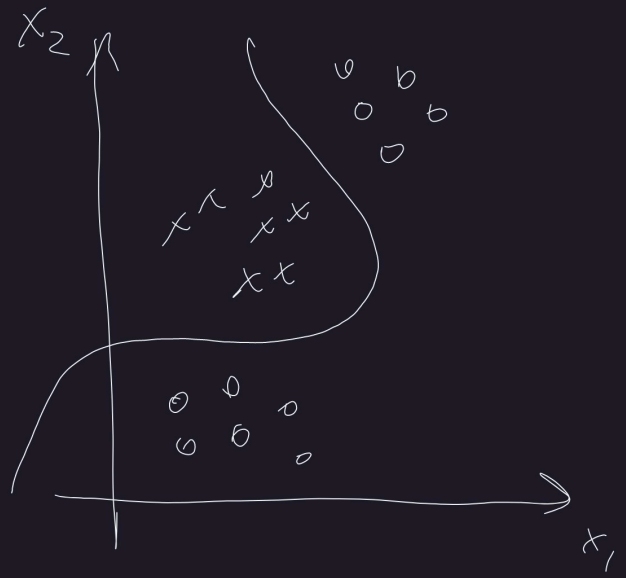
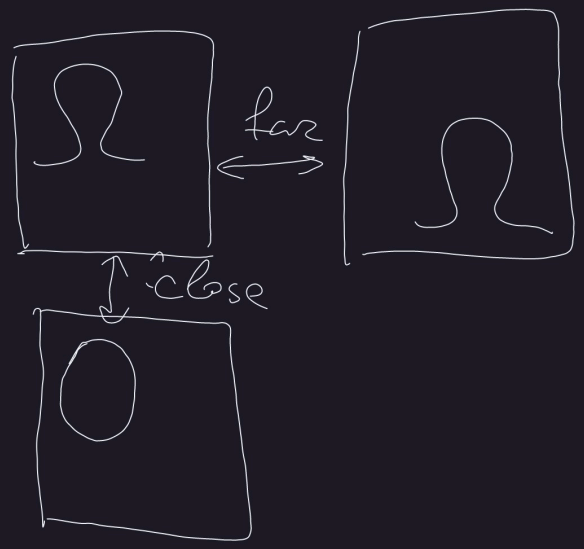
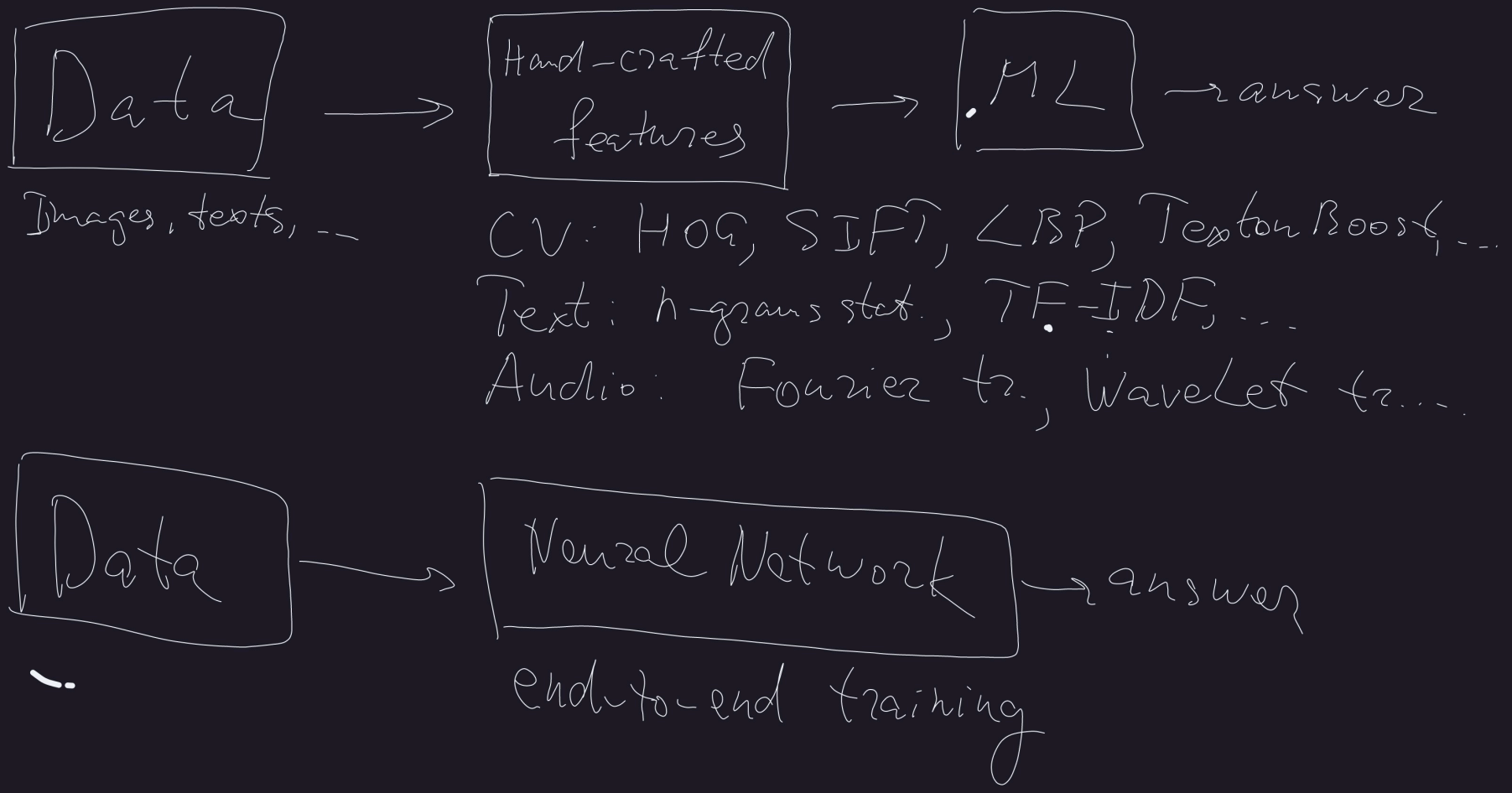
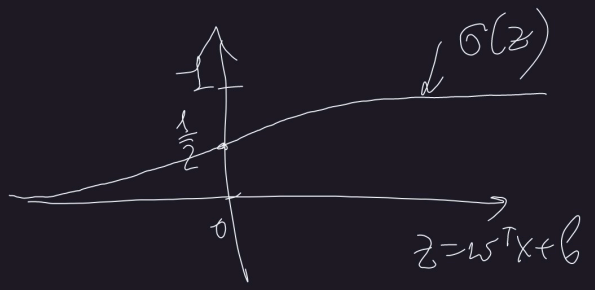

