向量链式法则

- 标量链式法则

$$y = f(u),\ u = g(x) \qquad \frac{\partial y}{\partial x} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial x}$$

- 拓展到向量

$$\underset{(1,n)}{\frac{\partial y}{\partial \mathbf{x}}} = \underset{(1,)}{\frac{\partial y}{\partial u}}\ \underset{(1,n)}{\frac{\partial u}{\partial \mathbf{x}}} \qquad \underset{(1,n)}{\frac{\partial y}{\partial \mathbf{x}}} = \underset{(1,k)}{\frac{\partial y}{\partial \mathbf{u}}}\ \underset{(k,n)}{\frac{\partial \mathbf{u}}{\partial \mathbf{x}}} \qquad \underset{(m,n)}{\frac{\partial \mathbf{y}}{\partial \mathbf{x}}} = \underset{(m,k)}{\frac{\partial \mathbf{y}}{\partial \mathbf{u}}}\ \underset{(k,n)}{\frac{\partial \mathbf{u}}{\partial \mathbf{x}}}$$

例子 1

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

假设

$$\mathbf{x}, \mathbf{w} \in \mathbb{R}^n, \quad y \in \mathbb{R}$$

$$z = (\langle \mathbf{x}, \mathbf{w} \rangle - y)^2$$

计算

$$\frac{\partial z}{\partial \mathbf{w}}$$

分解

$$a = \langle \mathbf{x}, \mathbf{w} \rangle$$

$$b = a - y$$

$$z = b^2$$

$$\begin{aligned}
\frac{\partial z}{\partial \mathbf{w}} &= \frac{\partial z}{\partial b} \frac{\partial b}{\partial a} \frac{\partial a}{\partial \mathbf{w}} \\
&= \frac{\partial b^2}{\partial b} \frac{\partial a - y}{\partial a} \frac{\partial \langle \mathbf{x}, \mathbf{w} \rangle}{\partial \mathbf{w}} \\
&= 2b \cdot 1 \cdot \mathbf{x}^T \\
&= 2 \left(\langle \mathbf{x}, \mathbf{w} \rangle - y \right) \mathbf{x}^T
\end{aligned}$$

例子 2

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

假设

$$\mathbf{X} \in \mathbb{R}^{m \times n}, \quad \mathbf{w} \in \mathbb{R}^n, \quad \mathbf{y} \in \mathbb{R}^m$$

$$z = \|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2$$

计算

$$\frac{\partial z}{\partial \mathbf{w}}$$

分解

$$\mathbf{a} = \mathbf{X}\mathbf{w}$$
$$\mathbf{b} = \mathbf{a} - \mathbf{y}$$
$$z = \|\mathbf{b}\|^2$$

$$\begin{aligned}
\frac{\partial z}{\partial \mathbf{w}} &= \frac{\partial z}{\partial \mathbf{b}} \frac{\partial \mathbf{b}}{\partial \mathbf{a}} \frac{\partial \mathbf{a}}{\partial \mathbf{w}} \\
&= \frac{\partial \|\mathbf{b}\|^2}{\partial \mathbf{b}} \frac{\partial \mathbf{a} - \mathbf{y}}{\partial \mathbf{a}} \frac{\partial \mathbf{X}\mathbf{w}}{\partial \mathbf{w}} \\
&= 2\mathbf{b}^T \times \mathbf{I} \times \mathbf{X} \\
&= 2\left(\mathbf{X}\mathbf{w} - \mathbf{y}\right)^T \mathbf{X}
\end{aligned}$$

自动求导

- 自动求导计算一个函数在指定值上的导数
- 它有别于
  - 符号求导

```
In[1]:= D[4 x^3 + x^2 + 3, x]

Out[1]= 2 x + 12 x^2
```

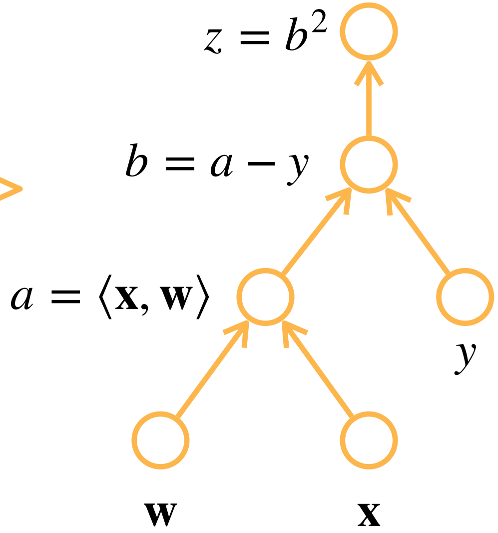
  - 数值求导

$$\frac{\partial f(x)}{\partial x} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

