Introduction to intelligent systems

Automatic differentiation

Mikkel N. Schmidt

Technical University of Denmark,
DTU Compute, Department of Applied Mathematics and Computer Science.

Overview

1. Automatic differentiation
2. Dual numbers
3. PyTorch
4. Tasks

Feedback group

- Mathias Kræmer Eberhardt Sørensen
- Oskar Gotthardt Bak
- Christian Ludvig Meinert Sørensen
- Alexander Baumkirchner

Learning objectives

I Automatic differentiation: Forward and reverse accumulation.

II Computation graphs.

II Automatic differentiation in Pytorch.

II Implementation of neural networks in Pytorch.

I Understand the concepts and definitions, and know their application. Reason about the concepts in the context of an example. Use correct technical terminology.

II As above plus: Read, manipulate, and work with technical definitions and expressions (mathematical and Python code). Carry out practical computations. Interpret and evaluate results.

Automatic differentiation

Gradient descent

Initialize x_0
Repeat, $t = 0, 1, 2, \ldots$

$$\underbrace{x_{t+1}}_{\text{new parameter value}} = \underbrace{x_t}_{\text{old parameter value}} - \underbrace{\alpha}_{\text{step size}} \cdot \underbrace{\nabla f(x_t)}_{\text{gradient}}$$

until convergence

Definition of gradient

$$\nabla f(x, y, \ldots) = \begin{bmatrix} \dfrac{\partial f(x, y)}{\partial x} \\ \\ \dfrac{\partial f(x, y)}{\partial y} \\ \\ \vdots \end{bmatrix}$$

Symbolic, numerical, and automatic differentiation

Symbolic Automatic manipulation of mathematical expressions to get derivatives (e.g. Mathematical, Maple)

Numerical Approximation of derivatives by finite differences

Automatic Automatic computation of the derivative of a compound expression by applying the chain rule

Ryan Adams: You should be using automatic differentiation, https://www.youtube.com/watch?v=sq2gPzlrM0g&t=985s

The chain rule

Derivative of composition of functions, $z(x) = (f \circ g)(x) = f\big(g(x)\big)$

$$z' = (f \circ g)' = (f' \circ g) \cdot g'$$

In Leibnitz's notation

$$\frac{dz}{dx} = \frac{df}{dy} \cdot \frac{dy}{dx}$$

where $z = f(y)$ and $y = g(x)$

Chain rule for functions of multiple variables

Function of two variables $z(t) = f\big(x(t), y(t)\big)$

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

Exercise: Chain rule

Compute the derivative $\frac{dz}{dt}$ of the following function

$$z(t) = f(x, y) = xy + x^2$$

where

$$x(t) = \sin(t)$$
$$y(t) = t^2$$

Exercise: Chain rule

Compute the derivative $\frac{dz}{dt}$ of the following function

$$z(t) = f(x, y) = xy + x^2$$

where

$$\begin{aligned} x(t) &= \sin(t) \\ y(t) &= t^2 \end{aligned}$$

Solution

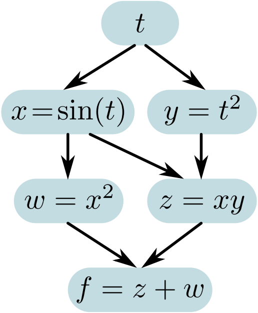
$$\begin{aligned} \frac{dz}{dt} &= \frac{\partial f}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dt} \\ &= (y + 2x) \cdot \cos(t) + x \cdot (2t) \end{aligned}$$

Computation graph

$$f(t) = \sin(t)t^2 + \sin^2(t)$$

Computation graph

$$f(t) = \sin(t)t^2 + \sin^2(t)$$



Exercise: Computation graph

Draw the computation graph for the function

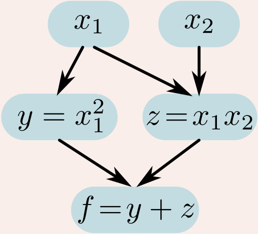
$$f(x_1, x_2) = x_1^2 + x_1 \cdot x_2$$

Exercise: Computation graph

Draw the computation graph for the function

$$f(x_1, x_2) = x_1^2 + x_1 \cdot x_2$$

Solution

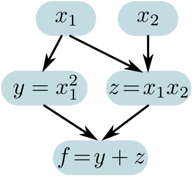


Forward accumulation

Function and derivatives

$$f(x_1, x_2) = x_1^2 + x_1 \cdot x_2$$

$$\frac{\partial f}{\partial x_1} = 2x_1 + x_2, \quad \frac{\partial f}{\partial x_2} = x_1, \quad \nabla f(3,4) = \begin{bmatrix} 10 \\ 3 \end{bmatrix}$$



