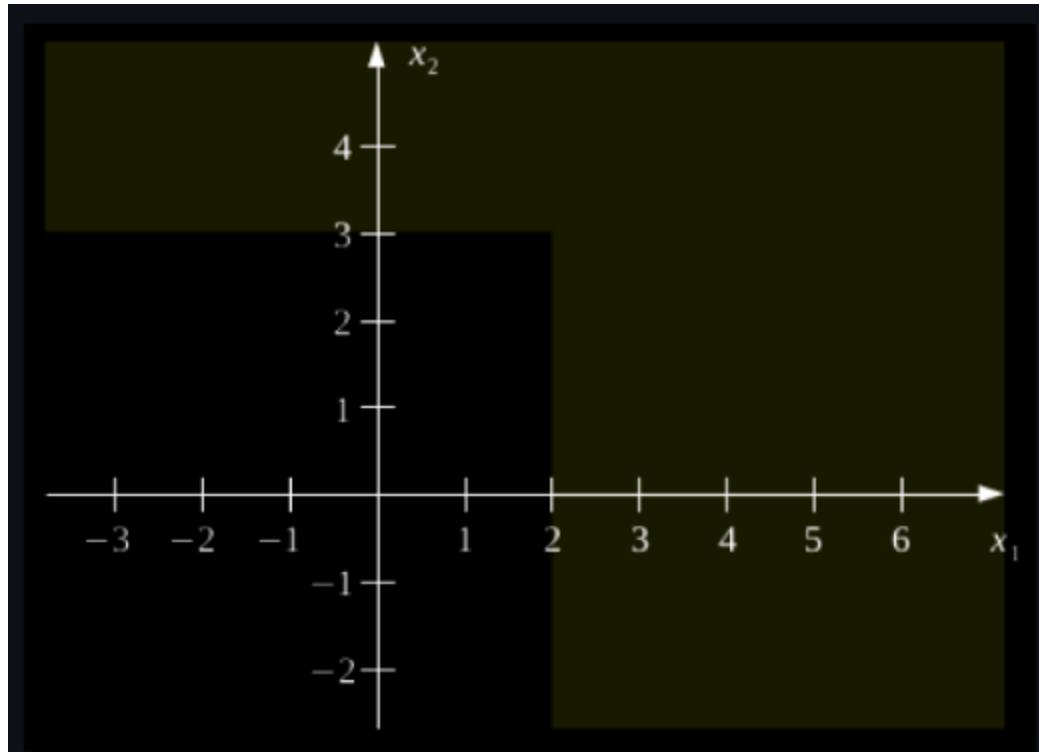


Übungsblatt: MLP

NN.MLP.01: Perzeptron-Netze

- Das war schwierig weil irgendwie keiner der Bereiche blau-grau ist (?)



- Wir nehmen an der Bereich ist $x_1 > 2$ oder $x_2 > 3$, alles andere ist -1
- Also OR-Verknüpfung der beiden Ungleichungen
- Perzeptron 1 für $x_1 > 2$
 - $y_1 = \text{sign}(w_1x_1 + w_2x_2 + b)$
 - Mit $w_1 = 1, w_2 = 0, b = -2$
 - Ergibt $\rightarrow y_1 = \text{sign}(x_1 - 2)$
 - Weil $+1$, wenn $x_1 > 2$ und -1 , wenn $x_1 < 2$
- Perzeptron 2 für $x_2 > 3$

- $y_2 = \text{sign}(0 * x_1 + 1 * x_2 - 3) = \text{sign}(x_2 - 3)$
- Weil $+1$, wenn $x_2 > 3$ und -1 , wenn $x_2 < 3$
- Heißt zusammen $\rightarrow +1$, wenn mindestens ein $y = +1$

NN.MLP.02: Vorwärtstlauf im MLP

- Dimensionen

- $W^{[1]} \in \mathbb{R}^{64 \times 25}$
- $W^{[2]} \in \mathbb{R}^{32 \times 64}$
- $W^{[3]} \in \mathbb{R}^{4 \times 32}$
- $b^{[1]} \in \mathbb{R}^{64 \times 1}$
- $b^{[2]} \in \mathbb{R}^{32 \times 1}$
- $b^{[3]} \in \mathbb{R}^{4 \times 1}$