计算图

- 将代码分解成操作子
- 将计算表示成一个无环图

$$z = (\langle \mathbf{x}, \mathbf{w} \rangle - y)^2$$



计算图

- 将代码分解成操作子
- 将计算表示成一个无环图
- 显示构造

```python
from mxnet import sym

a = sym.var()
b = sym.var()
c = 2 * a + b
# bind data into a and b later
```

计算图

- 将代码分解成操作子
- 将计算表示成一个无环图
- 显式构造
  - Tensorflow/Theano/MXNet
- 隐式构造
  - PyTorch/MXNet

```python
from mxnet import autograd, nd

with autograd.record():
    a = nd.ones((2,1))
    b = nd.ones((2,1))
    c = 2 * a + b
```

自动求导的两种模式

- 链式法则：$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u_n} \frac{\partial u_n}{\partial u_{n-1}} \ldots \frac{\partial u_2}{\partial u_1} \frac{\partial u_1}{\partial x}$

- 正向累积 $\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u_n} \left(\frac{\partial u_n}{\partial u_{n-1}} \left(\cdots \left(\frac{\partial u_2}{\partial u_1} \frac{\partial u_1}{\partial x} \right) \right) \right)$

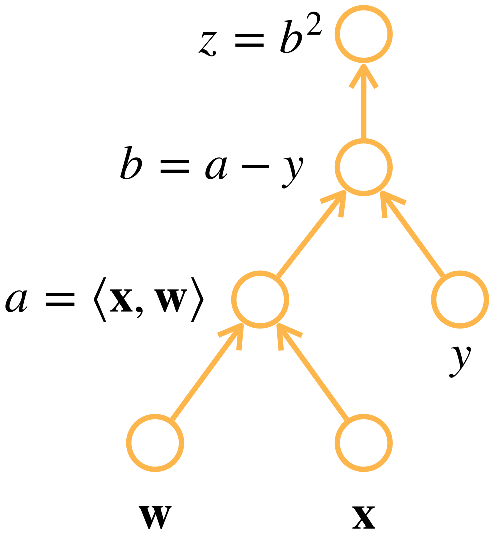
- 反向累积、又称反向传递

$$\frac{\partial y}{\partial x} = \left(\left(\left(\frac{\partial y}{\partial u_n} \frac{\partial u_n}{\partial u_{n-1}} \right) \ldots \right) \frac{\partial u_2}{\partial u_1} \right) \frac{\partial u_1}{\partial x}$$