x_2
x_1


far
close


Data
Hand-crafted features
ML
answer
Images, texts, ...
CV: HOG, SIFT, LBP, TextonBoost, ...
Text: n-grams stat., TF-IDF, ...
Audio: Fourier tr., Wavelet tr. ...
Data
Neural Network
answer
end-to-end training



$\{x_i, y_i\}_{i=1}^{N}$, $x_i \in \mathbb{R}^d$, $y_i \in \{-1, +1\}$

$y_i \in \{1, 2, \dots, K\}$

Decision rule: $y(x) = \text{sign}(w^T x + b)$

$1 \times d \quad d \times 1 \quad 1 \times 1$

$$y(x) = \underset{k \in \{1,\dots,K\}}{\text{argmax}} (w_k^T x + b_k)$$

$$p(y=+1 \mid w, b, x) = \frac{1}{1 + \exp(-w^T x - b)} = \sigma(w^T x + b)$$

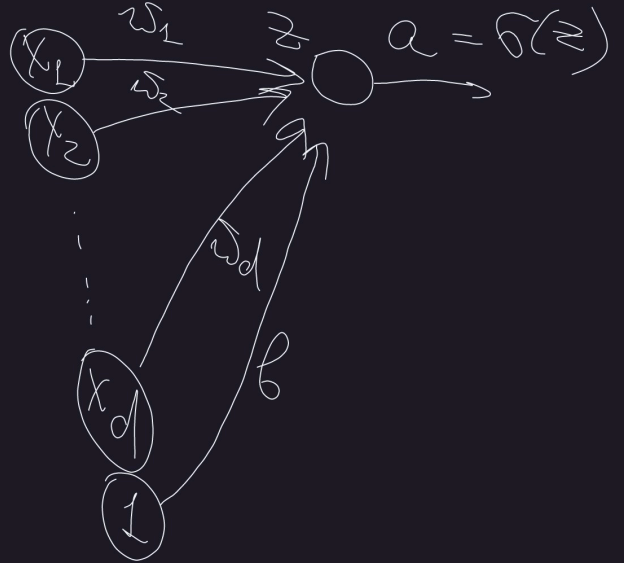
- $$p(y=k \mid W, b, x) = \frac{\exp(w_k^T x + b_k)}{\sum_{j=1}^{K} \exp(w_j^T x + b_j)} = \text{Softmax}(Wx + b)$$

