1. (2 Punkte) Ein Netz gibt die Wahrscheinlichkeit $\hat{y}$ für die Klasse 1 aus. Berechne jeweils die Wahrscheinlichkeit p, die das Netz der richtigen Klasse y gibt, und den Fehler $L = -\ln(p)$.
   (a) $\hat{y} = 0.95$, $y = 1$
   (b) $\hat{y} = 0.6$, $y = 0$
   (c) $\hat{y} = 0.02$, $y = 1$

2. (4 Punkte) Zwei Netze A und B sollen vier Bilder klassifizieren. Die richtigen Klassen und die Ausgaben $\hat{y}$ der Netze sind

Bild	1	2	3	4
richtige Klasse y	1	0	1	0
$\hat{y}$ von Netz A	0.99	0.02	0.97	0.9
$\hat{y}$ von Netz B	0.8	0.25	0.75	0.3

   (a) Bestimme für jedes Bild die Wahrscheinlichkeit p, die das Netz der richtigen Klasse gibt.
   (b) Berechne für beide Netze den mittleren Abstand $1 - p$ und das Produkt der Wahrscheinlichkeiten p.
   (c) Berechne für beide Netze den mittleren Fehler $-\ln(p)$.
   (d) Welches Netz ist besser? Warum kommt der mittlere Abstand zu einem anderen Ergebnis?

3. (4 Punkte) Der Fehler eines Netzes für ein Beispiel ist $L = -\big(y \ln(\hat{y}) + (1-y)\ln(1-\hat{y})\big)$.
   (a) Vereinfache die Formel für $y = 1$ und für $y = 0$. Was gibt der Fehler in Worten an?
   (b) Bei mehreren Beispielen multipliziert man die Wahrscheinlichkeiten für die richtigen Klassen. Warum?
   (c) Warum nimmt man den Logarithmus und warum das Minus davor?
   (d) Warum ist $-\ln(p)$ als Fehler besser geeignet als der Abstand $1 - p$?

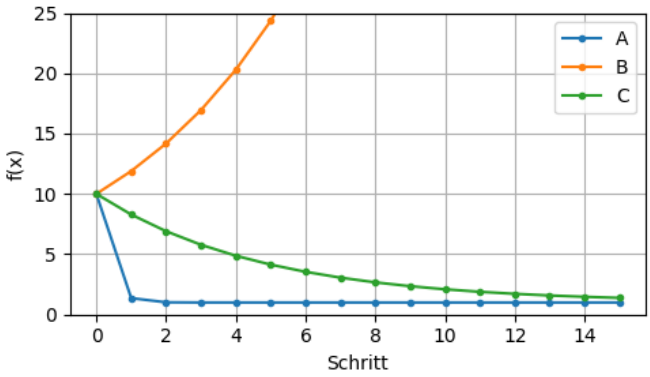
4. (3 Punkte) Ein Neuron mit einem Input hat $z = w \cdot x + b$ und $\hat{y} = \sigma(z)$. Für das Trainingsbeispiel $x = 1$, $y = 1$ und $b = 0$ ist der Fehler $L(w) = -\ln(\sigma(w))$. Am Anfang ist $w = 0$.
   (a) Berechne $L(0)$, $L(0.01)$ und $L(-0.01)$.
   (b) Berechne die numerische Ableitung $\dfrac{L(0.01) - L(-0.01)}{0.02}$. Soll w größer oder kleiner werden?
   (c) Mache einen Schritt des Gradientenabstiegs mit $lr = 1$ und berechne den neuen Fehler.

5. (4 Punkte) Die Funktion $f(x) = (x-3)^2 + 1$ hat die Ableitung $f'(x) = 2(x-3)$. Der Gradientenabstieg startet bei $x = 0$ und rechnet $x_{neu} = x - lr \cdot f'(x)$.
   (a) Berechne drei Schritte mit $lr = 0.25$.
   (b) Berechne zwei Schritte mit $lr = 1$ und zwei Schritte mit $lr = 1.25$.
   (c) Beschreibe, was in den drei Fällen passiert.

6. (3 Punkte) Gegeben ist $f(x_1, x_2) = x_1^2 + x_2^2$ an der Stelle $(2, 1)$.
   (a) Berechne die partiellen Ableitungen und den Gradienten an dieser Stelle. Welcher Input hat dort den größeren Einfluss auf f?
   (b) Gehe mit $lr = 0.25$ einen Schritt gegen den Gradienten. Wie groß ist f vorher und nachher?

7. (3 Punkte) Gegeben ist $f(a, b) = a^2 + 2b^2$. Der Gradientenabstieg startet bei $(a, b) = (2, 1)$ mit $lr = 0.1$.
   (a) Bestimme die partiellen Ableitungen $\frac{\partial f}{\partial a}$ und $\frac{\partial f}{\partial b}$.
   (b) Berechne zwei Schritte und jeweils den Funktionswert.

8. (3 Punkte) Für $f(x) = (x-3)^2 + 1$ wurde der Gradientenabstieg von $x = 0$ aus mit drei Lernraten durchgeführt: 0.05, 0.4 und 1.05. Das Bild zeigt den Funktionswert nach jedem Schritt.



(a) Ordne die Kurven A, B und C den Lernraten zu und begründe.

(b) Was ist beim Training eines neuronalen Netzes die Folge einer zu kleinen und einer zu großen Lernrate?

9. (4 Punkte) (a) Was gibt die Ableitung des Fehlers nach einem Parameter an?

(b) Was ist der Gradient und in welche Richtung zeigt er?

(c) Was ist eine Epoche?

(d) Was ist ein lokales Minimum und warum ist es ein Problem für den Gradientenabstieg?

10. (4 Punkte) Schreibe eine Funktion `abstieg(f, x, lr, anzahl)`, die für eine beliebige Funktion `f` von der Stelle `x` aus `anzahl` Schritte des Gradientenabstiegs macht und das Ende zurückgibt. Die Ableitung soll numerisch berechnet werden. Teste sie mit $g(x) = x^4 - 3x^2 + x$, $lr = 0.01$ und 200 Schritten, einmal ab $x = 2$ und einmal ab $x = -2$. Was fällt auf?