Automatic Differentiation (AD)

Jesse Michel

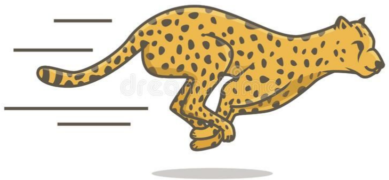
What is AD?

Derivative at a point, fast

Not numerical (finite differences)

Not symbolic (chunky expressions)

$$\frac{\partial f(x_0)}{\partial x}$$



Why AD?

Gradient Descent

Sensitivity Analysis

How does AD work?

- Forward mode (leaves up)
  - Easier to implement
  - Lighter memory footprint
  - Finds the derivative of all outputs wrt to a single input
- Reverse mode (root down)
  - Harder to implement
  - Often more desirable in practice
  - Finds the derivative of all inputs wrt a single output

$$\frac{\partial f}{\partial t} = \sum_i \frac{\partial f}{\partial u_i} \frac{\partial u_i}{\partial t}$$

$$\frac{\partial s}{\partial u} = \sum_i \frac{\partial w_i}{\partial u} \frac{\partial s}{\partial w_i}$$

Forward Mode Example

Forward Mode Code

Forward Mode

- Forward mode (leaves up)
  - Easier to implement
  - Lighter memory footprint
  - Finds the derivative of all outputs wrt to a single input variable

$$\frac{\partial f}{\partial t} = \sum_i \frac{\partial f}{\partial u_i} \frac{\partial u_i}{\partial t}$$

Reverse Mode Example

Reverse Mode Code

Reverse Mode

- Reverse mode (root down)
    - Harder to implement
    - Often more desirable in practice
    - Finds the derivative of all inputs wrt a single output

$$\frac{\partial s}{\partial u} = \sum_i \frac{\partial w_i}{\partial u} \frac{\partial s}{\partial w_i}$$

Forward Mode Another Perspective

Additional Concepts

- Compositions of elementary linear transformations
- Higher-order AD. Iterating?
- Non-differentiable functions

References

For the basics: https://rufflewind.com/2016-12-30/reverse-mode-automatic-differentiation

In a little more depth: https://arxiv.org/pdf/1411.0583.pdf