Blatt 5

Aufgabe 1.

(a) $\Psi = x^T A x + b^T x + \alpha$ ------ $\frac{\partial(\alpha)}{x} = 0$

mit chain rule and $\frac{\partial(x^T y)}{\partial x} = y$

$$\frac{d(f(x,y))}{dx} = \frac{\partial(f(x,y))}{\partial x} + \frac{d(y^T(x))}{dx} \cdot \frac{\partial(f(x,y))}{\partial y}$$

haben wir

① $\frac{d(b^T x)}{dx} = \frac{d(x^T b)}{dx} = b$

② $\frac{d(x^T A x)}{dx} = \frac{\partial(x^T y)}{\partial x} + \frac{d(y(x)^T)}{dx} \cdot \frac{\partial(x^T y)}{\partial y}$ mit $y = Ax$

Dann

$$= y + \frac{d(x^T A^T)}{dx} \cdot x$$

$$= y + A^T x$$

$$= (A + A^T) x$$

Zusammen:

$$\frac{\partial(\Psi)}{x} = (A + A^T) x + b$$

(b) $\tau = \sum_{i=1}^{n} (y_i - \sigma(x_i^T w + b))^2$ mit $\sigma(\alpha) = \frac{1}{1+\exp(-\alpha)}$

1) wenn $y_i = 1$, gilt $(1 - \sigma(x_i^T w + b))^2 = 1 - 2\sigma(x_i^T w + b) + \sigma(x_i^T w + b)^2$

wenn $y_i = 0$ gilt $(-\sigma(x_i^T w + b))^2 = \sigma(x_i^T w + b)^2$

zusammen $\tau = 1 - 2\sigma(x_i^T w + b) + 2\sigma(x_i^T w + b)^2$

2) $\sigma(\alpha) = \frac{1}{1+\exp(-\alpha)}$ Dann

$\sigma(x_i^T w + b) = \frac{1}{1+\exp(x_i^T w + b)}$

3) $\frac{d}{dw} \cdot \frac{d}{db} (\tau) = \frac{d^2}{dwdb} (-2\sigma(x_i^T w + b) + 2\sigma^2(x_i^T w + b))$

$$= -2 \cdot \frac{d}{db} \cdot \left(-\frac{x_i^T e^{x_i^T w + b}}{(1+e^{x_i^T w + b})^2}\right) + 2\frac{d^2}{dwdb} \sigma^2(x_i^T w + b)$$

$$= -2 \cdot \left(- \frac{x_i^T e^{x_i^T w + b} \cdot (-e^{x_i^T w + b} + 1)}{(1+e^{x_i^T w + b})^3} + 2\frac{d^2}{dwdb} \sigma^2(x_i^T w + b)\right.$$

$$= \frac{2x_i^T \cdot e^{x_i^T w + b}(-e^{x_i^T w + b} + 1)}{(1+e^{x_i^T w + b})^3} + 2\frac{d}{db}\left(\frac{-2x_i^T e^{x_i^T w + b}}{(1+e^{x_i^T w + b})^3}\right)$$

$$= \frac{2x_i^T \cdot e^{x_i^T w + b}(-e^{x_i^T w + b} + 1) \cdot (1+e^{x_i^T w + b})}{(1+e^{x_i^T w + b})^4} + \frac{(-4x_i^T e^{x_i^T w + b}(-2e^{x_i^T w + b} + 1)}{(1+e^{x_i^T w + b})^4}$$

$$= \frac{2x_i^T \cdot e^{x_i^T w + b}(1 - e^{2x_i^T w + 2b} - 2 - 4e^{w_i^T w + b})}{(1+e^{x_i^T w + b})^4}$$

