集(右下)-螺旋雙臂，特徵全選，隱藏層 1 層 /8 神經元
- 調整 **不同的批次大小** 後執行 500 次遞迴，看看學習效果有何不同？

## 實驗結果

- 批次大小很小時，雖然收斂過程非常不穩定，但平均而言會收斂到較好的結果



**\*\* 註：**實務上，批次大小如果極小，效果確實比較好，但計算時間會相當久，因此通常會依照時間需要而折衷



# 練習 6：切換學習速率

## 練習操作

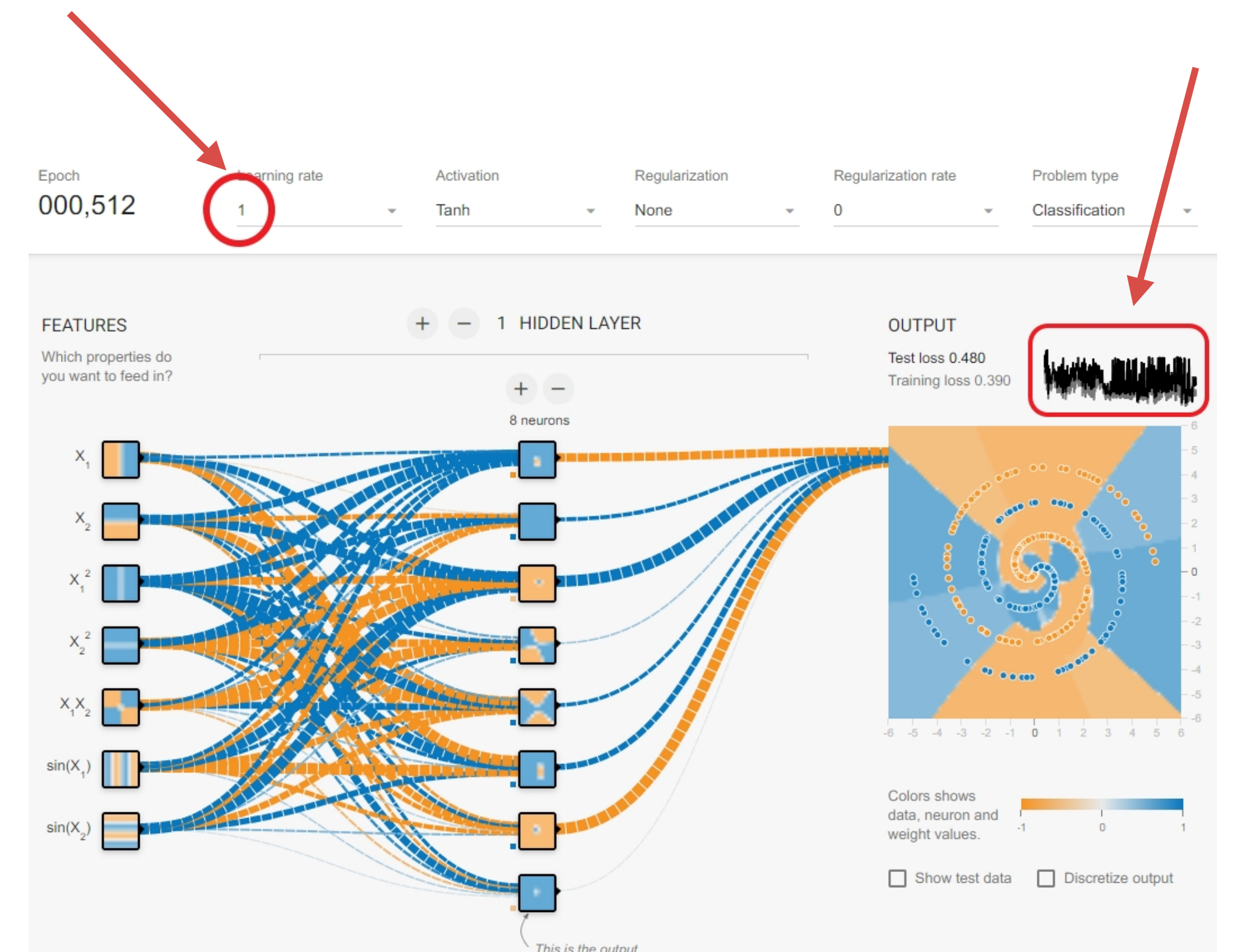
- 資料集切換：分類資料集(右下)-螺旋雙臂，特徵全選，隱藏層 1 層 /8 神經元，批次大小固定 10
- 調整 **不同的學習速率** 後執行 500 次遞迴，看看學習效果有何不同？