- Vorwärtstlauf

Eingabe

- $a^{[0]} = x \in R^{25 \times 1}$

Schicht 1

- $z^{[1]} = W^{[1]}a^{[0]} + b^{[1]}$
- $a^{[1]} = \text{ReLU}(z^{[1]})$

Schicht 2

- $z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}$
- $a^{[2]} = \text{ReLU}(z^{[2]})$

Ausgabeschicht

- $z^{[3]} = W^{[3]}a^{[2]} + b^{[3]}$
- $a^{[3]} = \text{ReLU}(z^{[3]})$

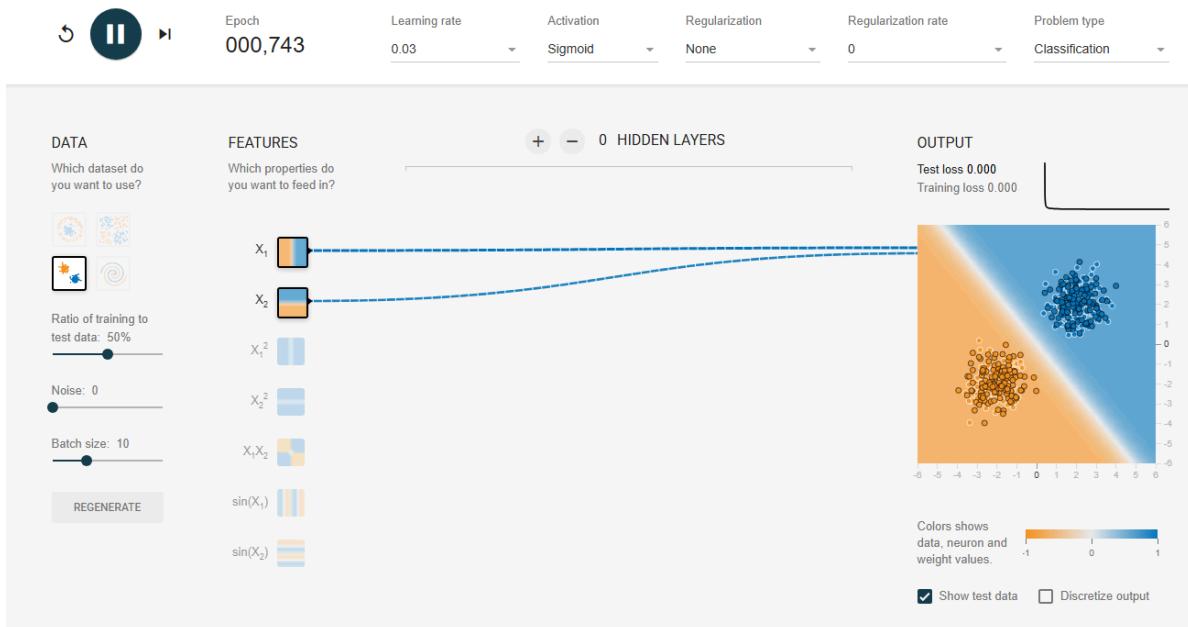
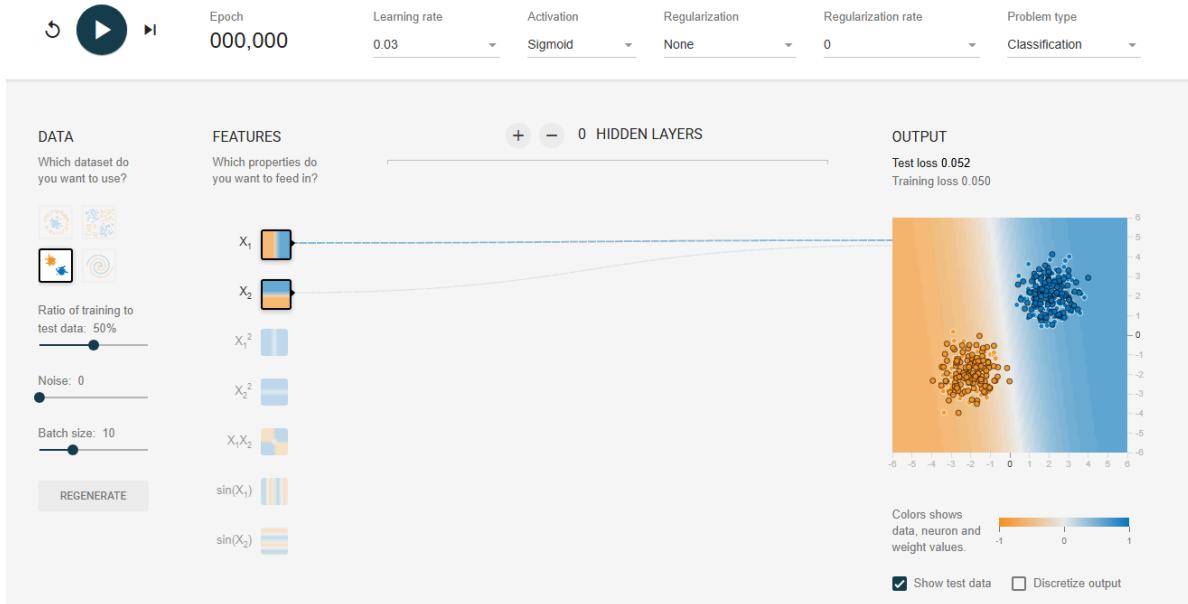
Ausgabe des Netzwerks

