Kompletterande problem, TATA83 Flervariabelanalys

K1. Bestäm funktionalmatriserna av $\bar{f} : \mathbb{R}^2 \to \mathbb{R}^2$, $\bar{g} : \mathbb{R}^6 \to \mathbb{R}^2$ och $\bar{h} : \mathbb{R}^3 \to \mathbb{R}^3$ som ges av

$$\bar{f}(\bar{x}) = \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 + 5 \\ 3x_1 + 4x_2 + 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 5 \\ 6 \end{pmatrix}$$

$$\bar{g}(\bar{w}) = \begin{pmatrix} g_1(w_1, w_2, w_3, w_4, w_5, w_6) \\ g_2(w_1, w_2, w_3, w_4, w_5, w_6) \end{pmatrix} = \begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \end{pmatrix} \begin{pmatrix} 7 \\ 8 \end{pmatrix} + \begin{pmatrix} w_5 \\ w_6 \end{pmatrix}$$

$$\bar{h}(\bar{t}) = \begin{pmatrix} h_1(t_1, t_2, t_3) \\ h_2(t_1, t_2, t_3) \\ h_3(t_1, t_2, t_3) \end{pmatrix} = \begin{pmatrix} e^{t_1} + t_3^2 \\ \sin t_2 \\ \ln t_3 \end{pmatrix}$$

K2. Antag att

$$\bar{t} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = \bar{f}(\bar{x}, w_1, w_2, w_3, w_4, w_5, w_6) = \begin{pmatrix} f_1(x_1, x_2, w_1, w_2, w_3, w_4, w_5, w_6) \\ f_2(x_1, x_2, w_1, w_2, w_3, w_4, w_5, w_6) \\ f_3(x_1, x_2, w_1, w_2, w_3, w_4, w_5, w_6) \end{pmatrix} = \begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \\ w_5 & w_6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\bar{u} = \bar{\sigma}(\bar{t}) = \begin{pmatrix} \sigma(t_1) \\ \sigma(t_2) \\ \sigma(t_3) \end{pmatrix}, \text{ där } \sigma(t) = \max(t, 0)$$

$$y = g(\bar{u}, w_7, w_8, w_9, w_{10}) = g(u_1, u_2, u_3, w_7, w_8, w_9, w_{10}) = (w_7, w_8, w_9) \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + w_{10}$$

a) Ange de fem funktionalmatriserna $\dfrac{\partial(\bar{t})}{\partial(\bar{x})} = \dfrac{\partial(t_1, t_2, t_3)}{\partial(x_1, x_2)}$,

$$\frac{\partial(\bar{t})}{\partial(w_1, w_2, w_3, w_4, w_5, w_6)} = \frac{\partial(t_1, t_2, t_3)}{\partial(w_1, w_2, w_3, w_4, w_5, w_6)}, \quad \frac{\partial(\bar{u})}{\partial(\bar{t})} = \frac{\partial(u_1, u_2, u_3)}{\partial(t_1, t_2, t_3)},$$

$$\frac{\partial(y)}{\partial(\bar{u})} = \frac{\partial(y)}{\partial(u_1, u_2, u_3)} \quad \text{och} \quad \frac{\partial(y)}{\partial(w_7, w_8, w_9, w_{10})}$$

b) Om $\begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \\ w_5 & w_6 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -3 & 1 \\ 2 & 3 \end{pmatrix}$, $(w_7, w_8, w_9) = (-1, 2, 1)$, $w_{10} = 3$ och $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$,

beräkna $\bar{t}$, $\bar{u}$ och y, samt de fem funktionalmatriserna ovan.

c) Med värdena i b), använd kedjeregeln för att beräkna funktionalmatrisen

$$\frac{\partial(y)}{\partial(w_1, w_2, w_3, w_4, w_5, w_6)} \quad \text{samt ange} \quad \frac{\partial(y)}{\partial(\bar{w})} = \frac{\partial(y)}{\partial(w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10})}$$

K3. Vi ska ta ett (eller två) steg med gradientmetoden för att träna upp ett litet neuronnät $y = f(\bar{x})$. Antag att det finns två träningspunkter, $\tilde{x} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ ska ge $\tilde{y} = -2$ och $\hat{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ ska ge $\hat{y} = 2$. Neuronnätet har två lager, $f(\bar{x}) = f_2(\bar{f}_1(\bar{x}))$, där

$\bar{f}_1(\bar{x}) = \bar{\sigma}(A_1\bar{x})$ med $A_1 = \begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \end{pmatrix}$ och $\bar{\sigma}(\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}) = \begin{pmatrix} \sigma(t_1) \\ \sigma(t_2) \end{pmatrix} = \begin{pmatrix} \max(t_1, 0) \\ \max(t_2, 0) \end{pmatrix}$,

$f_2(\bar{u}) = A_2\bar{u}$ med $A_2 = \begin{pmatrix} w_5 & w_6 \end{pmatrix}$.

