Derivative Calculation I

- Automatic differentiation (AD) is a method to calculate the derivatives
- In this set of slides, we will introduce the two modes of automatic differentiation: forward mode and reverse mode
- Most materials in this section are from Baydin et al. (2018)
- We consider the same example function

$$f(x_1, x_2) = \log x_1 + x_1 x_2 - \sin x_2$$

Forward Mode of AD I

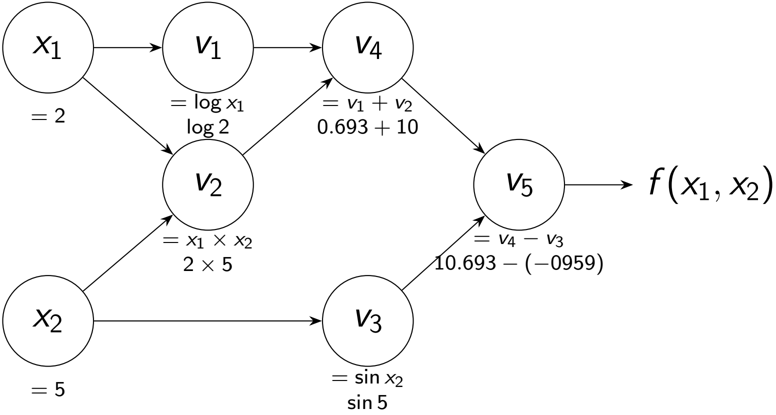
- The following table demonstrates the forward calculation of $f(2,5)$

$$
\begin{array}{lll}
x_1 & & = 2 \\
x_2 & & = 5 \\
\hline
v_1 & = \log x_1 & = \log 2 \\
v_2 & = x_1 \times x_2 & = 2 \times 5 \\
v_3 & = \sin x_2 & = \sin 5 \\
v_4 & = v_1 + v_2 & = 0.693 + 10 \\
v_5 & = v_4 - v_3 & = 10.693 + 0.959 \\
\hline
y & = v_5 & = 11.652
\end{array}
$$

Forward Mode of AD II

- Variables x_1 and x_2 are the input nodes
- Each v_i, $i = 1, 2, 3, \ldots,$ is related to a simple operation
- As shown in the table, a sequence of simple operations leads to the function evaluation
- The table corresponds to the following computational graph

Forward Mode of AD III



Forward Mode of AD IV

- For example, x_1 is used in calculating v_1 and v_2. So x_1 has two links to other nodes
- This computational graph tells us the dependencies of variables
- Remember we would like to calculate the derivative
- In particular, we check the partial derivative with respect to x_1

$$\frac{\partial y}{\partial x_1} = \frac{\partial v_5}{\partial x_1}$$

Forward Mode of AD V

- Here, for any variable v, we denote

$$\dot{v} = \frac{\partial v}{\partial x_1}$$

- We will explain that using the chain rule can give us the following forward derivative calculation

Forward Mode of AD VI

$$
\begin{array}{lll}
\dot{x}_1 & = \partial x_1/\partial x_1 & = 1 \\
\dot{x}_2 & = \partial x_2/\partial x_1 & = 0 \\
\hline
\dot{v}_1 & = \dot{x}_1/x_1 & = 1/2 \\
\dot{v}_2 & = \dot{x}_1 \times x_2 + \dot{x}_2 \times x_1 & = 1 \times 5 + 0 \times 2 \\
\dot{v}_3 & = \dot{x}_2 \times \cos x_2 & = 0 \times \cos 5 \\
\dot{v}_4 & = \dot{v}_1 + \dot{v}_2 & = 0.5 + 5 \\
\dot{v}_5 & = \dot{v}_4 - \dot{v}_3 & = 5.5 - 0 \\
\hline
\dot{y} & = \dot{v}_5 & = 5.5
\end{array}
$$

- The table starts from $\dot{x}_1 = 1$ and $\dot{x}_2 = 0$
- Based on $\dot{x}_1$ and $\dot{x}_2$, we can calculate other values

Forward Mode of AD VII

- For example,

$$
\begin{aligned}
v_1 &= \log x_1 \\
\frac{\partial v_1}{\partial x_1} &= \frac{\partial(\log x_1)}{\partial x_1} \times \frac{\partial x_1}{\partial x_1} \\
&= \frac{1}{x_1} \times \frac{\partial x_1}{\partial x_1} = \frac{\dot{x}_1}{x_1}
\end{aligned}
$$

- Clearly, the chain rule plays an important role here. The calculation of other $\dot{v}_i$ is similar

Reverse Mode of AD I

- Next we discuss the reverse mode of automatic differentiation
- We denote

$$\bar{v} = \frac{\partial y}{\partial v}$$

- Note that earlier, in the forward mode, we considered

$$\dot{v} = \frac{\partial v}{\partial x_1},$$

so the focus is on the derivatives of all variables with respect ot one input variable