Evaluate $f(3,4)$

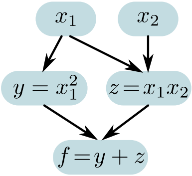
$$\begin{aligned} x_1 &= 3 \\ x_2 &= 4 \\ y &= x_1^2 \\ z &= x_1 x_2 \\ f &= y + z \end{aligned}$$

Forward accumulation

Function and derivatives

$$f(x_1, x_2) = x_1^2 + x_1 \cdot x_2$$

$$\frac{\partial f}{\partial x_1} = 2x_1 + x_2, \quad \frac{\partial f}{\partial x_2} = x_1, \quad \nabla f(3,4) = \begin{bmatrix} 10 \\ 3 \end{bmatrix}$$



Evaluate $f(3,4)$

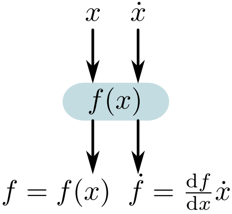
$$\begin{aligned} x_1 &= 3 \\ x_2 &= 4 \\ y &= x_1^2 \\ z &= x_1 x_2 \\ f &= y + z \end{aligned}$$

Evaluate $\nabla_{x_1} f(3,4)$

$$\begin{aligned} \dot{x}_1 &= \frac{\partial x_1}{\partial x_1} = 1 \\ \dot{x}_2 &= \frac{\partial x_2}{\partial x_1} = 0 \\ \dot{y} &= \frac{\partial y}{\partial x_1}\dot{x}_1 = 2x_1 \cdot \dot{x}_1 = 2 \cdot 3 \cdot 1 = 6 \\ \dot{z} &= \frac{\partial z}{\partial x_1}\dot{x}_1 + \frac{\partial z}{\partial x_2}\dot{x}_2 = x_2 \cdot \dot{x}_1 + x_1 \cdot \dot{x}_2 = 4 \cdot 1 + 3 \cdot 0 = 4 \\ \dot{f} &= \frac{\partial f}{\partial y}\dot{y} + \frac{\partial f}{\partial z}\dot{z} = 1 \cdot \dot{y} + 1 \cdot \dot{z} = 1 \cdot 6 + 1 \cdot 4 = \underline{10} \end{aligned}$$

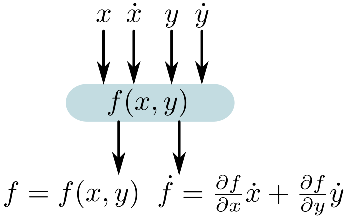
Forward accumulation

Function of one variable



Forward accumulation

Function of multiple variables



Forward accumulation

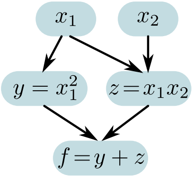
- Direct application of chain rule (going from input to output)
- Computation involves one forward pass through the graph per derivative
- Computationally expensive with many inputs

Reverse accumulation

Function and derivatives

$$f(x_1, x_2) = x_1^2 + x_1 \cdot x_2$$

$$\frac{\partial f}{\partial x_1} = 2x_1 + x_2, \quad \frac{\partial f}{\partial x_2} = x_1, \quad \nabla f(3,4) = \begin{bmatrix} 10 \\ 3 \end{bmatrix}$$



Evaluate $f(3,4)$

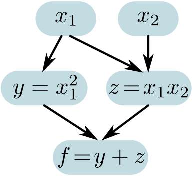
$$\begin{aligned} x_1 &= 3 \\ x_2 &= 4 \\ y &= x_1^2 \\ z &= x_1 x_2 \\ f &= y + z \end{aligned}$$