Vi vill minimera $\ell(\bar{w}) = \frac{1}{2}[(f(\tilde{x}, \bar{w}) - \tilde{y})^2 + (f(\hat{x}, \bar{w}) - \hat{y})^2]$ där $\bar{w} = (w_1, w_2, w_3, w_4, w_5, w_6)$.
Som startpunkt i $\mathbb{R}^6$ har vi $\bar{w}_{(0)} = (1, 2, -1, 1, 1, -2)$.

Ange matriserna A_1 och A_2 i startpunkten och beräkna $\tilde{t} = A_1\tilde{x}$, $\hat{t} = A_1\hat{x}$, $\tilde{u} = \bar{\sigma}(\tilde{t})$, $\hat{u} = \bar{\sigma}(\hat{t})$, $f(\tilde{x}) = A_2\tilde{u}$ och $f(\hat{x}) = A_2\hat{u}$. Vad blir $\ell(\bar{w}_{(0)})$?

Beräkna $\nabla\ell(\bar{w}_{(0)})$ och bestäm en ny punkt $\bar{w}_{(1)} = \bar{w}_{(0)} - \eta\nabla\ell(\bar{w}_{(0)})$ för $\eta = 0.1$.
I punkten $\bar{w}_{(1)}$, ange $\tilde{t}$, $\hat{t}$, $\tilde{u}$, $\hat{u}$, $f(\tilde{x})$, $f(\hat{x})$, $\ell(\bar{w}_{(1)})$ och $\nabla\ell(\bar{w}_{(1)})$.

Extra: Ta steget till nästa punkt $\bar{w}_{(2)}$ och beräkna $\ell(\bar{w}_{(2)})$. Jämför resultaten för $\eta = 0.1$ och $\eta = 0.05$ i detta steg.

K4. Beräkna största och minsta värde av $f(x, y) = 2x - y$ under bivillkoret att $g(x, y) = 4x^2 + y^2 = 8$.

K5. Beräkna, eller visa divergens av, följande generaliserade integraler :

a) $\iint_D \frac{dxdy}{x^2(1 + y^2)}$, $D = \{(x, y) \in \mathbb{R}^2; x \geq 1\}$

b) $\iint_D \frac{y}{(x^2 + y^2)^2} dxdy$, $D = \{(x, y) \in \mathbb{R}^2; y \geq 0,\ x^2 + y^2 \geq a > 0\}$

c) $\iint_D \frac{dxdy}{(1 + x + y)^2}$, $D =$ första kvadranten

Svar / lösningar

K1. $\bar{f}'(\bar{x}) = \dfrac{\partial(\bar{f})}{\partial(\bar{x})} = \dfrac{\partial(f_1, f_2)}{\partial(x_1, x_2)} = \begin{pmatrix} (f_1)'_{x_1} & (f_1)'_{x_2} \\ (f_2)'_{x_1} & (f_2)'_{x_2} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

$\begin{pmatrix} g_1(w_1, w_2, w_3, w_4, w_5, w_6) \\ g_2(w_1, w_2, w_3, w_4, w_5, w_6) \end{pmatrix} = \begin{pmatrix} 7w_1 + 8w_2 + w_5 \\ 7w_3 + 8w_4 + w_6 \end{pmatrix} \quad \Rightarrow$

$\dfrac{\partial(\bar{g})}{\partial(\bar{w})} = \dfrac{\partial(g_1, g_2)}{\partial(w_1, w_2, w_3, w_4, w_5, w_6)} = \begin{pmatrix} 7 & 8 & 0 & 0 & 1 & 0 \\ 0 & 0 & 7 & 8 & 0 & 1 \end{pmatrix}$

$\dfrac{\partial(\bar{h})}{\partial(\bar{t})} = \dfrac{\partial(h_1, h_2, h_3)}{\partial(t_1, t_2, t_3)} = \begin{pmatrix} e^{t_1} & 0 & 2t_3 \\ 0 & \cos t_2 & 0 \\ 0 & 0 & 1/t_3 \end{pmatrix}$

