

NN.MLP.01: Perzeptron-Netze:

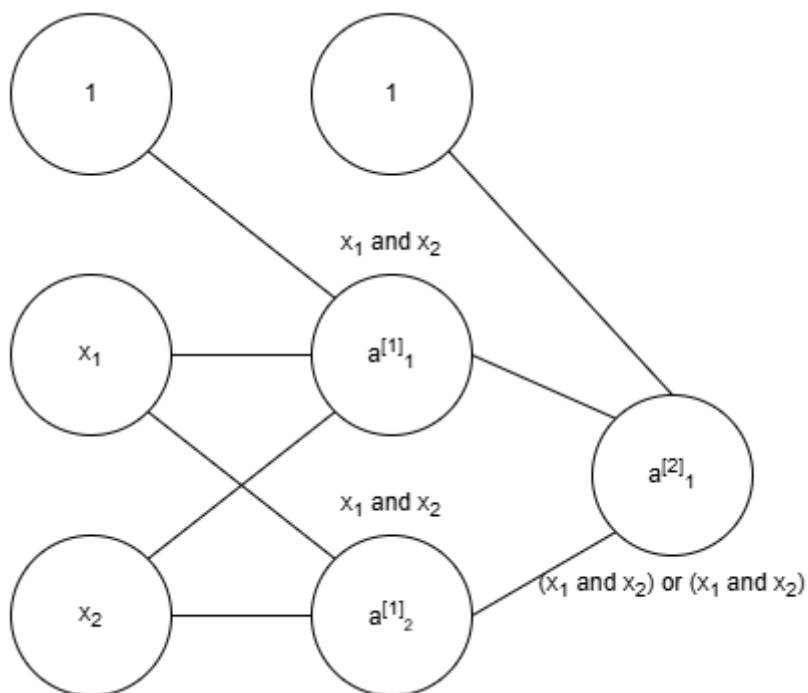
Beschreibung der blau-grauen Fläche:

- $x_1 > 2$
- $x_2 > 3$

Perzeptron Gewichtsvektor:

- P1: $-3 * x_2 + 0 * x_1 + 0$
- P2: $0 * x_2 - 2 * x_1 + 0$

Neuronales Netz Konzept:

**NN.MLP.02: Vorwärtslauf im MLP:**

- Dimensionen:
 - $W^{[1]} = (64 \times 25)$
 - $W^{[2]} = (32 \times 64)$
 - $W^{[3]} = (4 \times 32)$
 - $b^{[1]} = (64 \times 1)$
 - $b^{[2]} = (32 \times 1)$
 - $b^{[3]} = (4 \times 1)$
- Berechnungen:
 - Ausgabe: $a^{[4]}_i = \text{relu}(W^{[3]} * a^{[3]} + b^{[3]}_i)$

Forward Propagation:

- $a^{[0]} := x = \text{Eingabe}$
- $z^{[1]} = W^{[1]} * x + b^{[1]}$
 $a^{[1]} = \text{relu}(z^{[1]})$
- $z^{[2]} = W^{[2]} * a^{[1]} + b^{[2]}$
 $a^{[2]} = \text{relu}(z^{[2]})$
- $z^{[3]} = W^{[3]} * a^{[2]} + b^{[3]}$
 $a^{[3]} = \text{relu}(z^{[3]}) = y$

NN.MLP.03: Tensorflow Playground:

1.

Gaussian Datensatz:

- Entscheidungsgrenze ist eine Lineare Gerade die den Datensatz in zwei Teile aufteilt.
- Die Trainings und Testkosten sind null. Es entsteht keine Überanpassung
- Die Entscheidungsgrenze wird schnell in ein paar Epochen berechnet
- Es können alle Datenpunkte korrekt klassifiziert werden da die Gerade die Datenpunkte perfekt in zwei Klassen aufteilt.