$$p(y_1, \dots, y_N \mid X, w, b) = \prod_{i=1}^{N} p(y_i \mid x_i, w, b) \longrightarrow \max_{w,b}$$

$$-\log p(y_1, \dots, y_N \mid X, w, b) = -\sum_{i=1}^{N} \log p(y_i \mid x_i, w, b) =$$

$$= \sum_{i=1}^{N} \mathcal{L}(y_i, w^T x_i + b) \longrightarrow \min_{w,b}$$

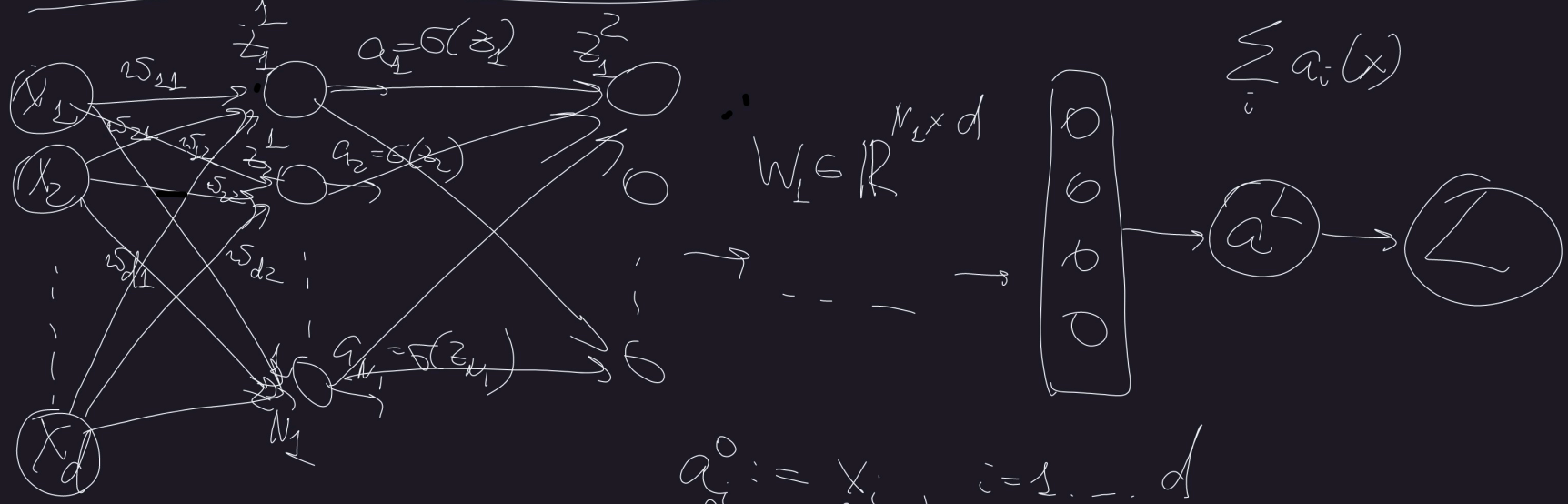
$$z = w^T x + b$$



$$z_2^1 = w_{12} \cdot x_1 + w_{22} \cdot x_2 + \dots + w_{d2} \cdot x_d + b_2$$

$$z_1^1 = w_{11} \cdot x_1 + w_{21} \cdot x_2 + \dots + w_{d1} \cdot x_d + b_1$$

Multi-Layer Perceptron (MLP)



$$a_i^0 := x_i, \quad i = 1 \dots d$$

for $l = 1, \dots, L$:

$$z^l = W^l a^{l-1} + b^l$$

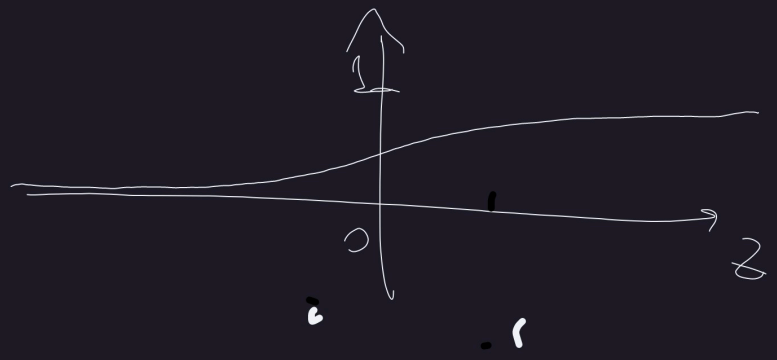
$$a^l = g(z^l)$$

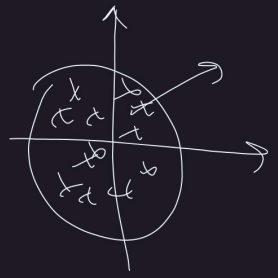
output a^L

$$F(W) = \sum_{i=1}^{N} L\left(y_i, a^L(x_i, W)\right) \to \min_W$$

Activation functions:

① $\sigma(z) = \dfrac{1}{1+\exp(-z)}$





$$\nabla_{a^i} L = \sum_{j=i+1}^{L} (W^{ij})^T \frac{\partial a_j}{\partial z_j} \cdot \nabla_{a^L} L$$

$$\frac{\partial a_j}{\partial z_j} = \operatorname{diag}(\sigma'(z))$$

$$\|\nabla_{a^i} L\| \le \sum_{j=i+1}^{L} \|W^{ij}\| \cdot \left\|\frac{\partial a_j}{\partial z_j}\right\| \cdot \|\nabla_{a^L} L\|$$