Reverse accumulation

Function and derivatives

$$f(x_1, x_2) = x_1^2 + x_1 \cdot x_2$$

$$\frac{\partial f}{\partial x_1} = 2x_1 + x_2, \quad \frac{\partial f}{\partial x_2} = x_1, \quad \nabla f(3,4) = \begin{bmatrix} 10 \\ 3 \end{bmatrix}$$



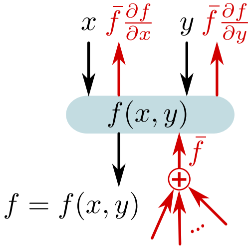
Evaluate $f(3,4)$

$$\begin{aligned} x_1 &= 3 \\ x_2 &= 4 \\ y &= x_1^2 \\ z &= x_1 x_2 \\ f &= y + z \end{aligned}$$

Evaluate $\nabla f(3,4)$

$$\begin{aligned} \bar{f} &= \frac{\partial f}{\partial f} = 1 \\ \bar{y} &= \bar{f}\frac{\partial f}{\partial y} = \bar{f} \cdot 1 = 1 \cdot 1 = 1 \\ \bar{z} &= \bar{f}\frac{\partial f}{\partial z} = \bar{f} \cdot 1 = 1 \cdot 1 = 1 \\ \bar{x}_1 &= \bar{y}\frac{\partial y}{\partial x_1} + \bar{z}\frac{\partial z}{\partial x_1} = \bar{y} \cdot 2 \cdot x_1 + \bar{z} \cdot x_2 = 1 \cdot 2 \cdot 3 + 1 \cdot 4 = 10 \\ \bar{x}_2 &= \bar{z}\frac{\partial z}{\partial x_2} = \bar{x}_2 + \bar{z} \cdot x_1 = 0 + 1 \cdot 3 = 3 \end{aligned}$$

Reverse accumulation



Reverse accumulation

- Direct application of chain rule (going from output to input)
- Computation involves one backward pass through the graph to compute all derivatives
- Requires a bit more “book-keeping” to keep track of dependencies and trace the graph backwards

Dual numbers

Complex and dual numbers

$\mathbb{C}$: *Complex numbers*

$$a + ib$$
$$i^2 = -1$$

Addition

$$(a + ib) + (c + id) = a + c + i(b + d)$$

Multiplication

$$\begin{aligned}(a + ib)(c + id) &= ac + iad + ibc + i^2 bd \\ &= (ac - bd) + i(ad + bc)\end{aligned}$$

$\mathbb{D}$: *Dual numbers*

$$a + \epsilon b$$
$$\epsilon^2 = 0$$

Addition

$$(a + \epsilon b) + (c + \epsilon d) = a + c + \epsilon(b + d)$$

Multiplication

$$\begin{aligned}(a + \epsilon b)(c + \epsilon d) &= ac + \epsilon ad + \epsilon bc + \epsilon^2 bd \\ &= ac + \epsilon(ad + bc)\end{aligned}$$

Example: $f(x) = x^2$

Example

$$f(x) = x^2$$

Example: $f(x) = x^2$

Example

$$f(x) = x^2$$

Evaluating $f(x)$ on $x = a + \epsilon$ we get

$$f(a+\epsilon) = (a+\epsilon)^2 = a^2 + 2a\epsilon + \epsilon^2 = \underbrace{a^2}_{f(a)} + \epsilon \underbrace{2a}_{f'(a)}$$

The dual part happens to be $f'(a) = 2a$. Coincidence?

Exercise: $f(x) = Ax^2 + Bx + C$

Consider the function

$$f(x) = Ax^2 + Bx + C$$

Evaluate the function on $x = a + \epsilon$

Exercise: $f(x) = Ax^2 + Bx + C$

Consider the function

$$f(x) = Ax^2 + Bx + C$$

Evaluate the function on $x = a + \epsilon$

Solution

$$\begin{aligned} f(a+\epsilon) &= A(a+\epsilon)^2 + B(a+\epsilon) + C \\ &= A(a^2 + 2a\epsilon + \epsilon^2) + B(a+\epsilon) + C \\ &= \underbrace{(Aa^2 + Ba + C)}_{f(a)} + \epsilon \underbrace{(2Aa + B)}_{f'(a)} \end{aligned}$$

