Guidebook to Exercise 2 (Derivatives)

Prof. Dr. Ivan Habernal
Yassine Thlija

2025-10-22

1 Basics of Differentiation

1.1 Definition

Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function. Its derivative with respect to x is defined as

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

The derivative measures how the function changes with respect to small changes in its input.

1.2 Common Derivatives

Function	**Derivative**
c	0
x	1
x^n	nx^{n-1}
e^x	e^x
$\ln(x)$	$\frac{1}{x}$
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$
$\sigma(x) = \frac{1}{1+e^{-x}}$	$\sigma(x)(1-\sigma(x))$
$\mathrm{ReLU}(x) = \max(0, x)$	$\begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$

Table 1: Basic functions and their derivatives

2 Partial Derivatives

When a function depends on several variables, $f(x, y, z, \dots)$, we can take the derivative with respect to one variable while keeping others constant.

$$\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x+h, y, z, \dots) - f(x, y, z, \dots)}{h}.$$

Example:

$$f(x, y) = 3x^2y + 2y^3 \quad \Rightarrow \quad \frac{\partial f}{\partial x} = 6xy, \quad \frac{\partial f}{\partial y} = 3x^2 + 6y^2.$$

3 Chain Rule

In deep learning, most functions are compositions of simpler functions. The chain rule allows us to compute the derivative of such compositions.

3.1 Single-variable Chain Rule

If $y = f(g(x))$, then

$$\frac{dy}{dx} = f'(g(x)) \cdot g'(x).$$

3.2 Multivariable Chain Rule

If $z = f(x, y)$ and both x and y depend on t, then

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

This rule generalizes naturally to vector-valued and matrix-valued functions, which are essential for understanding gradient propagation in neural networks.

4 Derivatives of Some Deep Learning Functions

4.1 Sigmoid Function

The sigmoid function is defined as

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$

Its derivative is given by

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)).$$

This derivative is widely used in logistic regression and as an activation function in neural networks.

4.2 ReLU Function

The Rectified Linear Unit (ReLU) is defined as

$$\text{ReLU}(x) = \max(0, x).$$

Its derivative is piecewise constant:

$$\text{ReLU}'(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

4.3 Softmax Function

For a vector $\mathbf{z} = (z_1, \ldots, z_K)$, the softmax is

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}.$$

The derivative of softmax with respect to z_j is

$$\frac{\partial \sigma(z_i)}{\partial z_j} = \sigma(z_i)(\delta_{ij} - \sigma(z_j)),$$

where δ_{ij} is the Kronecker delta.

4.4 Cross-Entropy Loss

For a one-hot encoded target vector $\mathbf{y}$ and predicted probabilities $\hat{\mathbf{y}}$, the cross-entropy loss is

$$L = -\sum_{i=1}^{K} y_i \log(\hat{y}_i).$$

Its derivative with respect to the logits z_i (inputs to the softmax) is

$$\frac{\partial L}{\partial z_i} = \hat{y}_i - y_i.$$