Sheet 6

Donnerstag, 27. November 2025 16:06

Task 1

$$y = \exp[\exp(x) + \exp(x)^2] + \sin[\exp(x) + \exp(x)^2]$$

a)
$$f_1 = \exp(x)$$
$$f_2 = f_1^2$$
$$f_3 = f_1 + f_2$$
$$f_4 = \exp(f_3)$$
$$f_5 = \sin(f_3)$$
$$y = f_4 + f_5$$

b)
$$\frac{\partial f_1}{\partial x} = \exp(x)$$
$$\frac{\partial f_2}{\partial x} = \frac{\partial f_2}{\partial f_1}\frac{\partial f_1}{\partial x} = 2\exp(x)\exp(x) = 2\exp(2x)$$
$$\frac{\partial f_3}{\partial x} = \frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial x} = \exp(x) + 2\exp(x)$$
$$\frac{\partial f_4}{\partial x} = \frac{\partial f_4}{\partial f_3}\frac{\partial f_3}{\partial x} = \exp(f_3)[\exp(x) + 2\exp(x)]$$
$$\frac{\partial f_5}{\partial x} = \frac{\partial f_5}{\partial f_3}\frac{\partial f_3}{\partial x} = \cos(f_3)[\exp(x) + 2\exp(x)]$$
$$\frac{\partial y}{\partial x} = \frac{\partial f_4}{\partial x} + \frac{\partial f_5}{\partial x} = \exp(f_3)[\exp(x) + 2\exp(x)] + \cos(f_3)[\exp(x) + 2\exp(x)]$$
$$= \exp[\exp(x) + \exp(x)^2][\exp(x) + 2\exp(x)] + \cos[\exp(x) + \exp(x)^2][\exp(x) + 2\exp(x)]$$

c)
$$\frac{\partial y}{\partial f_5} = f_4$$
$$\frac{\partial y}{\partial f_4} = f_5$$
$$\frac{\partial y}{\partial f_3} = \frac{\partial f_4}{\partial f_3} + \frac{\partial f_5}{\partial f_3} = \exp(f_3) + \cos(f_3)$$
$$\frac{\partial y}{f_2} = \frac{\partial f_4}{\partial f_2} + \frac{\partial f_5}{\partial f_2} = \frac{\partial f_4}{\partial f_3}\frac{\partial f_3}{\partial f_2} + \frac{\partial f_5}{\partial f_3}\frac{\partial f_3}{\partial f_2} = \exp(f_3) f_1 + \cos(f_3) f_1$$
$$\frac{\partial y}{\partial f_1} = \frac{\partial f_4}{\partial f_1} + \frac{\partial f_5}{\partial f_1} = \frac{\partial f_4}{\partial f_3}\frac{\partial f_3}{\partial f_1} + \frac{\partial f_5}{\partial f_3}\frac{\partial f_3}{\partial f_1} = \exp(f_3) f_2 + \cos(f_3) f_2$$
$$\frac{\partial y}{\partial x} = \frac{\partial f_4}{\partial x} + \frac{\partial f_5}{\partial x} = \frac{\partial f_4}{\partial f_3}\frac{\partial f_3}{\partial x} + \frac{\partial f_5}{\partial f_3}\frac{\partial f_3}{\partial x} = \frac{\partial f_4}{\partial f_3}\left(\frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial x}\right) + \frac{\partial f_5}{\partial f_3}\left(\frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial x}\right)$$
$$= \exp[\exp(x) + \exp(x)^2][\exp(x) + 2\exp(x)] + \cos[\exp(x) + \exp(x)^2][\exp(x) + 2\exp(x)]$$

Task 2

a) $\phi_{t+1} \leftarrow \phi_t - \alpha \frac{\tilde{m}_{t+1}}{\sqrt{\tilde{v}_{t+1}} + \varepsilon}$ normalization of the modified grad

$\tilde{m}_{t+1} \leftarrow \frac{m_{t+1}}{1-\beta^{t+1}}$, mod. grad to avoid 0 at beginning