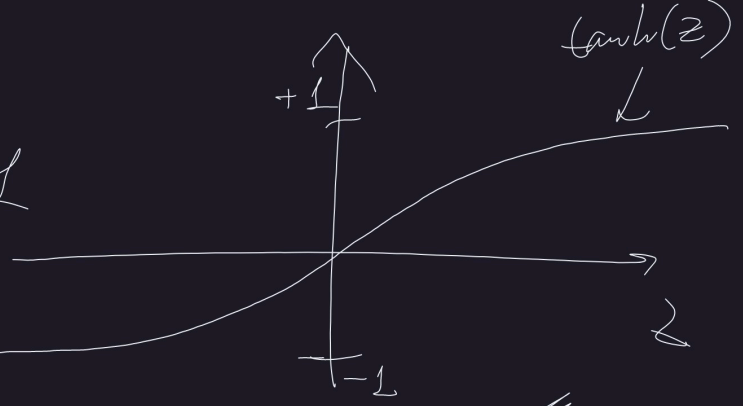
② $g(z) = \tanh(z) = 2\sigma(2z) - 1$

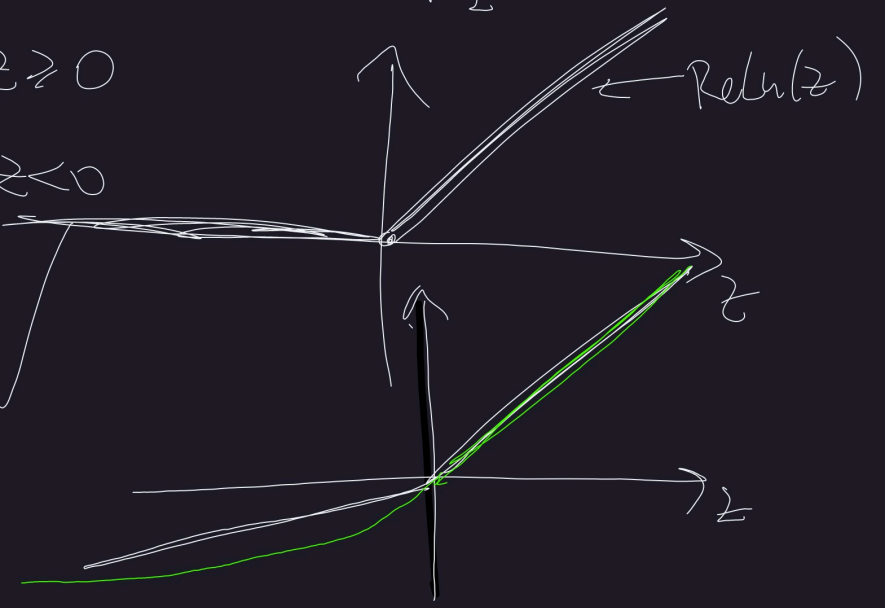
③ $g(z) = ReLu(z) = \max(0, z)$

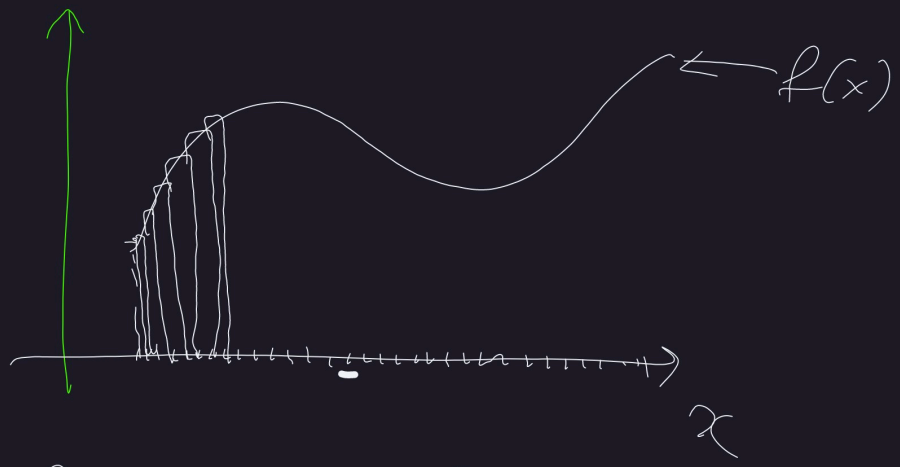
④ $g(z) = LeakyReLu(z) = \begin{cases} z, & z \geq 0 \\ \alpha z, & z < 0 \end{cases}$

$\begin{cases} z, & z \geq 0 \\ \alpha \exp(z) - 1, & z \leq 0 \end{cases}$

⑤ $g(z) = ELu(z)$







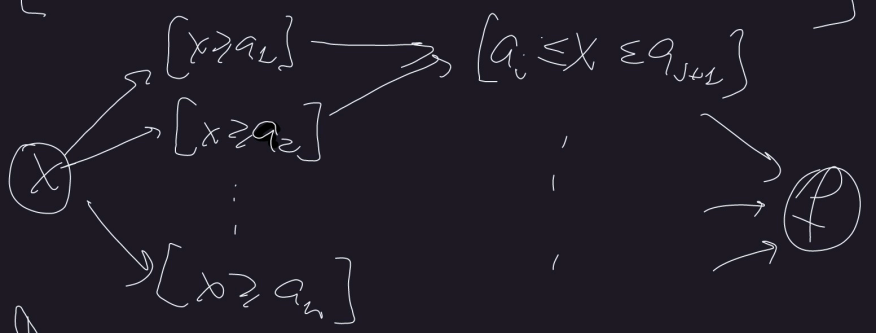
$$f(x) \approx \sum_i \mathbb{I}\left[x \in [a_i, a_{i+1}]\right] \cdot f(a_i)$$

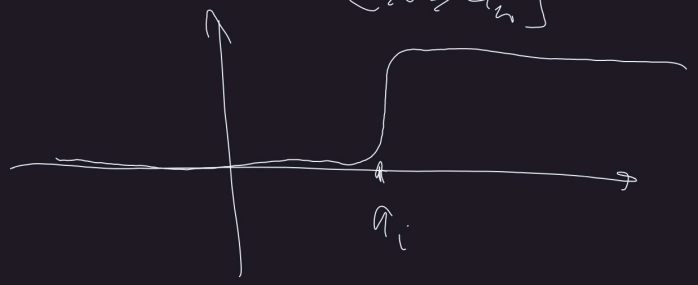
$$[x \geq a_i] \approx \sigma(k(x - a_i))$$