Circle Datensatz:

- Entscheidungsgrenze ist steigende Gerade wobei die Punkte unter der Geraden von 0 – 1 und über der Geraden mit 0 - (-1) Klassifiziert werden
- Trainings und Testverluste liegen bei 0.510 und 0.506 also recht hoch. Überanpassung ist nicht vorhanden
- Entscheidungsgrenze wird sehr schnell in ein paar Epochen berechnet
- Die Entscheidungsgrenze konnte die Datenpunkte nur sehr schlecht berechnen da der Datensatz auch gar nicht linear trennbar ist

Exclusive or Datensatz:

- Hier ist es ähnlich wie bei dem Circle Datensatz. Die Entscheidungsgrenze ist diesmal eine fallende Gerade die versucht den Datensatz Linear zu trennen. Alle Punkte über der Geraden werden mit positiven Werten klassifiziert und unter der geraden negativ. Genau auf der Gerade wird mit 0 klassifiziert
- Trainings und Testverluste liegen bei 0.510 und 0.521 also recht hoch. Überanpassung ist nicht vorhanden
- Entscheidungsgrenze wird sehr schnell in ein paar Epochen berechnet
- Die Entscheidungsgrenze konnte die Datenpunkte nur sehr schlecht berechnen da der Datensatz auch hier gar nicht linear trennbar ist

Spiral Datensatz:

- Auch hier ist es sehr ähnlich wie die vorherigen Datensätze. Die Entscheidungsgrenze ist hier genauso wie beim Circle Datensatz
- Trainings und Testverluste liegen bei 0.467 und 0.482 also recht hoch, aber ein bisschen besser als der Exclusive or und Circle Datensatz. Überanpassung ist nicht vorhanden

- Entscheidungsgrenze wird auch hier wieder sehr schnell in ein paar Epochen berechnet
- Die Entscheidungsgrenze konnte die Datenpunkte nur sehr schlecht berechnen da der Datensatz auch hier gar nicht linear trennbar ist

2.

Circle Datensatz: eine versteckten Schicht mit zwei Neuronen

- ReLU:
 - o Die Entscheidungsgrenze sind zwei Geraden, die die Blauen Datenpunkte gerade so mit erfassen.
 - o Trainings und Testverluste liegen bei 0.286 und 0.207. Überanpassung nicht vorhanden
 - o Entscheidungsgrenze wird in ca. 20 Epochen berechnet
 - o Die Entscheidungsgrenze konnte den Satz nur mittelmäßig gut klassifizieren. Es gibt viele Orange Punkte, die noch falsch klassifiziert werden und ein viele blaue die genau auf der Grenze liegen
 - o Die zwei Ausgaben der Neuronen der Versteckten Schicht sind etwas versetzt voneinander und was als 0 und als 1 klassifiziert ist unterschiedlich
- Sigmoid:
 - o Die Entscheidungsgrenze sieht hier ähnlich aus wie die aus der ReLU funktion aber etwas abgerundet
 - o Trainings und Testverluste liegen bei 0.281 und 0.205. Überanpassung nicht vorhanden
 - o Entscheidungsgrenze wird in ca. 200 Epochen berechnet, also recht langsam
 - o Die Entscheidungsgrenze konnte den Datensatz auch hier nur mittelmäßig gut berechnen
 - o Die Ausgaben der versteckten Neuronen sind zwei geraden die unterhalb der Geraden als 1 und über 0 klassifizieren, es wird nichts als -1 klassifiziert. Sie sind nur etwas versetzt zueinander
- Tanh:
 - o Entscheidungsgrenze auch hier sehr ähnlich zu den vorherigen Aktivierungsfunktionen
 - o Trainings und Testverluste liegen bei 0.284 und 0.204. Überanpassung nicht vorhanden
 - o Entscheidungsgrenze wird in ca. 40 Epochen berechnet
 - o Die Entscheidungsgrenze konnte den Datensatz auch hier nur mittelmäßig gut berechnen
 - o Die Ausgaben der versteckten Neuronen sind zwei geraden die unterhalb der Geraden als 1 und über -1 klassifizieren. Sie sind nur etwas versetzt zueinander