Taylor series

Taylor series around a

$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \frac{f'''(a)}{3!}(x - a)^3 + \cdots$$

Taylor series

Taylor series around a

$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \frac{f'''(a)}{3!}(x - a)^3 + \cdots$$

Inserting $x = a + \epsilon$

$$\begin{aligned} f(a + \epsilon) &= f(a) + \frac{f'(a)}{1!}\epsilon + \frac{f''(a)}{2!}\epsilon^2 + \frac{f'''(a)}{3!}\epsilon^3 + \cdots \\ &= f(a) + \epsilon f'(a) \end{aligned}$$

PyTorch

PyTorch

Demonstration

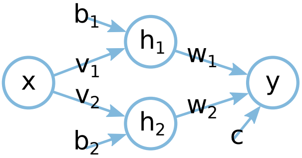
Demo: PyTorch

```
>>> import torch
>>> x1 = torch.tensor(3., requires_grad=True)
>>> x2 = torch.tensor(4., requires_grad=True)
>>> f = x1**2+x1*x2
>>> f.backward()
>>> x1.grad
tensor(10.)
>>> x2.grad
tensor(3.)
```

Neural network notebook

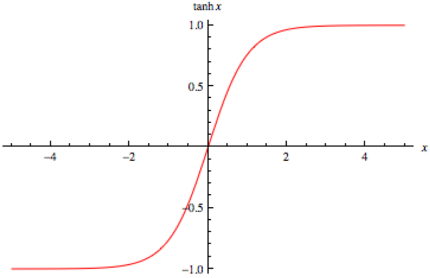
Cost function

$$E = \sum_{n=1}^{N} \big(y(n) - \hat{y}(n)\big)^2$$



Network structure

$$\begin{aligned} \hat{y}(n) &= w_1 h_1(n) + w_2 h_2(n) + c \\ h_1(n) &= \tanh\left(v_1 x(n) + b_1\right) \\ h_2(n) &= \tanh\left(v_2 x(n) + b_2\right) \end{aligned}$$



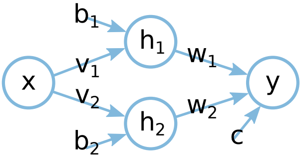
Model parameters

$$c,\ w_1,\ w_2,\ v_1,\ v_2,\ b_1,\ b_2$$

Neural network notebook

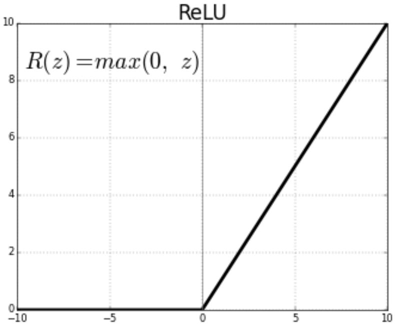
Cost function

$$E = \sum_{n=1}^{N} \left(y(n) - \hat{y}(n)\right)^2$$



Network structure

$$\hat{y}(n) = w_1 h_1(n) + w_2 h_2(n) + c$$
$$h_1(n) = \text{ReLU}\left(v_1 x(n) + b_1\right)$$
$$h_2(n) = \text{ReLU}\left(v_2 x(n) + b_2\right)$$



Model parameters

$$c,\ w_1,\ w_2,\ v_1,\ v_2,\ b_1,\ b_2$$

Tasks

Tasks

1. Work through introduction to PyTorch notebooks
   See `09-PyTorchTutorial1.ipynb` and `09-PyTorchTutorial2.ipynb` on the fileshare
2. Work through introduction to PyTorch notebooks
   See `09-TwoLayerNet-x.ipynb` on the fileshare
3. Experiment with the neural network challenge notebook
   See `09-NeuralNetworkChallenge.ipynb` on the fileshare
4. Today's feedback group
   - Mathias Kræmer Eberhardt Sørensen
   - Oskar Gotthardt Bak
   - Christian Ludvig Meinert Sørensen
   - Alexander Baumkirchner

Lab report

- Lab 4: Neural networks (Deadline: Thursday 9 November 20:00)