$$[x \leq a_{i+1}] = 1 - [x \geq a_{i+1}] \quad (k \gg 1)$$

$$\mathbb{I}[a_i \leq x \leq a_{i+1}] =$$

$$= \left[[x \geq a_i] + [x \leq a_{i+1}] \geq \frac{3}{2} \right]$$





Automatic Differentiation

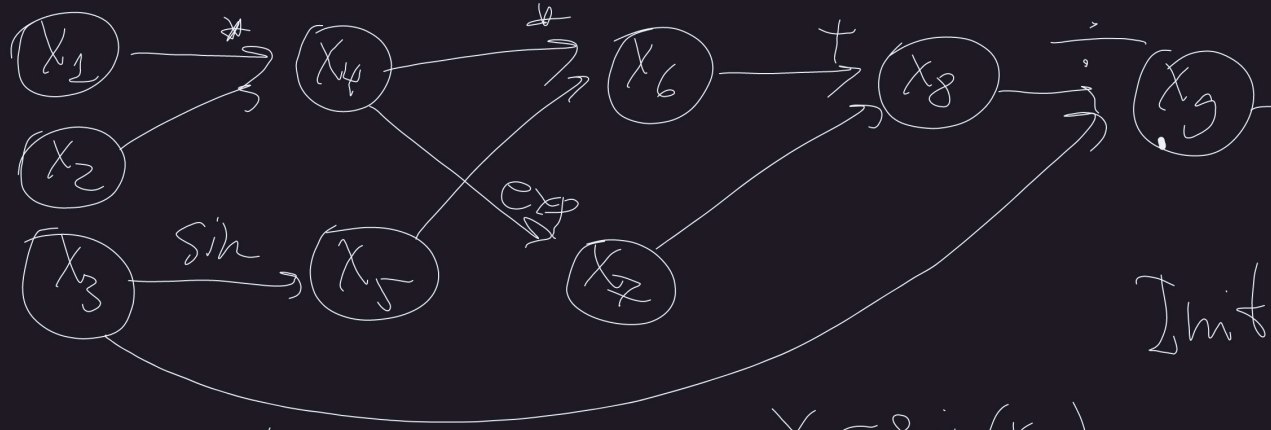
$f(x)$ Find: $\nabla_x f(x)$ $[0\ldots,0,\underset{\text{i-th pos.}}{1},0\ldots,0]$

Finite Diff: $\dfrac{\partial f}{\partial x_i} \approx \dfrac{f(x+\varepsilon e_i) - f(x-\varepsilon e_i)}{2\varepsilon}$ $\varepsilon = \varepsilon_m^{1/3}$

$i = 1, \ldots, d$

$$\frac{\partial f}{\partial x_4} = \frac{\partial f}{\partial x_6}\cdot\frac{\partial x_6}{\partial x_4} + \frac{\partial f}{\partial x_7}\cdot\frac{\partial x_7}{\partial x_4}$$

$$f(x_1, x_2, x_3) = \frac{x_1 \cdot x_2 \cdot \sin(x_3) + \exp(x_2 \cdot x_2)}{x_3}$$



Init: $\frac{\partial x_9}{\partial x_9} = 1$

$\frac{\partial f}{\partial x_8} = \frac{\partial f}{\partial x_9}\cdot\frac{\partial x_9}{\partial x_8} =$

$f = x_9 = x_8 / x_3$

$\frac{\partial f}{\partial x_9}\cdot\frac{1}{x_3}$

Init: $\frac{\partial x_1}{\partial x_1} = 1, \quad \frac{\partial x_2}{\partial x_1} = 0, \quad \frac{\partial x_3}{\partial x_1} = 0$

$x_5 = \sin(x_3)$

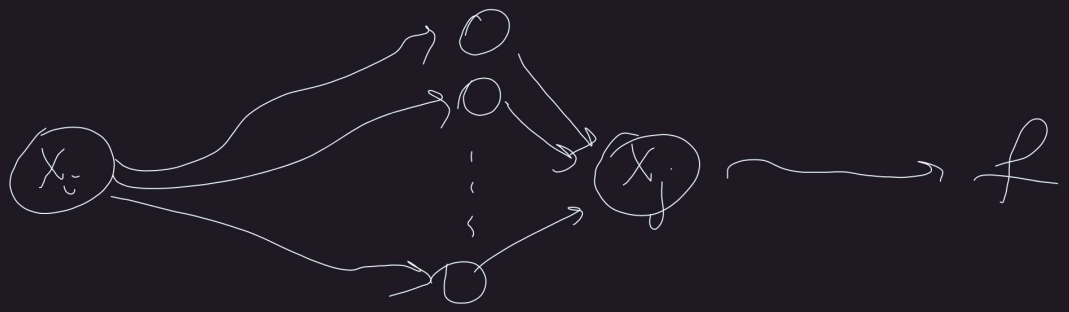
$\frac{\partial x_5}{\partial x_1} = \cos(x_3)\frac{\partial x_3}{\partial x_1}$