c), $x,y \in \mathbb{R}^n$, $A,B \in \mathbb{R}^{m\times m}$

From $\|x\|_2^2 = x^T x$

sei $\sigma(Bx) = c$

Dann $\|y - Ac\|_2^2 = (y-Ac)^T(y-Ac)$

$= y^Ty - (Ac)^Ty - y^TAc + c^TA^TAc$

$= y^Ty - 2(A^Ty)^Tc + c^TA^TAc$

Dann $\nabla(\|y-Ac\|_2^2) = -2A^Ty + (A^TA + (A^TA)^T)c$

mit $(A^TA)^T = A^T(A^T)^T = A^TA$

Dann $\nabla(\|y-Ac\|_2^2) = 2(A^TAc - A^Ty)$

$= 2(A^TA\sigma(Bx) - A^Ty)$

d) $d(\phi^\alpha) = \alpha\phi^{\alpha-1}d\phi$ same as power rule $f(x)=x^n \to f'(x) = nx^{n-1}$

IA. (base case). $n=1$, $f(x) = x$, $f'(x) = (1)x^{1-1} = x^0 = 1$

IS. assume $[x^n]' = nx^{n-1}$, prof $[x^{n+1}]' = (n+1)x^n$

$\Rightarrow f(x) = x^{n+1} = x^n \cdot x$, $f'(x) = (nx^{n-1})x + (x^n)\cdot(1) = nx^n + x^n$

$= (n+1)x^n$

$\square$

same

IA. $\alpha=1$ $d(\phi^1) = 1\cdot\phi^0 \cdot d\phi = 1$. $1 = \phi^0$ ✓

IS. $\alpha = n$ assume $d(\phi^\alpha) = \alpha\phi^{\alpha-1}d\phi$. prof $d(\phi^{\alpha+1}) = (\alpha+1)\phi^\alpha d\phi$

$\Rightarrow d(\phi^{\alpha+1}) = d(\phi^\alpha) * d(\phi^1) = \alpha\phi^{\alpha-1}d\phi \cdot 1\cdot\phi^0 \cdot d\phi$

$= \alpha\phi^\alpha + \phi^\alpha$

$= (\alpha+1)\phi^\alpha$

$\square$

Aufgabe 3. $p(y|X,w) = \mathcal{N}(y|Xw, \sigma^2 I) = \frac{1}{\sigma\sqrt{2\pi I}} \exp(-\frac{(y-Xw)^2}{2\sigma^2 I})$

mit $D = \{(x_1,y_1), \cdots (x_n,y_n)\}$

$$P_w(D) = \prod_{i=1}^{n} P_w(y_i|x_i)$$

Dann

$$w^* = \arg\max_w P_w(D)$$
$$= \arg\max_{w\in\mathbb{R}^d} \sum_{i=1}^{n} \log P_w(y_i|x_i)$$

$\Rightarrow$ mit log-likelihood.

$$\log P_w(D) = \sum_{i=1}^{n} \log P_w(y_i|x_i)$$
$$= \sum_{i=1}^{n} \log\left[\frac{1}{\sigma\sqrt{2\pi I}} \exp(-\frac{(y-Xw)^2}{2\sigma^2 I})\right]$$
$$= \underbrace{\sum_{i=1}^{n} \log\left[\frac{1}{\sigma\sqrt{2\pi I}}\right]}_{\text{independent of } w} + \sum_{i=1}^{n} \left(-\frac{(y-Xw)^2}{2\sigma^2 I}\right)$$

D.h. jetzt

$$w^* = \arg\min_{w\in\mathbb{R}^d} \sum_{i=1}^{n} (y - w^T X)^2 \quad \text{mit } X = \begin{pmatrix} - x_1 - \\ \vdots \\ - x_n - \end{pmatrix}$$

$$y = (y_1 \cdots y_n)^T$$

Dann

$$\sum_{i=1}^{n} (y - wX)^2 = (y - wX)^T (y - Xw)$$
$$= y^T y - 2w^T X^T y + w^T X^T X w$$

$\Leftrightarrow$ respect to w and equating to 0

$$2X^T X w - 2X^T y = 0$$
$$\Leftrightarrow X^T X w = X^T y$$
$$\Rightarrow w^* = (X^T X)^{-1} X^T y$$