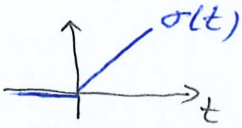
Neuronät (= neuralt nätverk)

= flervariabelfunktion $f(\bar x)$ som är en sammansättning av flera enklare funktioner: $f(\bar x) = (f_L \circ \dots \circ \bar f_2 \circ \bar f_1)(\bar x) = f_L(\dots \bar f_2(\bar f_1(\bar x))\dots)$. L = antal lager.
Varje $\bar f_j$ är i sig en sammansättning av en linjär eller affin avbildning $\underbrace{\underbrace{A_j \bar x}_{\text{linjär}} + \bar b_j}_{\text{affin}}$ och en enkel olinjär $\bar\sigma$.

Låt $\sigma(t) = (t)_+ = \max(t,0) = \begin{cases} t, & t \geq 0 \\ 0, & t < 0 \end{cases}$



och sätt

$$\bar\sigma(\bar t) = \begin{pmatrix} \sigma_1(t_1,\dots,t_n) \\ \vdots \\ \sigma_n(t_1,\dots,t_n) \end{pmatrix} \overset{\text{def}}{=} \begin{pmatrix} \sigma(t_1) \\ \vdots \\ \sigma(t_n) \end{pmatrix} = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}_+ \qquad \text{Ex } \bar\sigma\left(\begin{pmatrix} -1 \\ 2 \\ -3 \\ 4 \end{pmatrix}\right) = \begin{pmatrix} -1 \\ 2 \\ -3 \\ 4 \end{pmatrix}_+ = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 4 \end{pmatrix}$$

Ex Litet neuronnät $y = f(\bar x) = (f_3 \circ \bar f_2 \circ \bar f_1)(\bar x)$ $(L=3)$

$A_1 = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}$, $\bar b_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$, $A_2 = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ -3 & 1 \end{pmatrix}$, $A_3 = (1, -1, 1)$

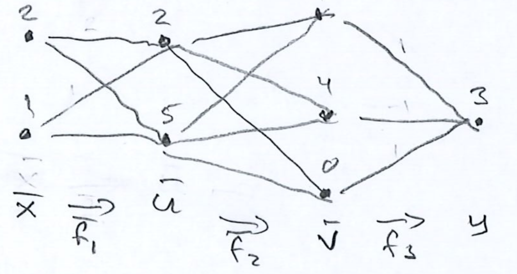
$\bar f_1(\bar x) = (A_1 \bar x + \bar b_1)_+$, $\bar f_2(\bar u) = (A_2 \bar u)_+$, $f_3(\bar v) = A_3 \bar v$

För t.ex. $\bar x = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \in \mathbb{R}^2$ fås $\bar u = \bar f_1(\bar x) = \left(A_1 \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \bar b_1\right)_+ = \begin{pmatrix} 2 \\ 5 \end{pmatrix}_+ = \begin{pmatrix} 2 \\ 5 \end{pmatrix} \Rightarrow$

$\bar v = \bar f_2(\bar u) = \bar f_2(\bar f_1(\bar x)) = \left(A_2 \begin{pmatrix} 2 \\ 5 \end{pmatrix}\right)_+ = \begin{pmatrix} 7 \\ 4 \\ -1 \end{pmatrix}_+ = \begin{pmatrix} 7 \\ 4 \\ 0 \end{pmatrix} \Rightarrow$

$y = f_3(\bar v) = f_3(\bar f_2(\bar f_1(\bar x))) = A_3 \bar v = (1, -1, 1)\begin{pmatrix} 7 \\ 4 \\ 0 \end{pmatrix} = 3$, så $f\left(\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right) = 3$

Grafisk repr.



Hur hitta bra neuronnät för viss tillämpning? Utnyttja att man har träningsdata, för visst antal givna $\bar x$ vet man vad $y = f(\bar x)$ ska bli (övervakad inlärning). Man vill bestämma alla matriser och vektorer i nätet så att felet

$$\ell = \underbrace{\frac{1}{N}}_{\text{[kunde skippas]}} \sum_{j=1}^{N} \Big(\underbrace{f(\overset{\text{träningspunkt}}{\bar x^{(j)}})}_{\text{utvärde som nätet ger då } \bar x^{(j)} \text{ matas in}} - \overset{\text{önskat utvärde då } \bar x^{(j)} \text{ matas in}}{y^{(j)}} \Big)^2$$

minimeras (minstakvadratproblem)

N = antal träningspunkter

[andra ℓ kan väljas]

Ex (mikronät) Skapa ett nät $f(\bar x) = (f_2 \circ \bar f_1)(\bar x)$ för $\bar x \in \mathbb{R}^2$ där
$\bar u = \bar f_1(\bar x) = (A_1 \bar x + \bar b_1)_+ \in \mathbb{R}^2$, och $y = f_2(\bar u) = A_2 \bar u \in \mathbb{R}$, så
$A_1 = \begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \end{pmatrix}$ 2×2, $\bar b_1 = \begin{pmatrix} w_5 \\ w_6 \end{pmatrix}$ och $A_2 = (w_7 \ w_8)$ söks.

