9 Backpropagation

Backpropagation is a method for efficiently computing the gradient vector of the loss function of a neural network. Backpropagation is essentially a systematic and efficient implementation of the chain rule.

131 Definition (Chain Rule)
Let $y = f(x)$ and $z = g(y)$. Then

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}.$$

Modern deep learning packages such as `TensorFlow` implement backpropagation using **computational graphs**. In a computational graph, each node is a tensor, i.e. a scalar, vector, matrix or higher dimensional variable. The computational graph is first generated in a forward pass and is used to evaluate the loss function. Then, in a backward pass, the graph is augmented with additional nodes consisting of partial derivatives. These partial derivatives are computed using the chain rule. Backpropagation is essentially an implementation of **dynamic programming**.

132 Example (Computational Graphs)
Consider the function

$$y = f(x_1, x_2, x_3) = \sin(\ln(x_1 x_2) + x_2) + x_3^2.$$

Compute the gradient vector

$$\nabla f(x_1, x_2, x_3) = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \frac{\partial y}{\partial x_3} \end{pmatrix}$$

by using the computational graph given below.