$\tilde{v}_{t+1} \leftarrow \frac{v_{t+1}}{1-\beta^{t+1}}$

$$m_{t+1} \leftarrow \frac{m_{t+1}}{1-\beta^{t+1}}$$

$$\tilde{v}_{t+1} \leftarrow \frac{v_{t+1}}{1-\gamma^{t+1}}$$, mod. grad to avoid 0 at beginning

$$m_{t+1} \leftarrow \beta m_t + (1-\beta) \frac{\partial L[\phi_t]}{\partial \phi}$$

$$v_{t+1} \leftarrow \gamma v_t + (1-\gamma) \left(\frac{\partial L[\phi_t]}{\partial \phi} \right)^2$$ add momentum

b) $$\phi_1 \leftarrow \phi_0 - \alpha \frac{\tilde{m}_1}{\sqrt{\tilde{v}_1}+\varepsilon} , \tilde{m}_1 \leftarrow \frac{m_1}{1-\beta} , \tilde{v}_1 \leftarrow \frac{v_1}{1-\gamma}$$

$$\leftarrow \phi_0 - \alpha \frac{m_1}{\sqrt{v_1}+\varepsilon} \frac{\sqrt{1-\gamma}}{1-\beta} , m_1 = (1-\beta) \frac{\partial L}{\partial \phi} , v_1 = (1-\gamma) \left(\frac{\partial L}{\partial \phi} \right)^2$$

$$\leftarrow \phi_0 - \alpha \frac{\partial L/\partial \phi}{|\partial L/\partial \phi|+\varepsilon} \approx \phi_0 - \alpha \,\mathrm{sgn}(\partial L/\partial \phi) \approx \phi_0 - \mathrm{sgn}(g)$$

$\Rightarrow$ update dep only on $\mathrm{sgn}(g)$

c) keep ε

d) i) put in loss: $g_t = \frac{\partial L}{\partial w} + \lambda w$

ii) direct weight decay: $w \leftarrow \alpha \lambda w$ (AdamW)

AdamW is better. $\Leftarrow$ matches desired effect of L_2 reg

Task 3

b) $$p(y|x,\theta) = \mathcal{N}(y \mid f_\mu(x), \sigma^2)$$

$$-\log[p(y|x,\theta)] = \frac{1}{2\sigma^2}[y - f_y(x)]^2 + \frac{1}{2}\log(\sigma^2) + \text{const} \quad , \sigma = \text{const}$$

$$= \frac{1}{C} \underbrace{[y - f_y(x)]^2}_{\text{MSE}} + \text{const}$$

e) $f_\mu(x_i) = \mu(x_i)$, $f_{\sigma^2}(x_i) = \sigma^2(x_i)$

$$p(y|x,\theta) = \mathcal{N}(y \mid f_\mu(x), f_{\sigma^2}(x))$$

$$\Rightarrow -\log p(y_i|x_i,\theta) = \frac{1}{2}\log(2\pi) + \frac{1}{2}\log \sigma_i^2 + \frac{(y_i - \mu_i)^2}{2\sigma_i^2}$$

$$\Rightarrow -\log p(y_i | x_i, \theta) = \frac{1}{2}\log(2\pi) + \frac{1}{2}\log\sigma_i^2 + \frac{(y_i-\mu_i)^2}{2\sigma_i^2}$$

$\Rightarrow$ total for data set:

$$L = -\frac{1}{2}\sum_{i=1}^{n}\left[\log(2\pi) + \log\sigma_i^2 + \frac{(y_i-\mu_i)^2}{\sigma_i^2}\right]$$
$$= -\frac{1}{2}\sum_{i=1}^{n}\left[\log\sigma_i^2 + \frac{(y_i-\mu_i)^2}{\sigma_i^2}\right] + \text{const}$$

Remark to nb: My export to .pdf is broken and I could not fix it in time. Therefor the nb will only be a jupyter nb.