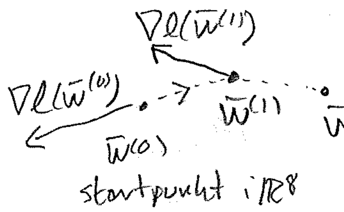
Träningsdata är endast 2 punkter, $\hat x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ in ska ge $\hat y = 1$ ut, $\tilde x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ ska ge $\tilde y = 5$.

Bestäm $\bar w = (w_1, \dots, w_8) \in \mathbb{R}^8$ så att

$$l(\bar w) = \frac{1}{2}\left((\underbrace{f(\hat x; \bar w)}_{\text{nytt skrivsätt för att visa att nätet } f \text{ också beror på matriserna}} - \hat y)^2 + (f(\tilde x; \bar w) - \tilde y)^2 \right) = \underbrace{\frac{1}{2}\left(\left(f\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \bar w\right) - 1\right)^2 + \left(f\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \bar w\right) - 5\right)^2 \right)}_{\text{minimeras}}$$

Min finns där $\nabla l(\bar w) = (l'_{w_1}, \dots, l'_{w_8}) = \bar 0$ [nästa föreläsn.] men försvårt lösa exakt eller numeriskt ($\bar w$ i mycket hög dim.).

Idé: $l(\bar w)$ avtar snabbast i riktn. $-\nabla l(\bar w)$. Ta steg i denna riktning för att minska $l(\bar w)$ [brantaste lutning i opt.].



etc. $l(\bar w^{(k)})$ bör minska då k ökar

$$\nabla(g^2) = 2g \nabla g \Rightarrow \nabla l(\bar w) = (f(\hat x, \bar w) - \hat y) \overbrace{\underbrace{\nabla_{\bar w}}_{\substack{\text{grad. på } \bar w \\ \text{(inte på } \bar x)}} f(\hat x, \bar w)}^{\text{Hur beräkna??}} + (f(\tilde x, \bar w) - \tilde y) \nabla_{\bar w} f(\tilde x, \bar w)$$

$(\nabla \hat y = \nabla \tilde y = \bar 0)$

Kedjeregeln för komponent u_k i $\bar u = \begin{pmatrix} u_1(\overbrace{f_1(t_1, \dots, t_q)}^{x_1}, \dots, \overbrace{f_n(t_1, \dots, t_q)}^{x_n}) \\ \vdots \\ u_m(\ - \ - \ - \ - \ - \ - \ ) \end{pmatrix}$

placera på rad k i matris

är
$$\underbrace{\frac{\partial u_k}{\partial t_j}}_{\substack{\text{hamnar då på} \\ \text{plats } k,j \text{ i matris}}} = \frac{\partial u_k}{\partial x_1} \frac{\partial x_1}{\partial t_j} + \dots + \frac{\partial u_k}{\partial x_n} \frac{\partial x_n}{\partial t_j} \Rightarrow$$

placera i kolonn j i matris

$$\underbrace{\begin{pmatrix} \frac{\partial u_1}{\partial t_1} & \cdots & \frac{\partial u_1}{\partial t_q} \\ \vdots & & \vdots \\ \frac{\partial u_m}{\partial t_1} & \cdots & \frac{\partial u_m}{\partial t_q} \end{pmatrix}}_{m \times q} = \underbrace{\begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \cdots & \frac{\partial u_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial u_m}{\partial x_1} & \cdots & \frac{\partial u_m}{\partial x_n} \end{pmatrix}}_{m \times n} \underbrace{\begin{pmatrix} \frac{\partial x_1}{\partial t_1} & \cdots & \frac{\partial x_1}{\partial t_q} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial t_1} & \cdots & \frac{\partial x_n}{\partial t_q} \end{pmatrix}}_{n \times q} \Rightarrow \frac{\partial(\bar u)}{\partial(\bar t)} = \frac{\partial(\bar u)}{\partial(\bar x)} \frac{\partial(\bar x)}{\partial(\bar t)}$$

mest allmänna kedjeregeln

$$\underset{\substack{\uparrow \\ \text{beteckn.}}}{=} \frac{\partial(u_1, \dots, u_m)}{\partial(t_1, \dots, t_q)} = \frac{\partial(\bar u)}{\partial(\bar t)} = \bar u'(t) \qquad = \frac{\partial(\bar u)}{\partial(\bar x)} \qquad = \frac{\partial(\bar x)}{\partial(\bar t)}$$

Matriserna kallas funktionalmatriser.