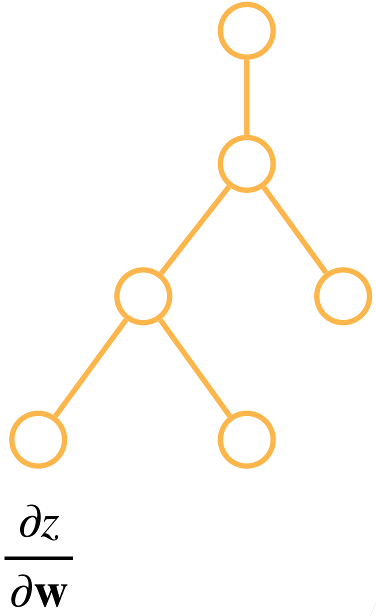
反向累积

$$z = \left(\langle \mathbf{x}, \mathbf{w} \rangle - y\right)^2$$

正向

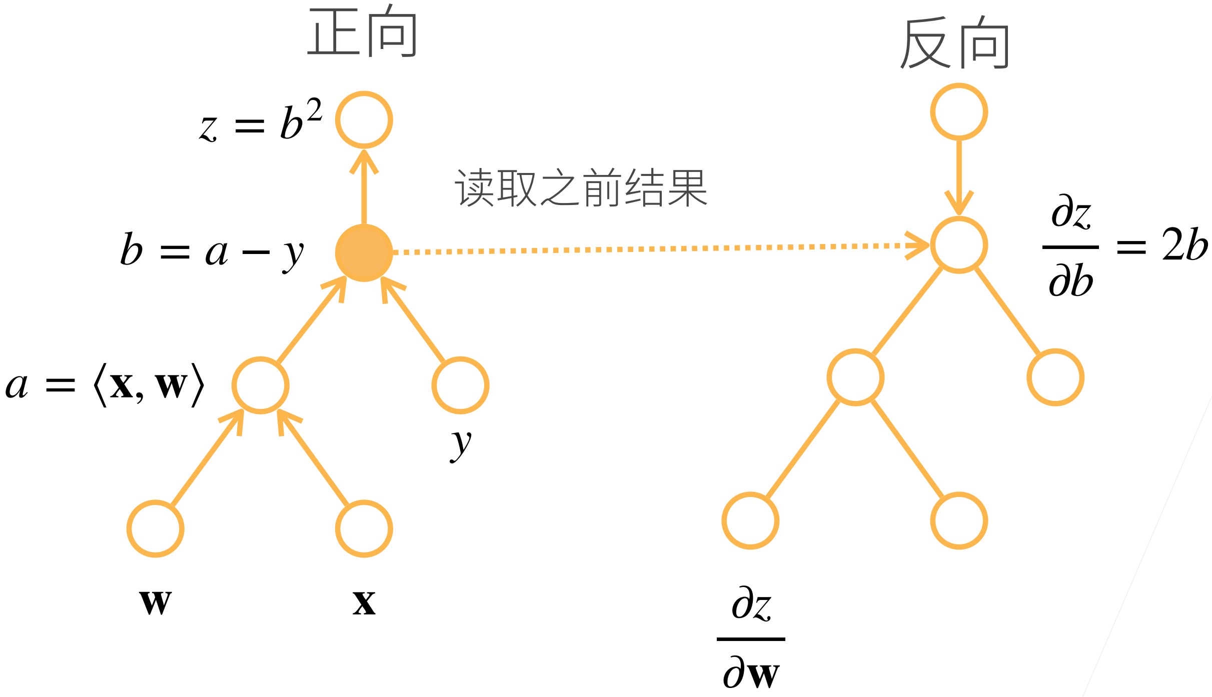


反向



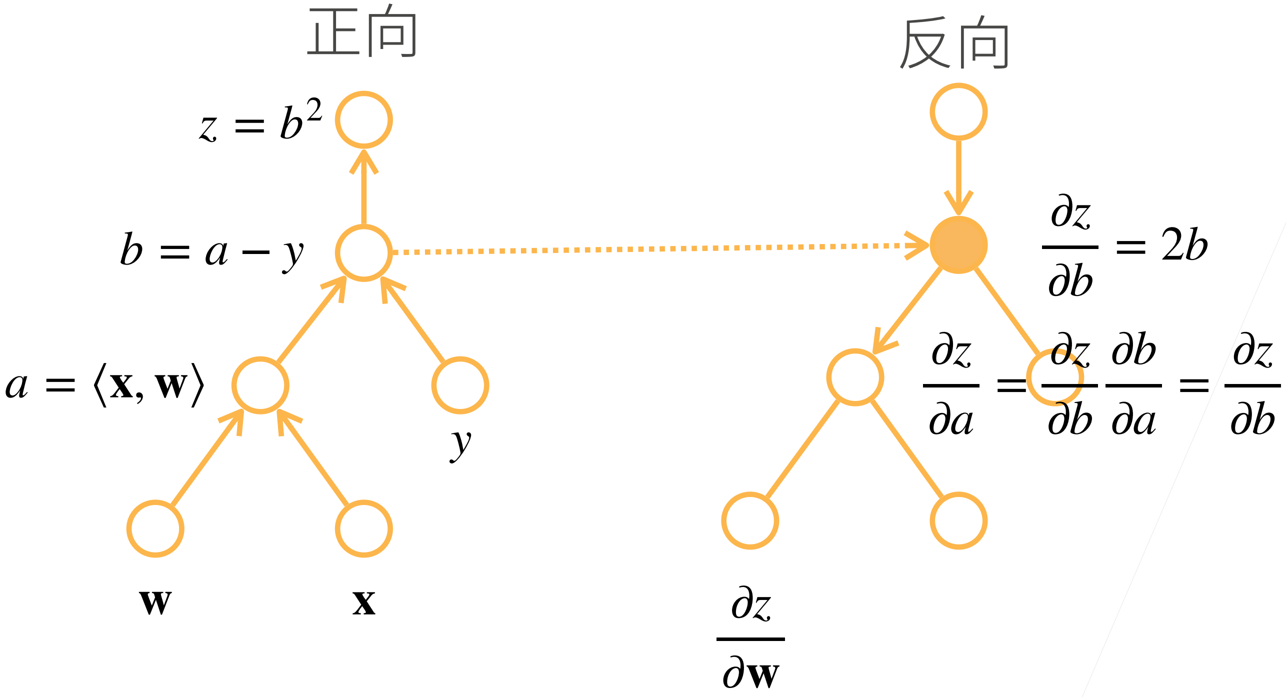
反向累积

$$z = \left(\langle \mathbf{x}, \mathbf{w} \rangle - y\right)^2$$



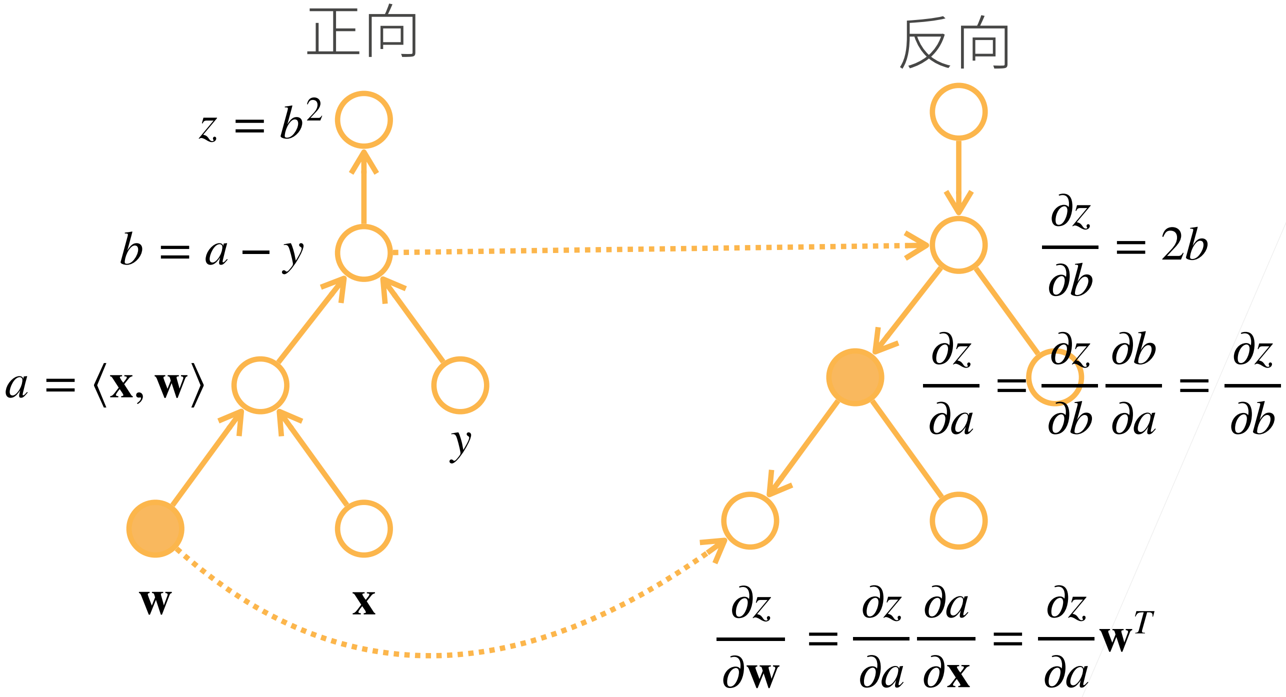
反向累积