K2. a) $\bar{t} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = \begin{pmatrix} w_1x_1 + w_2x_2 \\ w_3x_1 + w_4x_2 \\ w_5x_1 + w_6x_2 \end{pmatrix} \quad \Rightarrow \quad \dfrac{\partial(\bar{t})}{\partial(\bar{x})} = \dfrac{\partial(t_1, t_2, t_3)}{\partial(x_1, x_2)} = \begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \\ w_5 & w_6 \end{pmatrix}$ och

$\dfrac{\partial(\bar{t})}{\partial(w_1, w_2, w_3, w_4, w_5, w_6)} = \dfrac{\partial(t_1, t_2, t_3)}{\partial(w_1, w_2, w_3, w_4, w_5, w_6)} = \begin{pmatrix} x_1 & x_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & x_1 & x_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_1 & x_2 \end{pmatrix}$

$\dfrac{\partial(\bar{u})}{\partial(\bar{t})} = \dfrac{\partial(u_1, u_2, u_3)}{\partial(t_1, t_2, t_3)} = \begin{pmatrix} H(t_1) & 0 & 0 \\ 0 & H(t_2) & 0 \\ 0 & 0 & H(t_3) \end{pmatrix}$, där $H(t) = \begin{cases} 1\,, & t > 0 \\ 0\,, & t < 0 \end{cases}$

$y = w_7u_1 + w_8u_2 + w_9u_3 + w_{10} \;\Rightarrow\; \dfrac{\partial(y)}{\partial(\bar{u})} = \dfrac{\partial(y)}{\partial(u_1, u_2, u_3)} = (w_7, w_8, w_9)$ och

$\dfrac{\partial(y)}{\partial(w_7, w_8, w_9, w_{10})} = (u_1, u_2, u_3, 1)$

b) $\bar{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \Rightarrow \bar{t} = \begin{pmatrix} 1 & 2 \\ -3 & 1 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ -5 \\ 7 \end{pmatrix} \Rightarrow \bar{u} = \begin{pmatrix} 4 \\ 0 \\ 7 \end{pmatrix} \Rightarrow y = (-1, 2, 1)\begin{pmatrix} 4 \\ 0 \\ 7 \end{pmatrix} + 3 = 6 \Rightarrow$

$\dfrac{\partial(\bar{t})}{\partial(\bar{x})} = \begin{pmatrix} 1 & 2 \\ -3 & 1 \\ 2 & 3 \end{pmatrix}$, $\dfrac{\partial(\bar{t})}{\partial(w_1, w_2, w_3, w_4, w_5, w_6)} = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \end{pmatrix}$, $\dfrac{\partial(\bar{u})}{\partial(\bar{t})} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$,

$\dfrac{\partial(y)}{\partial(\bar{u})} = (-1, 2, 1)$ och $\dfrac{\partial(y)}{\partial(w_7, w_8, w_9, w_{10})} = (4, 0, 7, 1)$

c) Kedjeregeln ger

$$\frac{\partial(y)}{\partial(w_1, w_2, w_3, w_4, w_5, w_6)} = \frac{\partial(y)}{\partial(\bar{u})}\frac{\partial(\bar{u})}{\partial(\bar{t})}\frac{\partial(\bar{t})}{\partial(w_1, w_2, w_3, w_4, w_5, w_6)} =$$

$$= (-1, 2, 1)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \end{pmatrix} = (-2, -1, 0, 0, 2, 1) \;\Rightarrow$$

$$\frac{\partial(y)}{\partial(\bar{w})} = \frac{\partial(y)}{\partial(w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10})} = (-2, -1, 0, 0, 2, 1, 4, 0, 7, 1)$$