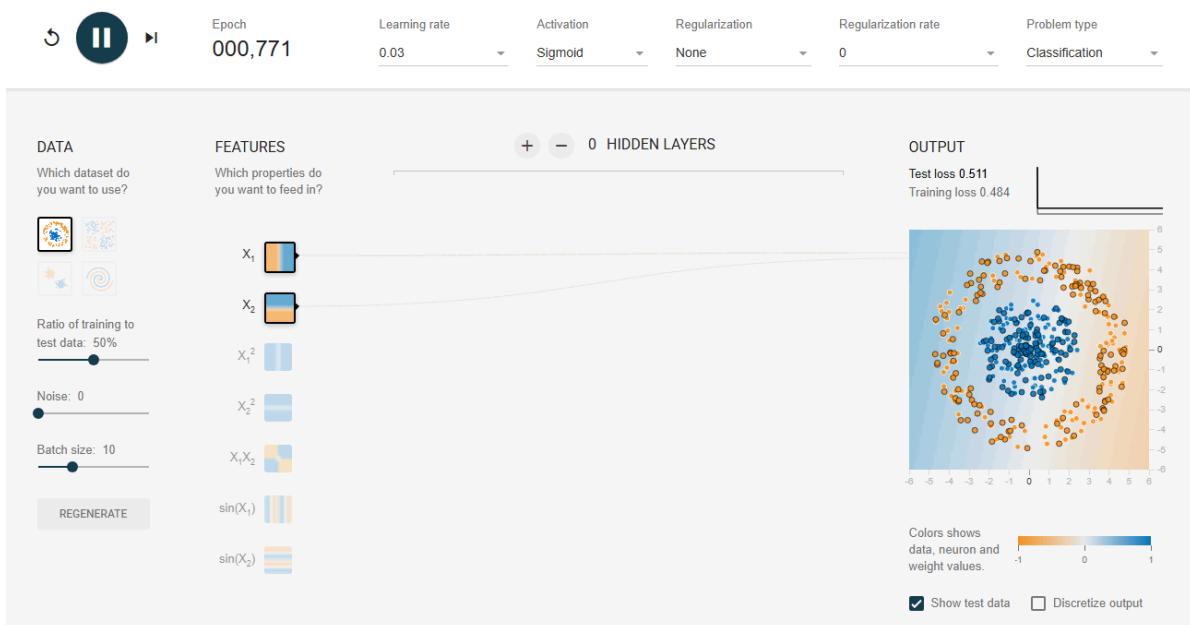
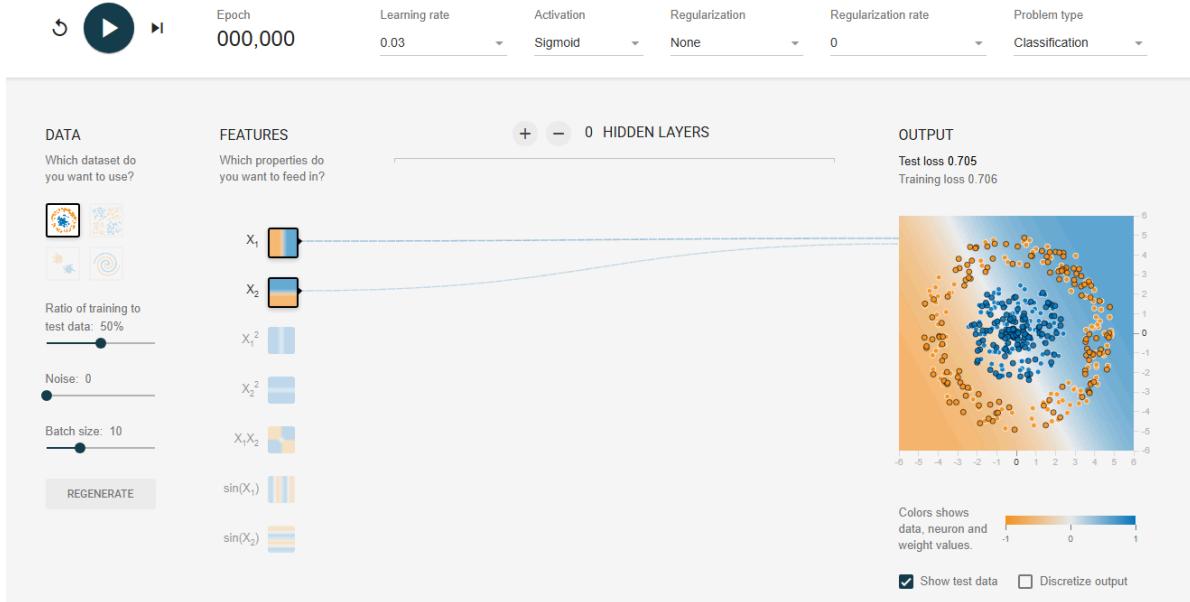
- $y = a^{[3]} \in \mathbb{R}^{4 \times 1}$
- Das ganze Netz könnte sich für Klassifikations und Regressionsprobleme mit vierdimensionaler Ausgabe eignen