Circle Datensatz: eine versteckte Schicht mit drei Neuronen

- ReLU:
 - o Die Entscheidungsgrenze ist ein Sechseck mit ungleich langen kanten

- Trainings und Testverluste liegen bei 0.017 und 0.016. Überanpassung nicht vorhanden
- Entscheidungsgrenze wird in ca. 50 Epochen berechnet
- Die Entscheidungsgrenze konnte den Datensatz recht gut berechnen, es liegen nur ein paar blaue punkte genau auf der Grenze
- Die Ausgaben der versteckten Neuronen sind geraden die entweder eine Seite mit 0 und die andere mit 1 klassifizieren
- Sigmoid:
 - Die Entscheidungsgrenze ähnelt hier einem Dreieck mit abgerundeten Ecken
 - Trainings und Testverluste liegen bei 0.014 und 0.014. Überanpassung nicht vorhanden
 - Die Entscheidungsgrenze wird in ca 300 berechnet, also sehr langsam
 - Die Entscheidungsgrenze konnte den Datensatz recht gut berechnen, es liegen nur ein paar blaue punkte genau auf der Grenze, wie bei der ReLU funktion
 - Die Ausgaben sind hier ähnlich wie bei der ReLU nur ist der Übergang von 0 zu 1 etwas größer
- Tanh:
 - Die Entscheidungsgrenze ähnelt hier einem Dreieck mit abgerundeten Ecken wie bei der Sigmoid Funktion
 - Trainings und Testverluste liegen bei 0.021 und 0.020, was schlechter als die anderen Funktionen ist aber immer noch gut. Überanpassung nicht vorhanden
 - Entscheidungsgrenze wird in ca. 200 Epochen berechnet
 - Die Entscheidungsgrenze konnte den Datensatz recht gut berechnen, es liegen nur ein paar blaue punkte genau auf der Grenze, wie bei der den anderen Funktionen
 - Die Ausgaben sind hier ähnlich wie bei den anderen nur wird hier schon zwischen -1 und 1 klassifiziert

Circle Datensatz: eine versteckte Schicht mit fünf Neuronen

- ReLU:
 - Die Entscheidungsgrenze ist hier wieder ein Sechseck aber etwas schief als bei drei Neuronen
 - Trainings und Testverluste liegen bei 0.011 und 0.007. Überanpassung nicht vorhanden
 - Entscheidungsgrenze wird in ca. 50 Epochen berechnet
 - Die Entscheidungsgrenze konnte den Datensatz recht gut berechnen, es liegen nun keine blauen punkte mehr genau auf der Grenze, aber immer noch anliegend
 - Das dritte Neuron ist hier eine Gerade, die im Gegensatz zu den anderen Neuronen fasst den ganzen Datensatz als größer 0 klassifiziert
- Sigmoid:

- Entscheidungsgrenze ist wieder ein Dreieck aber noch abgerundeter als mit drei Neuronen
- Trainings und Testverluste liegen bei 0.029 und 0.026
- Entscheidungsgrenze wird in ca. 400 Epochen berechnet, was bis jetzt am langsamsten ist
- Datensatz konnte gut klassifiziert werden, es werden ein paar blaue noch fasst als 0 klassifiziert
- Tanh:
 - Die Entscheidungsgrenze sieht hier fast so aus wie die von der Sigmoid Funktion
 - Trainings und Testverluste liegen bei 0.028 und 0.027
 - Entscheidungsgrenze wird in ca. 150 Epochen berechnet
 - Datensatz konnte gut klassifiziert werden, es werden ein paar blaue noch fasst als 0 klassifiziert wie bei der Sigmoid Funktion

Circle Datensatz: zwei versteckte Sichten mit jeweils 5 Neuronen

- ReLU:
 - Die Entscheidungsgrenze ähnelt hier einem Dreieck, aber nur ein wenig
 - Trainings und Testverluste liegen bei 0.005 und 0.005, was bis jetzt die besten Werte sind
 - Entscheidungsgrenze wird in ca. 100 Epochen berechnet
 - Der Datensatz wird sehr gut klassifiziert, es gibt nur noch wenig blaue Punkte, die der Grenze sehr nah kommen
- Sigmoid:
 - Die Entscheidungsgrenze ist fast ein Kreis
 - Trainings und Testverluste liegen bei 0.012 und 0.012
 - Entscheidungsgrenze wird in ca. 500 Epochen berechnet, was sehr schlecht ist
 - Der Datensatz wird auch sehr gut klassifiziert
- Tanh:
 - Die Entscheidungsgrenze ist ähnlich wie bei ReLU, Dreiecksform noch ansatzweise erkennbar
 - Trainings und Testverluste liegen bei 0.003 und 0.003, was sogar besser als ReLU ist
 - Entscheidungsgrenze wird in ca. 100 Epochen berechnet
 - Der Datensatz wird auch sehr gut klassifiziert

