Some Notes on Automatic Differentiation

Chih-Jen Lin
National Taiwan University

Last updated: May 25, 2020

Here we give some notes on the slides at https://www.cs.toronto.edu/~rgrosse/courses/csc321_2018/slides/lec10.pdf

P6 I

- The expression on the right means

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathcal{L}} &= 1 \\
\frac{\partial \mathcal{L}}{\partial y} &= y - t \\
\frac{\partial \mathcal{L}}{\partial z} &= \frac{\partial \mathcal{L}}{\partial y}\sigma'(z) \\
\frac{\partial \mathcal{L}}{\partial w} &= \frac{\partial \mathcal{L}}{\partial z} \cdot x \\
\frac{\partial \mathcal{L}}{\partial b} &= \frac{\partial \mathcal{L}}{\partial z} \cdot 1
\end{aligned}
$$

P6 II

- "transform the left-hand side into the right-hand side": we want to calculate

$$\frac{\partial \mathcal{L}}{\partial w}$$

  and it can be replaced by

$$\frac{\partial \mathcal{L}}{\partial z} \cdot x$$

- What we have discussed is the so called reverse mode of automatic differentiation

P6 III

- We notice that in every expression we deal with

$$\frac{\partial \mathcal{L}}{\partial(\text{something})}$$

- Note that the backward setting to calculate the gradient of neural networks is a special case of the reverse mode automatic differentiation
- There is also forward mode of automatic differentiation, in which every node is

$$\frac{\partial(\text{something})}{\partial(\text{variables})}$$

P6 IV

- We will probably see an example of forward mode later in discussing Newton methods

P13 I

- Things shown on this slide are more general
- Even for scalar we do the same thing
- We will see this on p15

P15 I

- We want to calculate

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial y}\frac{\partial y}{\partial x}$$

- So we need

$$\frac{\partial}{\partial y} \text{ and } \frac{\partial y}{\partial x}$$

- Further

$$\frac{\partial y}{\partial x}$$

  may involve y or x

P15 II

- That's why it says to get

$$\frac{\partial}{\partial x}$$

we need

$$\frac{\partial}{\partial y}, x, \text{ and } y$$

P15 III

- Example:

$$
\begin{aligned}
y &= \exp(x) \\
\frac{\partial}{\partial x} &= \frac{\partial}{\partial y}\frac{\partial y}{\partial x} \\
&= \frac{\partial}{\partial y}\exp(x) \\
&= \frac{\partial}{\partial y} y
\end{aligned}
$$

P15 IV

- For this case we need

$$\frac{\partial}{\partial y} \text{ and } y$$

P17 I

- The lines

```
for argnum, parent in zip(argnums, node.pare
    vjp = primitive_vjps[fun][argnum]
    parent_grad = vjp(outgrad, value, *args,
    outgrads[parent] = add_outgrads(outgrads
```

roughly correspond to

$$\frac{\partial}{\partial x_j} = \sum_i \frac{\partial}{\partial y_i} \frac{\partial y_i}{\partial x_j}$$

on p13

P20 I

- We do not discuss pages 20-23