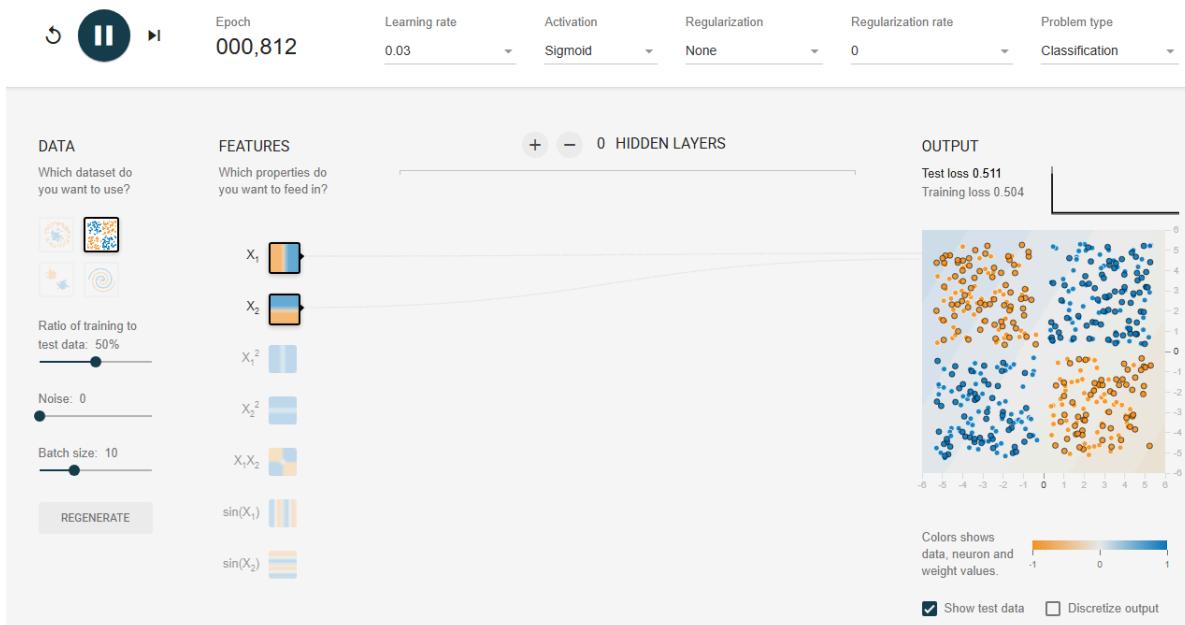
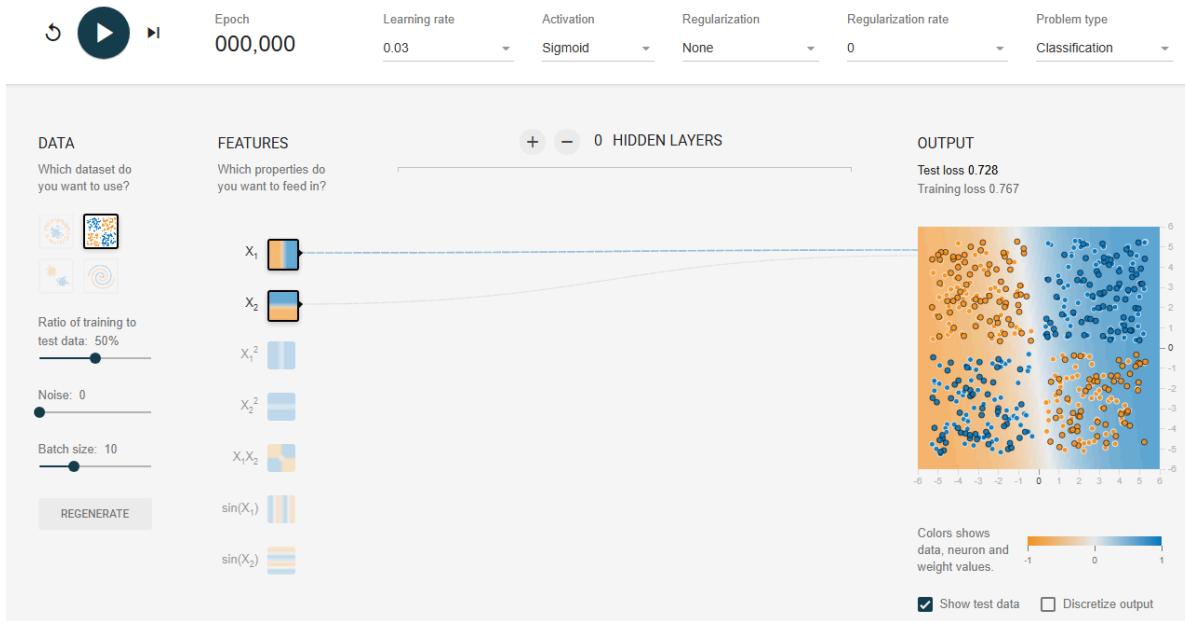
NN.MLP.03: Tensorflow Playground

