Gradient:

$$\text{grad}(f) = \frac{\partial f}{\partial x_1} + \ldots + \frac{\partial f}{\partial x_n} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ + \\ \vdots \\ + \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Sammlung aller partiellen Ableitungen.

⇒ Zeigt in die Richtung der größten Steigung.

$$\text{grad}(f)(x,y) = 2x^2 - y^2 = \begin{matrix} x: 4x \\ y: -2y \end{matrix} \Rightarrow \nabla f = \begin{pmatrix} 4x \\ -2y \end{pmatrix}$$

Richtungsableitung:

$$D_v f(X_0) = \lim_{h \to 0} \frac{f(x_0 + h \cdot v) - f(x_0)}{h}$$

Lokale Änderungsrate im Punkt

$$f(x,y) = x^2 + 3xy^2 \qquad X_0 = (1,1) \qquad v = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

$$\nabla(f)(x,y) = \begin{pmatrix} 2x + 3y^2 \\ 6xy \end{pmatrix}$$

$$\nabla(f)(1,1) = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$$

$$\nabla(f)(1,1)^t \cdot v = \begin{pmatrix} 5 \\ 6 \end{pmatrix}^t \cdot \begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

Differentialgleichung:

$$x_1(t)' = x_1(t) + 12x_2(t)$$
$$x_2(t)' = 3x_1(t) + x_2(t)$$

Aus einer Ableitung eine Funktion bilden

$$(A - \lambda I) = 0 \Leftrightarrow Av = \lambda v: \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} 1 & 12 \\ 3 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$$

$$\det \begin{pmatrix} 1-\lambda & 12 \\ 3 & 1-\lambda \end{pmatrix} = (1-\lambda) \cdot (1-\lambda) - 3 \cdot 12 = \lambda^2 - \underset{p}{2} \cdot \lambda \underset{q}{- 35}$$

$$-\frac{-2}{2} \pm \sqrt{\left(\frac{-2}{2}\right)^2 - (-35)} = \begin{matrix} \lambda_1 = 7 \\ \lambda_2 = -5 \end{matrix}$$

$$\lambda_1 = \begin{pmatrix} 1-7 & 12 \\ 3 & 1-7 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \left(\begin{array}{c|c} -6v_1 + 12v_2 & 0 \\ 3v_1 - 6v_2 & 0 \end{array}\right) \quad | \cdot 2$$

$$\left(\begin{array}{c|c} -6v_1 + 12v_2 & 0 \\ 0 + 0 & 0 \end{array}\right)$$

$$\Rightarrow -6v_1 + 12v_2 = 0$$

$$x_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$\lambda_2 = \begin{pmatrix} 1+5 & 12 \\ 3 & 1+5 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \left(\begin{array}{c|c} 6v_1 + 12v_2 & 0 \\ 3v_1 + 6v_2 & 0 \end{array}\right) \quad | \cdot (-2)$$

$$\left(\begin{array}{c|c} 6v_1 + 12v_2 & 0 \\ 0 + 0 & 0 \end{array}\right)$$

$$\Rightarrow 6v_1 + 12v_2 = 0$$

$$x_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

$$x(t) = c_1 \cdot e^{7t} \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} + c_2 \cdot e^{-5t} \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

Kettenregel:

Eine Verkettung differenzierbarer Funktionen lässt sich als einzelne differenzierbare Funktionen zusammensetzen.

$$h(x) = f(g(x)) \Rightarrow h'(x) = f'(g(x)) \cdot g'(x)$$

Eine Kompisition zweier Funktionen lässt sich also so ausdrücken:

$g: \mathbb{R} \to \mathbb{R}^2$
$f: \mathbb{R}^2 \to \mathbb{R}$

$$df(x,y) = \left[\frac{\partial}{\partial x} f(x,y) \ \frac{\partial}{\partial x} f(x,y)\right]$$

$$\frac{d}{dx}(f \circ g)(x) = df(g(x)) \cdot dg(x)$$

$\gamma: [0, 2\pi] \to \mathbb{R}^2$

$$\gamma(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} \qquad \gamma'(t) = \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix}$$

$f: \mathbb{R}^2 \to \mathbb{R}$

$$f(x,y) = xy \qquad f'(x,y) = \begin{pmatrix} y \\ x \end{pmatrix}$$

$$\frac{d}{dt}(f \circ \gamma)(t) = (\sin(t), \cos(t)) \cdot \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix}$$

$$= \sin(t) \cdot (-\sin(t)) + \cos(t) \cdot \cos(t)$$

$$= -\sin(t)^2 + \cos(t)^2$$

Extrema:

$$f(x,y) = x^2 + y^2$$

1. Gradient bilden:

$$\nabla f(x,y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$$

2. Kritischen Punkt bestimmen:

$$\begin{array}{l} \text{I} \ \ 2x = 0 \\ \text{II} \ \ 2y = 0 \end{array} \Rightarrow P = (0,0)$$