$$z = \left(\langle \mathbf{x}, \mathbf{w} \rangle - y\right)^2$$



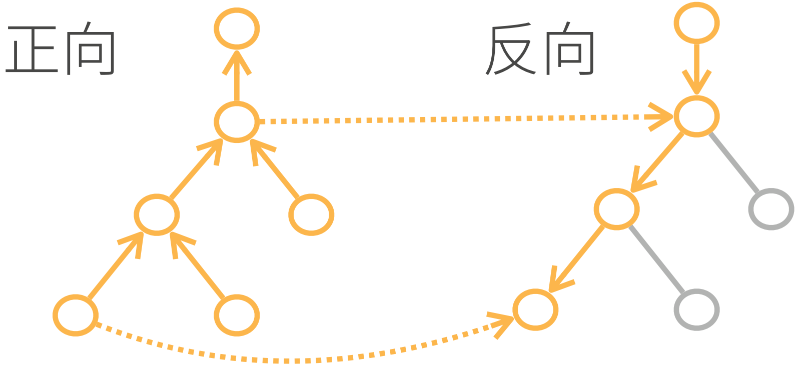
反向累积

$$z = \left(\langle \mathbf{x}, \mathbf{w} \rangle - y\right)^2$$



反向累积总结

- 构造计算图
- 前向：执行图，存储中间结果
- 反向：从相反方向执行图
  - 去除不需要的枝



复杂度

- 计算复杂度：O(n), n 是操作子个数
  - 通常正向和方向的代价类似
- 内存复杂度: O(n)，因为需要存储正向的所有中间结果
- 跟正向累积对比：
  - O(n) 计算复杂度用来计算一个变量的梯度
  - O(1) 内存复杂度