Circle Datensatz: drei versteckte Schichten mit jeweils 7 Neuronen

- ReLU:
 - Entscheidungsgrenze wieder ein Sechseck
 - Trainings und Testverluste liegen bei 0.001 und 0.001, besser als das Experiment davor
 - Entscheidungsgrenze wird in ca. 50 Epochen berechnet
 - Datensatz wird sehr gut klassifiziert, es gibt keine blauen Punkte die nah an der Grenze liegen
- Sigmoid:
 - Entscheidungsgrenze nicht erkennbar oder nicht vorhanden

- Trainings und Testverluste liegen bei 0.5 und 0.499
- Entscheidungsgrenze wird in ein oder zwei Epochen berechnet
- Datensatz wird sehr schlecht klassifiziert bzw. eigentlich gar nicht, also alles wird als 0 klassifiziert
- Tanh:
 - Entscheidungsgrenze ähnelt mehr einem Kreis
 - Trainings und Testverluste liegen bei 0.001 und 0.001, wie bei der ReLU-Funktion
 - Entscheidungsgrenze wird in ca. 100 Epochen berechnet
 - Datensatz wird sehr gut klassifiziert und vor allem schnell

Circle Datensatz: vier versteckte Schichten mit jeweils 7 Neuronen

- ReLU:
 - Entscheidungsgrenze ist ein Neuneck mit ungleichen Kanten
 - Trainings und Testverluste liegen bei 0.000 und 0.000, also keine Verluste vorhanden
 - Entscheidungsgrenze wird in ca. 100 Epochen berechnet
 - Datensatz wird perfekt und schnell klassifiziert
- Sigmoid:
 - Entscheidungsgrenze nicht vorhanden, alles wird als 0 klassifiziert
 - Trainings und Testverluste liegen bei 0.5 und 0.5
 - Entscheidungsgrenze wird gar nicht berechnet
 - Datensatz wird sehr schlecht bzw. gar nicht klassifiziert
- Tanh:
 - Entscheidungsgrenze ähnelt einem Kreis
 - Trainings und Testverluste liegen bei 0.002 und 0.002, also nur sehr wenig Verluste
 - Entscheidungsgrenze wird in ca. 50 Epochen berechnet
 - Datensatz wird sehr gut klassifiziert

Spiral Datensatz: eine versteckte Schicht mit zwei Neuronen

- ReLU:
 - Entscheidungsgrenze ist eine fallende Gerade die durch den Mittelpunkt der Spirale läuft
 - Trainings und Testverluste liegen bei 0.504 und 0.469, also sehr hohe Verluste
 - Entscheidungsgrenze wird in ca. 100 Epochen berechnet
 - Datensatz wird sehr schlecht klassifiziert, eine Gerade kann diese Daten nicht wirklich trennen
 - Die beiden Ausgaben der versteckten Neuronen sind auch nur Geraden, die beide sehr ähnlich aussehen
- Sigmoid:
 - Auch hier ist die Entscheidungsgrenze eine Gerade
 - Trainings und Testverluste liegen bei 0.485 und 0.475, also auch hier sehr hohe Verluste

- Entscheidungsgrenze wird in ca. 50 Epochen berechnet
- Datensatz wird sehr schlecht klassifiziert, eine Gerade kann diese Daten nicht wirklich trennen
- Die beiden Ausgaben der versteckten Neuronen sind auch hier nur Geraden, die beide sehr ähnlich aussehen
- Tanh:
 - Entscheidungsgrenze ähnelt einem Schrägen V' alles im V' wird als 0 klassifiziert alles über dem V' als -1 und unter mit +1
 - Trainings und Testverluste liegen bei 0.466 und 0.429, also sehr hohe Verluste
 - Entscheidungsgrenze wurde in ca. 2000 Epochen berechnet also sehr langsam
 - Datensatz wird auch hier nur sehr schlecht klassifiziert, viele Punkte werden als 0 klassifiziert
 - Die Ausgaben der versteckten Neurone bilden die zwei Linien zum V