3. Hessematrix bestimmen

$f_{xx}: \ \frac{\partial f}{\partial x}(2x) = 2$
$f_{yy}: \ \frac{\partial f}{\partial y}(2y) = 2$
$f_{xy} = f_{yx} = \frac{\partial f}{\partial y}(2) = 0$

$$H_f(x,y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

4. Determinante von $P(0,0)$ berechnen:

$$\det\begin{pmatrix} 2-\lambda & 0 \\ 0 & 2-\lambda \end{pmatrix} = (2-\lambda) \cdot (2-\lambda) = \lambda^2 \underbrace{-4}_{p} \cdot \lambda \underbrace{+4}_{q}$$

$$-\frac{-4}{2} \pm \sqrt{\left(\frac{-4}{2}\right)^2 - 4} = 2$$

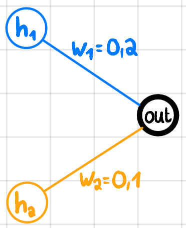
$> 0 \Rightarrow$ Minimum
$< 0 \Rightarrow$ Maximum

Gradient Descent:

$$b = a - \gamma \nabla f(a)$$

Backpropagation:

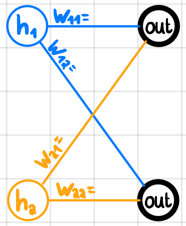
- Wie stark trägt ein Gewicht zum Fehler bei?



$e = 0{,}3$

$$e_{h1} = \frac{w_1}{w_1+w_2} \cdot e_{out}$$

$$e_{h2} = \frac{w_2}{w_1+w_2} \cdot e_{out}$$



$e_1 =$

$e_2 =$

$$\left.\begin{aligned} e_{h1} &= \frac{w_{11}}{w_{11}+w_{21}} \cdot e_1 + \frac{w_{12}}{w_{12}+w_{22}} \cdot e_2 \\ e_{h2} &= \frac{w_{21}}{w_{11}+w_{21}} \cdot e_1 + \frac{w_{22}}{w_{12}+w_{22}} \cdot e_2 \end{aligned}\right\} \sum_{i,j}^{k} \left(\frac{w_{ij}}{w_{ii}+w_{jj}} \cdot e_k \right) \hat{=} L_D(\omega) = \sum_{x_i, y_i} \left(f(\omega, x_i) - y_i \right)^2$$

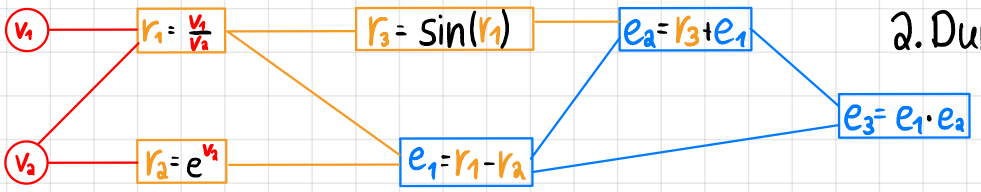
Algorithmus:
1. Setze $k=0$ und Gewichte ω_0
2. Setze Genauigkeit $\varepsilon > 0$
3. Solange $\|\nabla L_D(\omega)\| > \varepsilon$ sind:
   Berechne $d_k = -\nabla L_D(\omega_k)$ und Schrittweite α_k
   Setze $\omega_{k+1} = \omega_k + \alpha_k d_k$
4. $k = k+1$

3. Nur lokale Extrema / α_k schwer zu wählen

Automatisches Differenzieren:

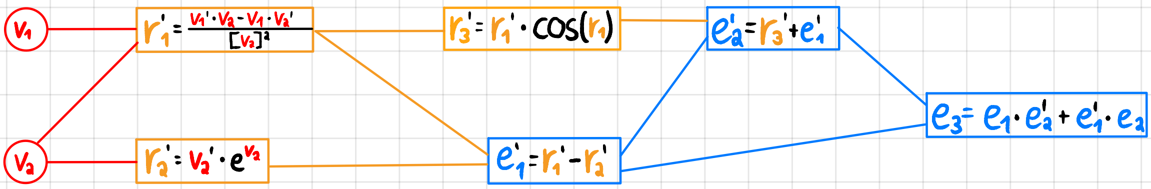
$$f(x_1,x_2) = \left[\underbrace{\boxed{\sin\left(\underbrace{\boxed{\tfrac{x_1}{x_2}}}_{r_1}\right)}}_{r_3} + \underbrace{\boxed{\tfrac{x_1}{x_2}}}_{r_1} - \underbrace{\boxed{e^{x_2}}}_{r_2} \right] \cdot \left[\underbrace{\boxed{\tfrac{x_1}{x_2}}}_{r_1} - \underbrace{\boxed{e^{x_2}}}_{r_2} \right]$$

1. Variable als Funktion: $x_1 = v_1(t)$
   $x_2 = v_2(t)$



2. Durch Kettenregel einzeln ableitbar:

$$\frac{\delta f}{\delta x} = \frac{\delta f}{\delta a} \cdot \frac{\delta a}{\delta x}$$



Eigenwert: $\det(A-\lambda E)=0 \quad \Rightarrow c_1 \cdot e^{\lambda t} \cdot \begin{pmatrix} w_1 \end{pmatrix}$

Eigenvektor: $(A-\lambda E)\cdot \vec{v}=0$

Hauptvektor Stufe 2: $(A-\lambda I)\cdot h_2 = v_1 \quad | \quad (A-I)^2 \cdot v = 0$