## 實驗結果

- 小於 0.3 時 學習速率較大時，收斂過程會越不穩定，但會收斂到較好的結果
- 大於 1 時 因為過度不穩定而導致無法收斂

選1時無法收斂

崩壞的學習曲線



# 練習 7：切換啟動函數

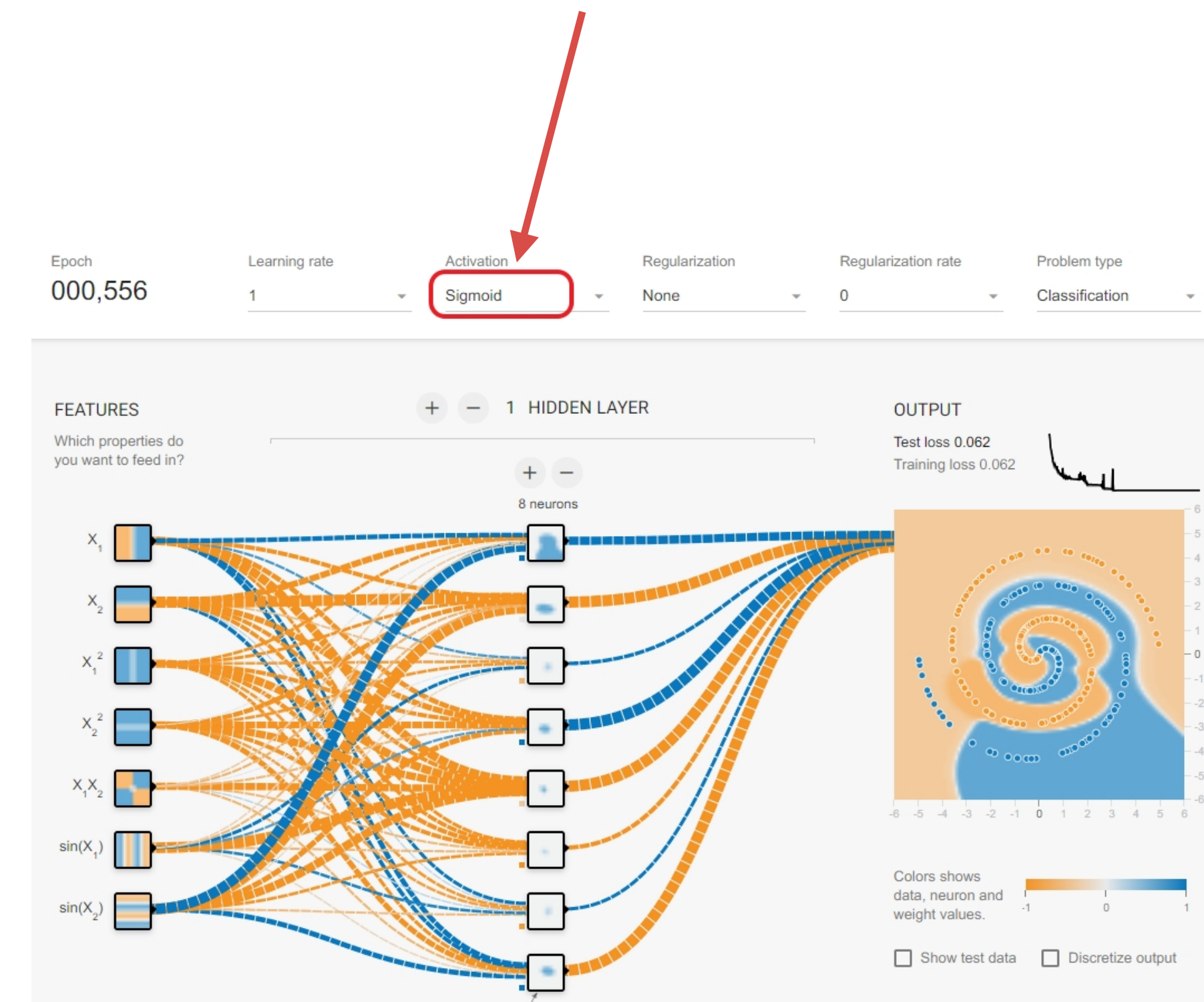
## 練習操作

- 資料集切換：分類資料集(右下)-螺旋雙臂，特徵全選，隱藏層 1層 /8 神經元，批次大小固定 10，學習速率固定 1
- 調整 **不同的啟動函數** 後執行 500 次遞迴，看看學習效果有何不同？

## 實驗結果

- 在這種極端的情形下，Tanh 會無法收斂，Relu 很快就穩定在很糟糕的分類狀態，惟有 Sigmoid 還可以收斂到不錯的結果
- 但實務上，Sigmoid 需要大量計算時間，而 Relu 則相對快得很多，這也是需要取捨的，在本例中因位只有一層，所以狀況不太明顯