Spiral Datensatz: eine versteckte Schicht mit drei Neuronen

- ReLU:
 - Entscheidungsgrenze ähnelt keiner geometrischen Form
 - Trainings und Testverluste liegen bei 0.460 und 0.555. Eine Überanpassung ist vorhanden erkennbar daran, dass Testverluste nach ein paar Epochen anfangen zu steigen
 - Entscheidungsgrenze wurde in ca. 500 Epochen berechnet also recht langsam
 - Datensatz wird schlecht klassifiziert und das Neuronale Netz fängt an die Punkte auswendig zu lernen
 - Die Ausgaben der versteckten Schicht sind einfache Geraden mit unterschiedlicher Steigung
- Sigmoid:
 - Entscheidungsgrenze ähnelt einem Schrägen S'
 - Trainings und Testverluste liegen bei 0.460 und 0.555. Eine Überanpassung ist auch hier vorhanden erkennbar daran, dass Testverluste nach ein paar Epochen anfangen zu steigen
 - Entscheidungsgrenze wurde in ca. 400 Epochen berechnet also recht langsam
 - Datensatz wird auch hier schlecht klassifiziert und das Neuronale Netz fängt an die Punkte auswendig zu lernen
- Tanh:
 - Entscheidungsgrenze ähnelt keiner Geometrischen Form
 - Trainings und Testverluste liegen bei 0.398 und 0.475. Eine Überanpassung ist auch hier vorhanden erkennbar daran, dass Testverluste nach ein paar Epochen anfangen zu steigen
 - Entscheidungsgrenze wurde in ca. 400 Epochen berechnet also recht langsam
 - Datensatz wird schlecht klassifiziert und das Neuronale Netz fängt an die Punkte auswendig zu lernen

Spiral Datensatz: eine versteckte Schicht mit fünf Neuronen

- ReLU:

- Entscheidungsgrenzen ähnelt zwei fallenden Geraden die parallel zueinander verlaufen
- Trainings und Testverluste liegen bei 0.506 und 0.442. Eine Überanpassung ist hier noch ein bisschen vorhanden, die Verlust Werte schwanken viel über die Epochen
- Entscheidungsgrenze wurde in ca. 200 Epochen berechnet
- Datensatz wird schlecht klassifiziert
- Sigmoid:
 - Entscheidungsgrenze ist hier wieder eine fallende Gerade die durch den Mittelpunkt der Spirale läuft
 - Trainings und Testverluste liegen bei 0.473 und 0.488, Überanpassung ist nicht vorhanden
 - Entscheidungsgrenze wird in ca. 100 Epochen berechnet
 - Datensatz wird trotzdem nur schlecht klassifiziert, eine Gerade kann diese Daten nur schlecht trennen
- Tanh:
 - Entscheidungsgrenze ähnelt keiner geometrischen form
 - Trainings und Testverluste liegen bei 0.506 und 0.442. Eine Überanpassung ist hier leicht vorhanden
 - Entscheidungsgrenze wird in ca. 1000 Epochen berechnet
 - Datensatz wird schlecht klassifiziert

Spiral Datensatz: zwei versteckte Schichten mit fünf Neuronen

- ReLU:
 - Entscheidungsgrenze nähert sich der Spiral form an ist aber nicht perfekt und sehr eckig
 - Trainings und Testverluste liegen bei 0.015 und 0.034. Überanpassung ist nicht vorhanden
 - Entscheidungsgrenze wird in ca. 1000 Epochen berechnet
 - Datensatz wird recht gut klassifiziert es gibt aber noch einige Punkte, die genau auf der grenze liegen oder falsch klassifiziert werden
- Sigmoid:
 - Entscheidungsgrenze versucht der Spiralform näher zu kommen, es passt aber nicht wirklich
 - Trainings und Testverluste liegen bei 0.221 und 0.344. Überanpassung ist nicht vorhanden. Die Verlust werte werden über die Zeit immer besser
 - Entscheidungsgrenze wird in ca. 10000 Epochen berechnet, also sehr sehr langsam
 - Datensatz wird recht schlecht klassifiziert es gibt eine große grenze wo viele Punkte als 0 klassifiziert werden
- Tanh:
 - Die Entscheidungsgrenze nähert sich langsam der Spiral Form, passt aber auch hier nicht perfekt
 - Trainings und Testverluste liegen bei 0.192 und 0.049. Überanpassung ist nicht vorhanden