$x_4 = x_2 x_2$

$\frac{\partial x_4}{\partial x_1} = \frac{\partial x_2}{\partial x_1}\cdot x_2 + x_2 \cdot \frac{\partial x_2}{\partial x_1}$

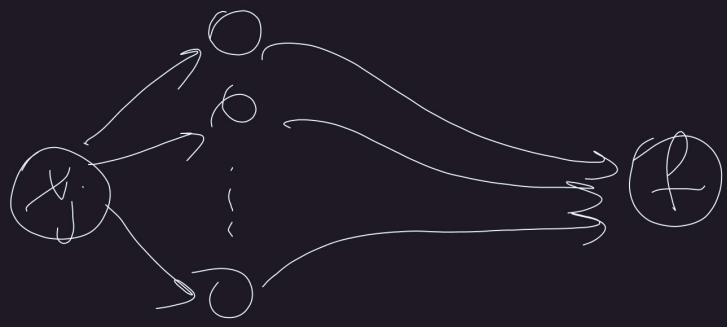
Forward mode

$\frac{\partial x_j}{\partial x_i}, \quad j = 1 \ldots 9, \quad i = 1 \ldots 3$



$$\frac{\partial x_j}{\partial x_i} = \sum_{k:(k,j)\in\mathcal{E}} \frac{\partial x_j}{\partial x_k} \left(\frac{\partial x_k}{\partial x_i} \right)$$

from prev. it.



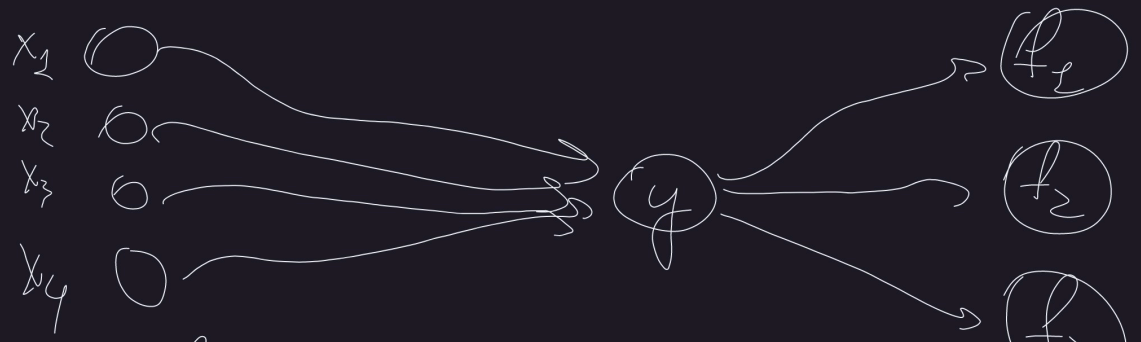
$$\frac{\partial f}{\partial x_j} = \sum_{(j,k)\in\mathcal{E}} \frac{\partial f}{\partial x_k} \cdot \frac{\partial x_k}{\partial x_j}$$

Backward mode

$$\frac{\partial f}{\partial x_j}, \; j = 9, 8, \ldots, 1$$

Cross Mode

$$\left(\frac{\partial f}{\partial x}\right)_{ij} = \frac{\partial f_i}{\partial x_j}$$



$$\frac{\partial f_i}{\partial x_i} = \frac{\partial y}{\partial x_i} \cdot \frac{\partial f_i}{\partial y}$$

$$f(X) = tr(AX^TB) : \mathbb{R}^{n\times n} \to \mathbb{R}$$

$$\frac{\partial f}{\partial X} = \frac{\partial}{\partial X} tr(AX^TB) = tr\Big(\underbrace{\frac{\partial}{\partial X}}_{n\times n} \underbrace{AX^TB}_{n\times n}\Big)$$

4d tensor $n \times n \times n \times n$

Def $f : U \to V$

$$f(x+h) - f(x) = df(x)[h] + \bar{o}(\|h\|)$$

linear op. w.r.t. h

$f : \mathbb{R} \to \mathbb{R}$ $\quad df(x)[h] = f'(x) \cdot h$

$f : \mathbb{R}^d \to \mathbb{R}$ $\quad df(x)[h] = \nabla f(x)^T h$

$df(x)[h] =$

$f : \mathbb{R}^{n\times m} \to \mathbb{R}$

$= \langle \nabla_X f(x), h \rangle =$

$= tr(\nabla_X f(x)^T h)$

$$f(X) = tr(A X^{-1} B)$$

$$df = d\,tr(A X^{-1} B) = tr\left(d(A X^{-1} B)\right) =$$

$$= tr(A \cdot dX^{-1} \cdot B) = tr\left(-A \cdot X^{-1} dX \cdot X^{-1} B\right) =$$

