HW Assignment 9, Theory

1 Activations Functions (10 points)

Compute the derivative of the logistic sigmoid $\sigma(x)$, hyperbolic tangent $tanh(x)$, and ReLU's $ramp(x)$ activation functions. For logistic sigmoid and hyperbolic tangent, express the derivative in terms of the original function.

2 Gradients & Computation Graphs (30 points)

Consider a 3D vector $\mathbf{x} = [x_1, x_2, x_3]^T$ and let $\mathbf{x} \circ \mathbf{x} = [x_1^2, x_2^2, x_3^2]^T$ be the element-wise square of $\mathbf{x}$. Let $h(\mathbf{x})$ be a function computed as follows:

$$\begin{aligned} h(x) &= \sigma(v_1 a_1(\mathbf{x}) + v_2 a_2(\mathbf{x})) \\ a_1(\mathbf{x}) &= z_1^2(\mathbf{x}) \\ z_1(\mathbf{x}) &= \mathbf{w}^T \mathbf{x} \\ a_2(\mathbf{x}) &= \sigma(z_2(\mathbf{x})) \\ z_2(\mathbf{x}) &= \mathbf{u}^T (\mathbf{x} \circ \mathbf{x}) \end{aligned}$$

where $\mathbf{w} = [w_1, w_2, w_3]^T$, $\mathbf{u} = [u_1, u_2, u_3]^T$.

1. Show the computation graph of $h(\mathbf{x})$, similar to how was done in class.
2. Use the chain rule to compute the gradient of h with respect to x_2. Show all your derivation steps and the final formula for the gradient as a product of various factors resulting from the application of the chain rule.

3 Submission

Submit your responses on Canvas as one file named `theory.pdf`. It is recommended to use an editor such as Latex or Word or Jupyter-Notebook that allows editing and proper formatting of equations. Alternatively, if you choose to write your solutions on paper, submit an electronic scan / photo of it on Canvas. Make sure that your writing is legible and the scan has good quality.