- Entscheidungsgrenze wird in ca. 5000 Epochen berechnet. Ab ca. 3000 fängt Sie an stark zu schwanken, beruhigt sich aber nach 5000 Epochen
- Datensatz wird mittelmäßig gut klassifiziert es gibt viele Punkte die Falsch klassifiziert sind

Spiral Datensatz: drei versteckte Schichten mit jeweils 7 Neuronen:

- ReLU:
 - Entscheidungsgrenze nähert sich weiter der Spiralform
 - Trainings und Testverluste liegen bei 0.056 und 0.008. Überanpassung ist nicht vorhanden
 - Entscheidungsgrenze wird in ca. 400 Epochen berechnet.
 - Datensatz wird recht gut klassifiziert es gibt trotzdem in den Testdaten einige die Falsch klassifiziert sind da die Entscheidungsgrenze keine durchgängige Spirale ist
- Sigmoid:
 - Entscheidungsgrenze hier wieder eine Gerade
 - Trainings und Testverluste liegen bei 0.467 und 0.502. Überanpassung ist nicht vorhanden
 - Entscheidungsgrenze wird in ca. 500 Epochen berechnet.
 - Datensatz wird schlecht klassifiziert, Gerade kann die Spirale nicht gut trennen
- Tanh:
 - Entscheidungsgrenze nähert sich sehr wenig der Spirale an
 - Trainings und Testverluste liegen bei 0.058 und 0.151. Überanpassung ist nicht vorhanden
 - Entscheidungsgrenze wird in ca. 1000 Epochen berechnet.
 - Datensatz wird ganz okay klassifiziert, es werden aber viele noch falsch klassifiziert

Spiral Datensatz: vier versteckte Schichten mit jeweils sieben Neuronen

- ReLU:
 - Entscheidungsgrenze ähnelt sehr einer Spiralform
 - Trainings und Testverluste liegen bei 0.032 und 0.015. Überanpassung ist nicht vorhanden
 - Entscheidungsgrenze wird in ca. 1000 Epochen berechnet. Sie schwankt viel nach ca. 500 Epochen
 - Datensatz wird sehr gut klassifiziert bis jetzt am besten. Es gibt wenige die noch falsch klassifiziert sind
- Sigmoid:
 - Entscheidungsgrenze ist eine fallende Gerade
 - Trainings und Testverluste liegen bei 0.467 und 0.504. Überanpassung ist nicht vorhanden
 - Entscheidungsgrenze wird in ca. 1500 Epochen berechnet.
 - Datensatz wird schlecht klassifiziert wegen Gerade
- Tanh:
 - Entscheidungsgrenze nähert sich einer Spiralform

- Trainings und Testverluste liegen bei 0.009 und 0.026. Überanpassung ist nicht vorhanden
- Entscheidungsgrenze wird in ca. 2000 Epochen berechnet.
- Datensatz wird sehr gut klassifiziert und von allem Kombination für diese Daten am besten. Ist aber nicht perfekt ein paar falsch klassifizierten Punkte gibt es noch

Die ReLU-Funktion hat meistens eine eher eckige Entscheidungsgrenze, während die anderen beiden Aktivierungsfunktionen mehr abgerundete Formen haben. Die ReLU hat meistens ganz gute Werte und berechnet die Entscheidungsgrenze auch recht schnell. Die Sigmoidfunktion scheint am langsamsten zu sein und bringt meistens eher schlechtere Verlustwerte als die beiden anderen. Die Tanh scheint auch sehr gut zu sein und bringt zu mindestens bei dem Spiral Datensatz die besten Verlustwerte.

3.

Es kann von einer Überanpassung gesprochen werden, wenn die Testverlustwerte ansteigen, obwohl die Trainingsverlustwerte weiter sinken und sozusagen ‚besser‘ werden. Das Neuronale Netz wird aber nicht wirklich besser, sondern fängt an die Trainingsdatenpunkte auswendig zu lernen