$$-X^{-1} dX X^{-1}$$

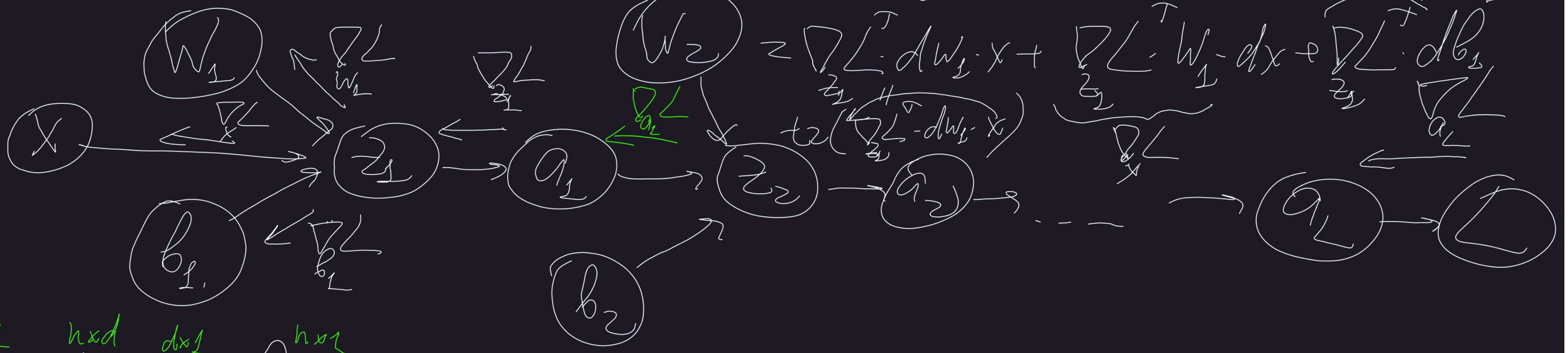
$$= tr(\nabla f(X)^T dX) =$$

$$\nabla f(X) = -X^{-T} A^T B^T X^{-T} = tr\underbrace{\left(-X^{-1} B A X^{-1} dX\right)}_{\nabla f(X)^T}$$

$$dL = \nabla_{z_2} L^T (dW_1 \cdot x + W_1 \cdot dx + db_1) = \nabla_{b_1} L$$

$$= \nabla_{z_1} L^T dW_1 \cdot x + \nabla_{z_2} L^T W_1 \cdot dx + \nabla_{z_1} L^T db_1$$

$$tr(\nabla_{z_1} L^T \cdot dW_1 \cdot x)$$



$$\underset{n\times 1}{z_1} = \underset{n\times d}{W_1} \cdot \underset{d\times 1}{x} + \underset{n\times 1}{b_1}$$

$$a_1 = g(z_1)$$

$$dz_1 = d(W_1 x + b_1) = d(W_1 x) + db_1 =$$

$$= dW_1 \cdot x + W_1 \cdot dx + db_1$$

$$dL = \nabla_{z_1} L^T dz_1 = \nabla_x L^T dx + \nabla_{b_1} L^T db_1 + tr(\nabla_{W_1} L^T dW_1)$$

$$\underset{n\times d}{\nabla_{W_1} L} = \underset{n\times 1}{\nabla_{z_1} L} \cdot \underset{1\times d}{x^T} \qquad \underset{n\times 1}{\nabla_{b_1} L} = \underset{n\times 1}{\nabla_{z_1} L}$$

$$\underset{d\times 1}{\nabla_x L} = \underset{d\times n}{W_1^T} \cdot \underset{n\times 1}{\nabla_{z_1} L}$$

$$a_1 = g(z_1)$$

$$da_1 = \frac{\partial a_1}{\partial z_1} \cdot dz_1$$

$$= \mathrm{diag}(g'(z))$$

$$dL = \nabla_{a_1} L^T \cdot da_1 = \underbrace{\nabla_{a_1} L^T \cdot \frac{\partial a_1}{\partial z_1}}_{\nabla_{z_1} L^T} \cdot dz_1$$

$$\nabla_{z_1} L = \left(\frac{\partial a_1}{\partial z_1}\right)^T \cdot \nabla_{a_1} L$$