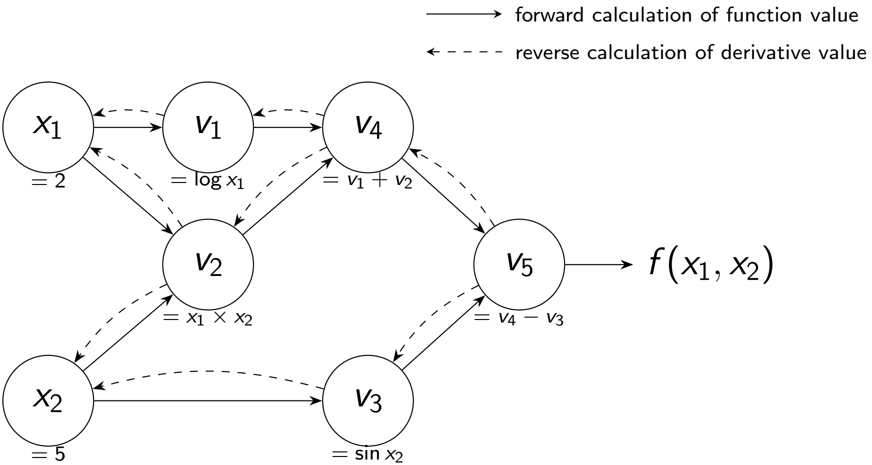
Reverse Mode of AD II

- In contrast, the reverse mode focuses on the partial derivative of one output with respect to all variables
- Here, we illustrate the calculation of

$$\frac{\partial y}{\partial x_2} = \bar{x}_2$$

- By checking the variable x_2 in the computational graph, we see that variable x_2 affects y by affecting v_2 and v_3

Reverse Mode of AD III



Reverse Mode of AD IV

- By the chain rule, we have

$$\frac{\partial y}{\partial x_2} = \frac{\partial y}{\partial v_2}\frac{\partial v_2}{\partial x_2} + \frac{\partial y}{\partial v_3}\frac{\partial v_3}{\partial x_2}$$

  or

$$\bar{x}_2 = \bar{v}_2\frac{\partial v_2}{\partial x_2} + \bar{v}_3\frac{\partial v_3}{\partial x_2}$$

- Assume that $\bar{v}_2$ and $\bar{v}_3$ have been available
- Then we need to calculate

$$\frac{\partial v_2}{\partial x_2} \text{ and } \frac{\partial v_3}{\partial x_2}$$

Reverse Mode of AD V

- From the relation $v_3 = \sin x_2$, we know

$$\frac{\partial v_3}{\partial x_2} = \cos(x_2).$$

Similarly, from $v_2 = x_1 \times x_2$, we have

$$\frac{\partial v_2}{\partial x_2} = x_1$$

Reverse Mode of AD VI

- Then the evaluation of $\bar{x}_2$ is done in two steps:

$$\bar{x}_2 \leftarrow \bar{v}_3 \frac{\partial v_3}{\partial x_2}$$

$$\bar{x}_2 \leftarrow \bar{x}_2 + \bar{v}_2 \frac{\partial v_2}{\partial x_2}$$

- These steps are part of the sequence of a reverse traversal, shown in the following table

Reverse Mode of AD VII

$$
\begin{array}{llll}
\bar{x}_1 & = \bar{x}_1 & & = 5.5 \\
\bar{x}_2 & = \bar{x}_2 & & = 1.716 \\
\hline
\bar{x}_1 & = \bar{x}_1 + \bar{v}_1 \frac{\partial v_1}{\partial x_1} & = \bar{x}_1 + \bar{v}_1 / x_1 & = 5.5 \\
\bar{x}_2 & = \bar{x}_2 + \bar{v}_2 \frac{\partial v_2}{\partial x_2} & = \bar{x}_2 + \bar{v}_2 \times x_1 & = 1.716 \\
\bar{x}_1 & = \bar{v}_2 \frac{\partial v_2}{\partial x_1} & = \bar{v}_2 \times x_2 & = 5 \\
\bar{x}_2 & = \bar{v}_3 \frac{\partial v_3}{\partial x_2} & = \bar{v}_3 \times \cos x_2 & = -0.284 \\
\bar{v}_2 & = \bar{v}_4 \frac{\partial v_4}{\partial v_2} & = \bar{v}_4 \times 1 & = 1 \\
\bar{v}_1 & = \bar{v}_4 \frac{\partial v_4}{\partial v_1} & = \bar{v}_4 \times 1 & = 1 \\
\bar{v}_3 & = \bar{v}_5 \frac{\partial v_5}{\partial v_3} & = \bar{v}_5 \times (-1) & = -1 \\
\bar{v}_4 & = \bar{v}_5 \frac{\partial v_5}{\partial v_4} & = \bar{v}_5 \times 1 & = 1 \\
\hline
\bar{v}_5 & = \bar{y} & = 1 &
\end{array}
$$

Reverse Mode of AD VIII

- We begin with

$$\bar{v}_5 = \frac{\partial y}{\partial v_5} = \frac{\partial y}{\partial y} = 1$$

- By the sequence of a reverse traversal, in the end, we have

$$\frac{\partial y}{\partial x_1} = \bar{x}_1$$

$$\frac{\partial y}{\partial x_2} = \bar{x}_2$$

Reverse Mode of AD IX

- As we mentioned earlier, the reverse mode can obtain the derivatives with respect to all variables at the same time

References I

A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind. Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research*, 18(153):1–43, 2018.