最後的希望：Sigmoid





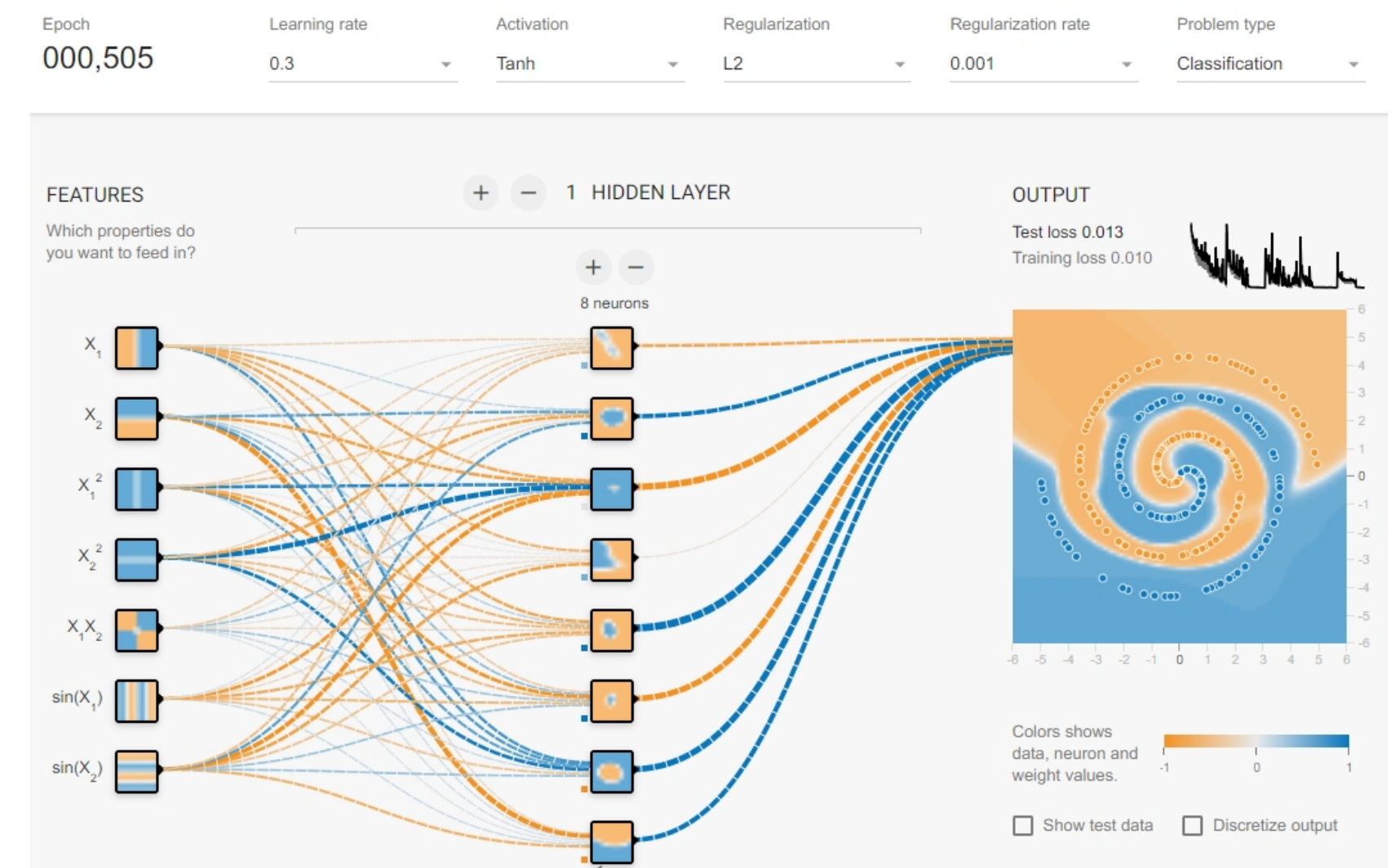
# 練習 8：切換正規化選項與參數

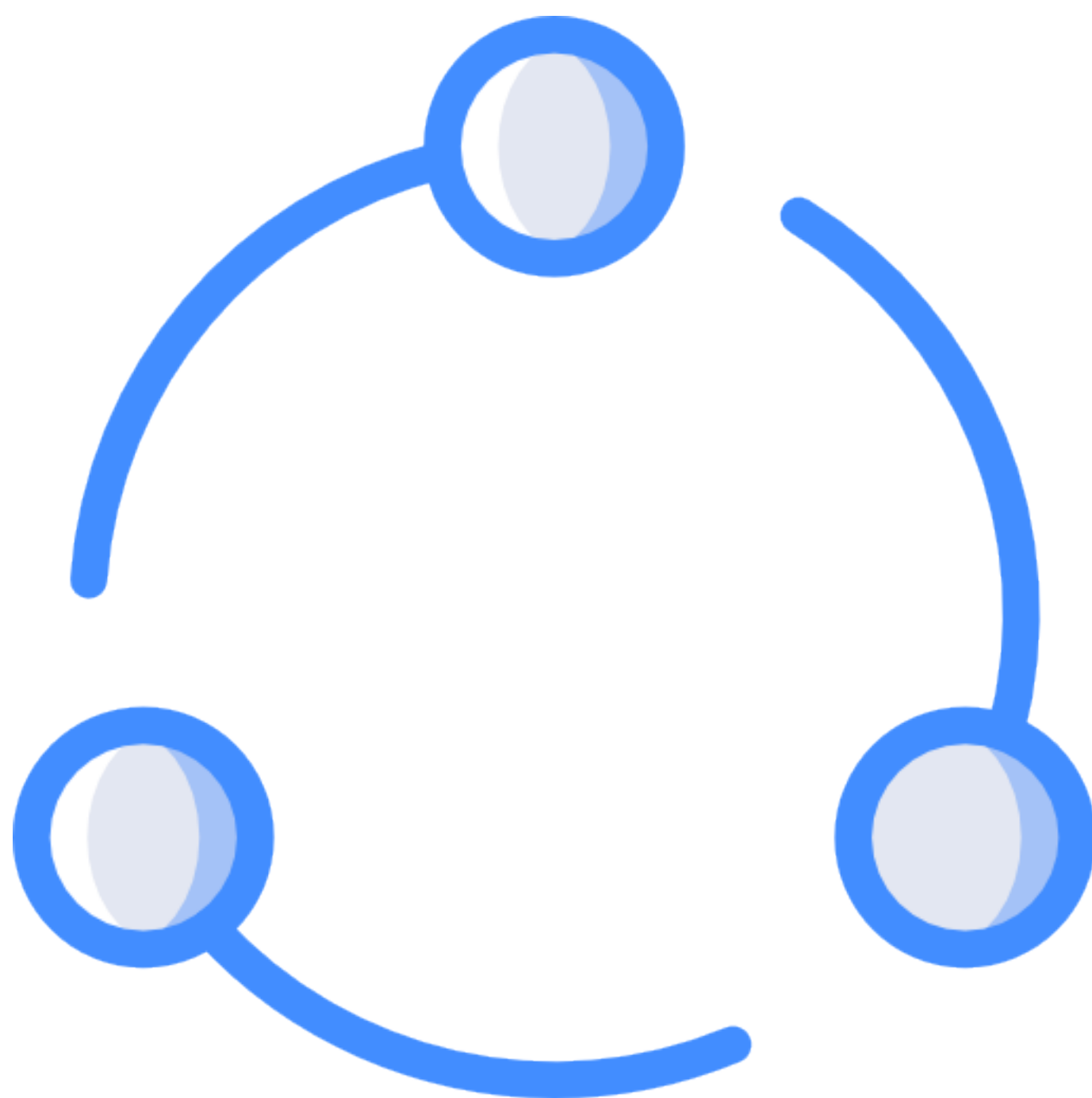
## 練習操作

- 資料集切換：分類資料集(右下)-螺旋雙臂，特徵全選，隱藏層1層/8神經元，批次大小固定 10，學習速率固定 0.3，啟動函數設為 Tanh
- 調整 不同的正規化選項與參數 後執行 500 次遞迴，看看學習效果有何不同？

## 實驗結果

- 我們已經知道上述設定本來就會收斂，只是在較小的 L1 / L2 正規劃參數下收斂比較穩定一點
- 但正規化參數只要略大，反而會讓本來能收斂的設定變得無法收斂，這點 L1 比 L2情況略嚴重，因此本例中最適合的正規化參數是 L2 + 參數 0.001
- 實務上：L1 / L2 較常使用在非深度學習上，深度學習上效果有限





- 批次大小越小：學習曲線越**不穩定**、但**收斂越快**
- 學習速率越大：學習曲線越**不穩定**、但**收斂越快**，但是與批次大小不同的是 - 學習速率大於一定以上時，有可能不穩定到**無法收斂**
- 當類神經網路層數不多時，啟動函數 Sigmoid / Tanh 的效果比 Relu 更好
- L1 / L2 正規化在非深度學習上效果較明顯，而**正規化參數較小才有效果**



# 解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業  
開始解題