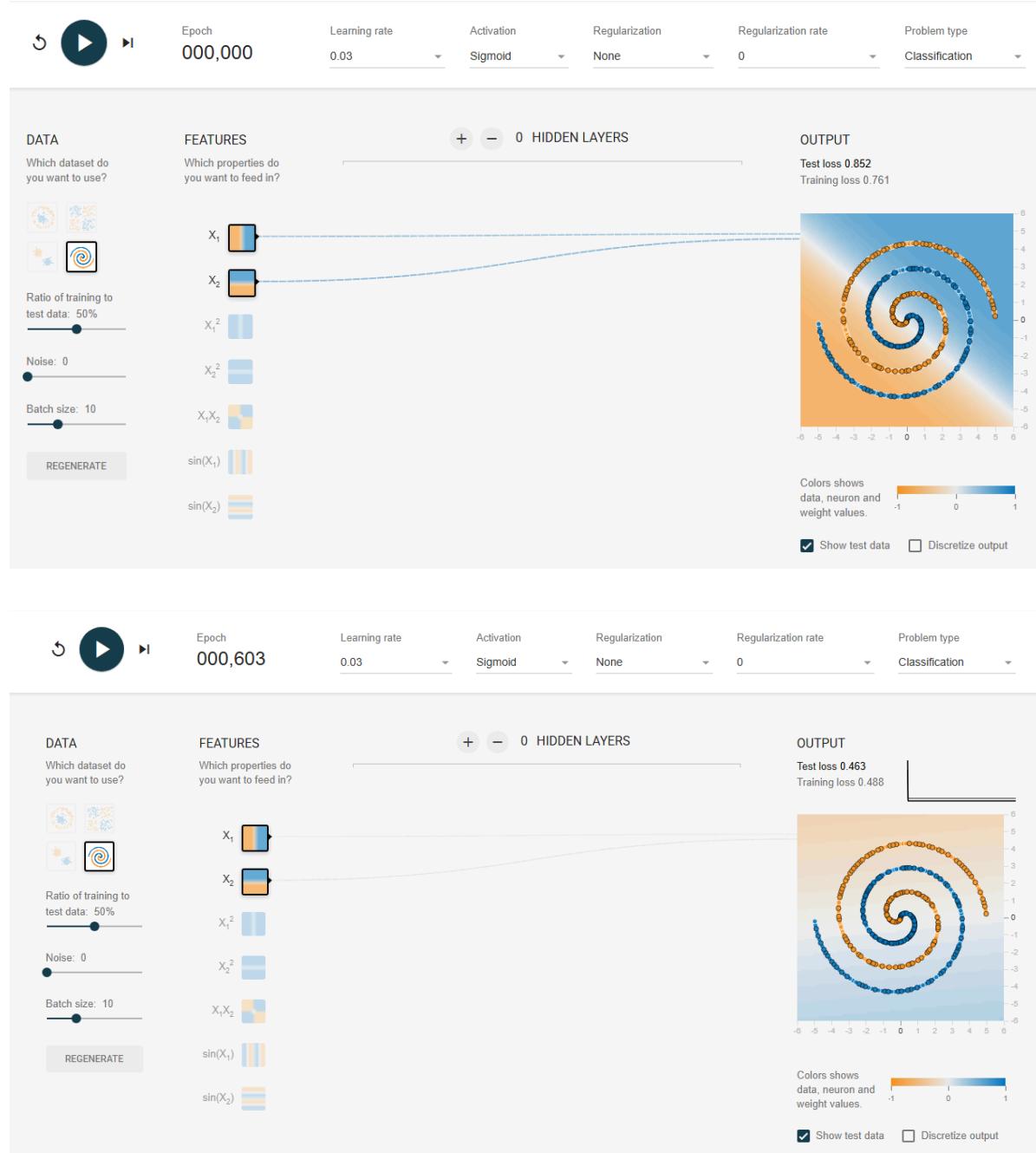
1. Logisches Regressionsmodell

- Wie verhält sich die Entscheidungsgrenze?
 - Bleiben immer linear, kann gaussian gut abbilden, aber nicht die anderen
- Was können Sie über Trainings- und Testkosten sagen? Entsteht eine Überanpassung?
 - Gaussian → Trainings und Testkosten sinken schnell und stark, gute Generalisierung
 - Andere Datensätze → Trainings- und Testkosten bleiben hoch, Underfitting
 - Kein Overfitting
- Wie schnell wird die Entscheidungsgrenze berechnet?
 - Sehr schnell, nach wenigen Sekunden, nach wenigen Iterationen
- Können alle Datenpunkte jedes mal korrekt klassifiziert werden? Warum?
 - Nein, weil lineare Regression nur bei linearen Entscheidungsgrenzen richtig funktioniert
- Untersuchen und vergleichen Sie die Ausgaben der Zellen in den versteckten Schichten, in dem Sie die Maus über die jeweilige Zelle bewegen. Bemerken Sie einen wesentlichen Unterschied in den Ausgaben der ersten Schicht im Vergleich zu der letzten Schicht?
 - Keine versteckten Schichten