K3. För $\bar{w}_{(0)}$ fås $A_1 = \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}$ och $A_2 = \begin{pmatrix} 1 & -2 \end{pmatrix} \Rightarrow \tilde{t} = A_1 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \Rightarrow \tilde{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$
$\Rightarrow f(\tilde{x}) = A_2 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = -3$. Liknande fås $\hat{t} = \begin{pmatrix} 1 \\ -4 \end{pmatrix} \Rightarrow \hat{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \Rightarrow f(\hat{x}) = 1$.
$\ell(\bar{w}_{(0)}) = 0.5 \cdot ((-3-(-2))^2 + (1-2)^2) = 0.5 \cdot (1^2 + 1^2) = 1$.
$y = A_2\bar{u} = u_1 w_5 + u_2 w_6$ ger $\dfrac{\partial(y)}{\partial(w_5, w_6)} = \begin{pmatrix} u_1 & u_2 \end{pmatrix} = \bar{u}^T$ och $\dfrac{\partial(y)}{\partial(\bar{u})} = \begin{pmatrix} w_5 & w_6 \end{pmatrix} = A_2$.
$\bar{u} = \bar{\sigma}(\bar{t})$ ger $\dfrac{\partial(\bar{u})}{\partial(\bar{t})} = \begin{pmatrix} H(t_1) & 0 \\ 0 & H(t_2) \end{pmatrix} = H(\bar{t})$ (där $H(t) = 1, t > 0, H(t) = 0, t < 0$).
$\bar{t} = A_1\bar{x} = \begin{pmatrix} x_1 w_1 + x_2 w_2 \\ x_1 w_3 + x_2 w_4 \end{pmatrix}$ ger $\dfrac{\partial(\bar{t})}{\partial(w_1, w_2, w_3, w_4)} = \begin{pmatrix} x_1 & x_2 & 0 & 0 \\ 0 & 0 & x_1 & x_2 \end{pmatrix} = M(\bar{x})$.
Kedjeregeln $\Rightarrow \dfrac{\partial(y)}{\partial(w_1, w_2, w_3, w_4)} = \dfrac{\partial(y)}{\partial(\bar{u})} \dfrac{\partial(\bar{u})}{\partial(\bar{t})} \dfrac{\partial(\bar{t})}{\partial(w_1, w_2, w_3, w_4)} = A_2 H(\bar{t}) M(\bar{x})$.
$\nabla \ell(\bar{w}_{(0)}) = (f(\tilde{x}, \bar{w}_{(0)}) - \tilde{y}) \nabla f(\tilde{x}, \bar{w}_{(0)}) + (f(\hat{x}, \bar{w}_{(0)}) - \hat{y}) \nabla f(\hat{x}, \bar{w}_{(0)}) =$
$= (-3-(-2)) \cdot (A_2 H(\tilde{t}) M(\tilde{x}), \tilde{u}^T) + (1-2) \cdot (A_2 H(\hat{t}) M(\hat{x}), \hat{u}^T) =$
$-\left(\begin{pmatrix} 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}, 1, 2 \right) - \left(\begin{pmatrix} 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & -1 & 0 & 0 \\ 0 & 0 & 3 & -1 \end{pmatrix}, 1, 0 \right)$
$= -(-1, 1, 2, -2, 1, 2) - (3, -1, 0, 0, 1, 0) = (-2, 0, -2, 2, -2, -2)$.

$\bar{w}_{(1)} = \bar{w}_{(0)} - 0.1 \cdot (-2, 0, -2, 2, -2, -2) = (1.2, 2, -0.8, 0.8, 1.2, -1.8)$, ger nya matriser
$A_1 = \begin{pmatrix} 1.2 & 2 \\ -0.8 & 0.8 \end{pmatrix}$, $A_2 = \begin{pmatrix} 1.2 & -1.8 \end{pmatrix} \Rightarrow \tilde{t} = \begin{pmatrix} 0.8 \\ 1.6 \end{pmatrix} \Rightarrow \tilde{u} = \begin{pmatrix} 0.8 \\ 1.6 \end{pmatrix} \Rightarrow f(\tilde{x}) = -1.92$,
$\hat{t} = \begin{pmatrix} 1.6 \\ -3.2 \end{pmatrix} \Rightarrow \hat{u} = \begin{pmatrix} 1.6 \\ 0 \end{pmatrix} \Rightarrow f(\hat{x}) = 1.92$ och $\ell(\bar{w}_{(1)}) = 0.5 \cdot (0.08^2 + 0.08^2) = 0.0064$.
Obs att $\ell(\bar{w}_{(1)}) < \ell(\bar{w}_{(0)})$ så de nya matriserna är bättre anpassade till träningsdata.
$\nabla \ell(\bar{w}_{(1)}) = 0.08 \cdot (-1.2, 1.2, 1.8, -1.8, 0.8, 1.6) - 0.08 \cdot (3.6, -1.2, 0, 0, 1.6, 0)$
$= (-0.384, 0.192, 0.144, -0.144, -0.064, 0.128)$.

Extra: $\eta = 0.1$ ger $\bar{w}_{(2)} = (1.2384, 1.9808, -0.8144, 0.8144, 1.2064, -1.8128)$ och $\ell(\bar{w}_{(2)}) \approx 0.0058$. $\eta = 0.05$ ger $\bar{w}_{(2)} = (1.2192, 1.9904, -0.8072, 0.8072, 1.2032, -1.8064)$ och $\ell(\bar{w}_{(2)}) \approx 0.0001$ (mycket bättre).

K4. Största värde är $f(1, -2) = 4$, minsta $f(-1, 2) = -4$

K5. a) π b) $2/\sqrt{a}$ c) divergent