Ex $\bar{f}(\bar{x};\bar{w}) = \underbrace{\begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{\bar{x}} + \underbrace{\begin{pmatrix} w_5 \\ w_6 \end{pmatrix}}_{\bar{b}} = \begin{pmatrix} w_1x_1 + w_2x_2 + w_5 \\ w_3x_1 + w_4x_2 + w_6 \end{pmatrix}$

$\Rightarrow \dfrac{\partial(\bar{f})}{\partial(\bar{x})} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \end{pmatrix} = A$ och $\dfrac{\partial(\bar{f})}{\partial(\bar{w})} = \begin{pmatrix} \frac{\partial f_1}{\partial w_1} & \cdots & \frac{\partial f_1}{\partial w_6} \\ \frac{\partial f_2}{\partial w_1} & \cdots & \frac{\partial f_2}{\partial w_6} \end{pmatrix} = \begin{pmatrix} \overbrace{x_1 \; x_2 \; 0 \; 0}^{\text{från } A\bar{x}} & \overbrace{1 \; 0}^{\text{från } \bar{b}} \\ 0 \; 0 \; x_1 \; x_2 & 0 \; 1 \end{pmatrix} = M(\bar{x})$

$\sigma(t) = (t)_+$ har $\sigma'(t) = H(t) = \begin{cases} 1 & t>0 \\ 0 & t<0 \end{cases} \Rightarrow \bar{\sigma}(\bar{t}) = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}_+$ har $\dfrac{\partial(\bar{\sigma})}{\partial(\bar{t})} = \begin{pmatrix} H(t_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & H(t_n) \end{pmatrix}$

diagonal, 0/1 på diagonalen

Åter till mikronätet:

Söker $\nabla_{\bar{w}} f(\bar{x};\bar{w}) = \left(\dfrac{\partial y}{\partial w_1}, \ldots, \dfrac{\partial y}{\partial w_8}\right)$

$y = A_2\bar{u} = (w_7 \; w_8)\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = w_7u_1 + w_8u_2 \Rightarrow \dfrac{\partial(y)}{\partial(w_7, w_8)} = \left(\dfrac{\partial y}{\partial w_7}, \dfrac{\partial y}{\partial w_8}\right) = (u_1, u_2) = \bar{u}^T$

Låt $\bar{t} = A_1\bar{x} + \bar{b}_1$ så $\bar{u} = (\bar{t})_+ \Rightarrow \dfrac{\partial(y)}{\partial(w_1, \ldots, w_6)} \underset{\substack{\uparrow \\ \text{kedjeregeln} \\ \text{2 gånger}}}{=} \underbrace{\dfrac{\partial(y)}{\partial(\bar{u})}}_{\substack{=A_2 \\ 1\times 2}} \underbrace{\dfrac{\partial(\bar{u})}{\partial(\bar{t})}}_{\substack{=H(\bar{t}) \\ 2\times 2}} \underbrace{\dfrac{\partial(\bar{t})}{\partial(w_1, \ldots, w_6)}}_{\substack{=M(\bar{x}) \\ 2\times 6}} \Rightarrow 1\times 6$ matris

$\Rightarrow \nabla_{\bar{w}} f(\bar{x},\bar{w}) = (\underbrace{A_2 H(\bar{t}) M(\bar{x})}_{6 \text{ komp.}}, \underbrace{\bar{u}^T}_{2 \text{ komp.}})$

Välj ett startnät (startpunkt i $\mathbb{R}^8$), t.ex. $\bar{w}^{(0)} = (\underbrace{1,-1,2,1}_{A_1}, \underbrace{1,-2}_{\bar{b}_1}, \underbrace{2,1}_{A_2})$

så $A_1 = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}$, $\bar{b}_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$, $A_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ till att börja med

Förbättra nätet genom att stega längs $-\nabla l(\bar{w})$. Kör igenom träningspunkterna:

$\hat{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow \hat{t} = A_1\hat{x} + \hat{b}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\hat{u} = (\hat{t})_+ = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $f(\hat{x}, \bar{w}^{(0)}) = A_2\hat{u} = 3$, jämför önskat $\hat{y} = 1$

Notera $H(\hat{t}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $M(\hat{x}) = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$ kan läsas av

$\tilde{x} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \Rightarrow \tilde{t} = A_1\tilde{x} + \tilde{b}_1 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$, $\tilde{u} = (\tilde{t})_+ = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$, $f(\tilde{x}, \bar{w}^{(0)}) = A_2\tilde{u} = 6$, jämför $\hat{y} = 5$

Notera $H(\tilde{t}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $M(\tilde{x}) = \begin{pmatrix} 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{pmatrix}$

Detta ger $l(\bar{w}^{(0)}) = \frac{1}{2}\left((3-1)^2 + (6-5)^2\right) = 2.5$ fel med startnät, och

$\nabla l(\bar{w}^{(0)}) = (3-1)\cdot\underbrace{(A_2 H(\hat{t}) M(\hat{x}), \hat{u}^T)}_{=(2,2,1,1,2,1,1,1)} + (6-5)\underbrace{(A_2 H(\tilde{t}) M(\tilde{x}), \tilde{u}^T)}_{(2,-2,0,0,2,0,3,0)} =$

$= (6,2,2,2,6,2,5,2)$

Flytta till nästa punkt (nytt nät) $\bar{w}^{(1)} = \bar{w}^{(0)} - \eta \nabla l(\bar{w}^{(0)}) =$

$= (\underbrace{0.7, -1.1, 1.9, 0.9}_{\text{ny } A_1}, \underbrace{0.7, -2.1}_{\text{ny } \bar{b}}, \underbrace{1.75, 0.9}_{\text{ny } A_2})$

Tag $\eta = 0.05$, svårt veta bästa steglängd

Nytt nät $A_1 = \begin{pmatrix} 0{,}7 & -1{,}1 \\ 1{,}9 & 0{,}9 \end{pmatrix}$, $\bar{b}_1 = \begin{pmatrix} 0{,}7 \\ -2{,}1 \end{pmatrix}$, $A_2 = (1{,}75 \ \ 0{,}9)$

Samma steg som ovan

$\hat{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ger $\hat{t} = \begin{pmatrix} 0{,}3 \\ 0{,}7 \end{pmatrix}$, $\hat{u} = \begin{pmatrix} 0{,}3 \\ 0{,}7 \end{pmatrix}$, $f(\hat{x}, \bar{w}^{(1)}) = 1.155$ H- och M-matriser som ovan

$\tilde{x} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ $\tilde{t} = \begin{pmatrix} 2{,}5 \\ -1{,}1 \end{pmatrix}$, $\tilde{u} = \begin{pmatrix} 2{,}5 \\ 0 \end{pmatrix}$, $f(\tilde{x}, \bar{w}^{(1)}) = 4{,}375$

Ger $\ell(\bar{w}^{(1)}) \approx 0{,}21$ stor förbättring

$\nabla \ell(\bar{w}^{(1)}) \approx (\dots)$

$\bar{w}^{(2)} = \bar{w}^{(1)} - 0{,}05 \nabla \ell(\bar{w}^{(1)}) \approx (0.74, \dots, 0{,}89)$ $\Rightarrow$ nytt nät har

$A_1 \approx \begin{pmatrix} 0{,}74 & -1.17 \\ 1{,}89 & 0.89 \end{pmatrix}$, $\bar{b}_1 \approx \begin{pmatrix} 0{,}74 \\ -2{,}11 \end{pmatrix}$, $A_2 \approx (1{,}83, 0{,}89)$

med $\ell(\bar{w}^{(2)}) \approx 0.09$ ny förbättring

Fortsätt till $\ell(\bar{w}^{(k)})$ acceptabelt litet. Då har man sitt färdiga neuronnät som kan användas för att studera $y = f(\bar{x})$ för $\bar{x}$ med okända y-värden. Man har tränat eller lärt upp nätet.

Allt enkelt koda. Användning av kedjeregeln kallas inom AI för "back propagation algorithm". Djupare nät kräver kedjeregeln i fler steg: $A_L H_{L-1} A_{L-1} \dots A_{j+1} H_j M_j$ för derivator på matriselement i lager j, $j \geq 1$. Vår σ kallas ReLU i AI.
Många fler aspekter, tillämpningar, problem, terminologi etc fås på utblick 27/4 10-12