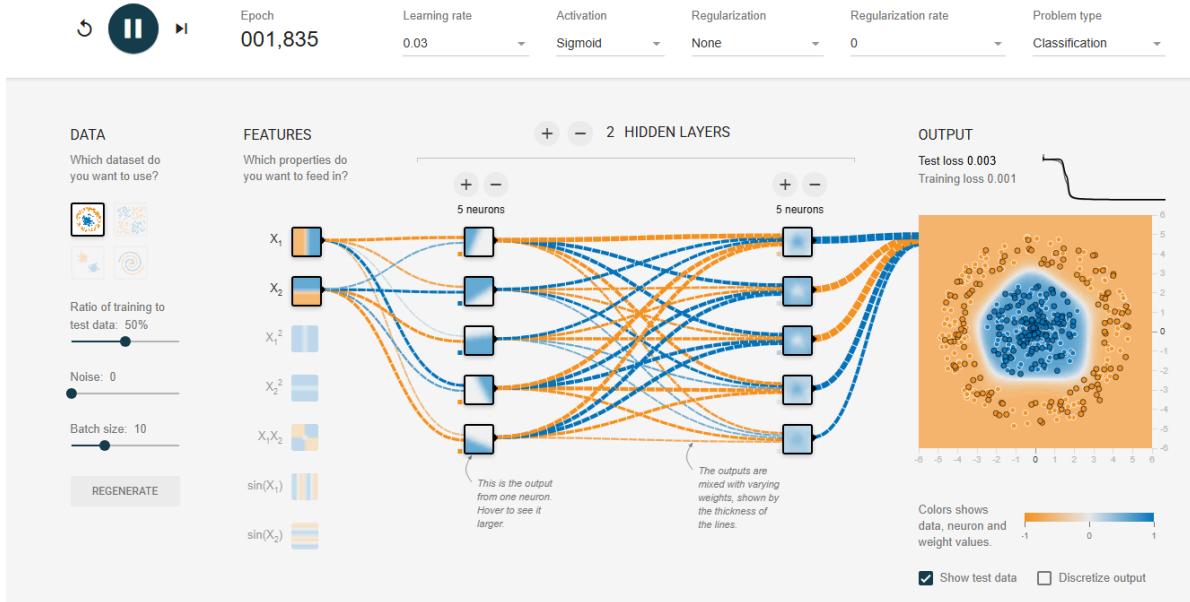




2. MLP

- Wie verhält sich die Entscheidungsgrenze?
 - ReLU → kantig
 - Tanh → glatt, weich
 - Sigmoid → sehr glatt

- Mehr Neuronen, Circle und Spiral besser trennbar, mehr nichtlinear und akkurater
- Was können Sie über Trainings- und Testkosten sagen? Entsteht eine Überanpassung?
 - Kosten sinken mit größerem Netz
 - Bei größerem Netz Überanpassung möglich
- Wie schnell wird die Entscheidungsgrenze berechnet?
 - ReLU > tanh > Sigmoid
 - Schnellstes → Langsamer
- Können alle Datenpunkte jedes mal korrekt klassifiziert werden? Warum?
 - Bei größeren Netzen ja, bei kleineren nicht
 - Bei sehr komplexen Grenzen kann schnell overfitting passieren
 - Grund → Das Modell lernt nur die Approximation, nicht die exakten Regeln
- Untersuchen und vergleichen Sie die Ausgaben der Zellen in den versteckten Schichten, in dem Sie die Maus über die jeweilige Zelle bewegen. Bemerken Sie einen wesentlichen Unterschied in den Ausgaben der ersten Schicht im Vergleich zu der letzten Schicht?
 - In der ersten Schicht sind die Muster einfacher und die Aktivierung ist global, in der letzten Schicht sind die Muster abstrakter und die Aktivierung ist lokalisiert



3. Noise-Level 15

- Wie verhält sich die Entscheidungsgrenze?
 - Glatt, Teil des Rauschens wird ignoriert, Zickzack
 - Von ReLU zu Sigmoid werden die Grenzen weniger kantig
- Was können Sie über Trainings- und Testkosten sagen? Entsteht eine Überanpassung?
 - Kosten bleiben relativ hoch
 - Da die Trainingskosten niedrig werden und die Testkosten höher bleiben, kann man wahrscheinlich auf Overfitting schließen
 - Das Rauschen wird gelernt anstatt der Struktur
- Wie schnell wird die Entscheidungsgrenze berechnet?
 - Langsamer durch das Rauschen, instabileres Training
- Können alle Datenpunkte jedes mal korrekt klassifiziert werden? Warum?
 - Besonders bei den kleineren Netzen oft nicht, weil die Daten durch das Rauschen zu unklar sind
- Untersuchen und vergleichen Sie die Ausgaben der Zellen in den versteckten Schichten, in dem Sie die Maus über die jeweilige Zelle

bewegen. Bemerken Sie einen wesentlichen Unterschied in den Ausgaben der ersten Schicht im Vergleich zu der letzten Schicht?

- Die erste Schicht ist glatter und strukturierter, wohingegen die letzte Schicht rauschanfälliger ist

