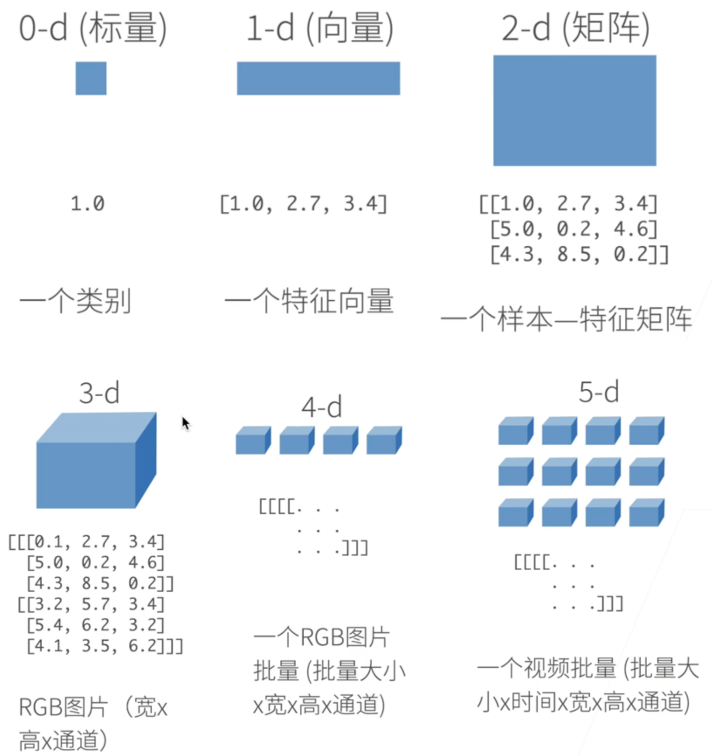
深度学习基础

数据操作和数据预处理

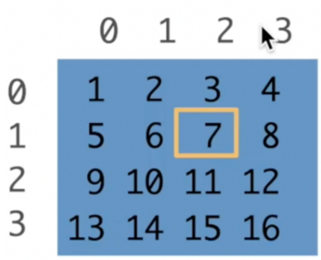
1. N 维数组样例

- N 维数组是机器学习和神经网络的主要数据结构

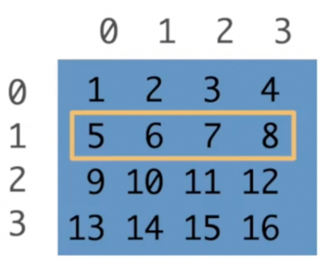


2. 访问元素

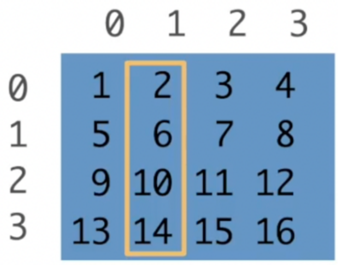
- 一个元素： **[1 , 2]**



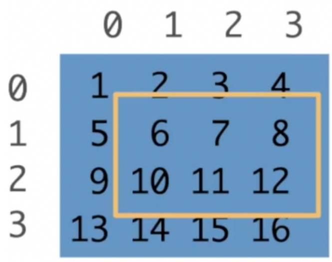
- 一行： **[1 , :]**



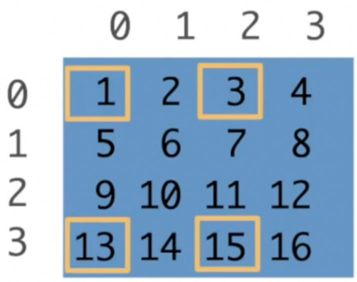
- 一列： **[: , 1]**



- 子区域： **[1:3 , 1:]**



- 子区域： **[::3 , ::2]**



导数

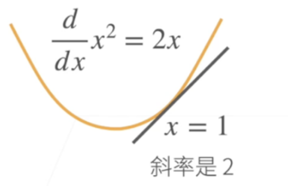
1. 标量导数:

y	a	x^n	$\exp(x)$	$\log(x)$	$\sin(x)$
$\frac{dy}{dx}$	0	nx^{n-1}	$\exp(x)$	$\frac{1}{x}$	$\cos(x)$

a 不是 x 的函数

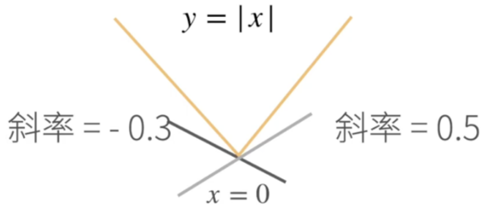
y	$u+v$	uv	$y=f(u),\ u=g(x)$
$\frac{dy}{dx}$	$\frac{du}{dx}+\frac{dv}{dx}$	$\frac{du}{dx}v+\frac{dv}{dx}u$	$\frac{dy}{du}\frac{du}{dx}$

导数是切线的斜率



2. 亚导数:

- 将导数拓展到不可微的函数



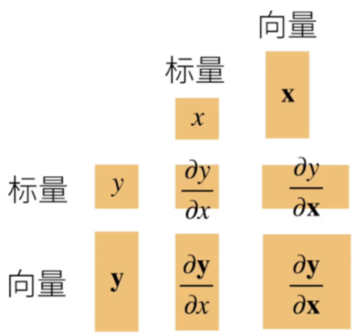
另一个例子

$$\frac{\partial |x|}{\partial x} = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \\ a & \text{if } x = 0, \quad a \in [-1,1] \end{cases}$$

$$\frac{\partial}{\partial x}\max(x,0) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x < 0 \\ a & \text{if } x = 0, \quad a \in [0,1] \end{cases}$$

3. 梯度:

- 将导数拓展到向量



注: x, y 均为向量，则输出为矩阵(右下角)

3.1 x 为向量

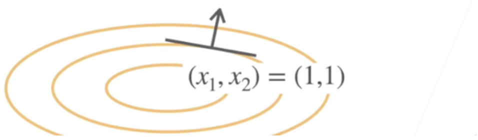
- 下图中，点（1，1）处的梯度即为方向（2，4），与等高线正交，**意味着梯度指向的是值变化最大的方向**。

$$\partial y / \partial \mathbf{x}$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \frac{\partial y}{\partial \mathbf{x}} = \left[\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \ldots, \frac{\partial y}{\partial x_n} \right]$$

$$\frac{\partial}{\partial \mathbf{x}} x_1^2 + 2x_2^2 = [2x_1, 4x_2]$$

方向 (2, 4) 跟等高线正交



- 样例:

y	a	au	sum($\mathbf{x}$)	$\|\mathbf{x}\|^2$
$\frac{\partial y}{\partial \mathbf{x}}$	$\mathbf{0}^T$	$a\frac{\partial u}{\partial \mathbf{x}}$	$\mathbf{1}^T$	$2\mathbf{x}^T$

a is not a function of $\mathbf{x}$

$\mathbf{0}$ and $\mathbf{1}$ are vectors

y	$u + v$	uv	$\langle \mathbf{u}, \mathbf{v} \rangle$
$\frac{\partial y}{\partial \mathbf{x}}$	$\frac{\partial u}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}}$	$\frac{\partial u}{\partial \mathbf{x}} v + \frac{\partial v}{\partial \mathbf{x}} u$	$\mathbf{u}^T \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^T \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$

3.2 y 为向量

$$\partial \mathbf{y} / \partial x$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \qquad \frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}$$

$\partial y / \partial \mathbf{x}$ 是行向量, $\partial \mathbf{y} / \partial x$ 是列向量

这个被称之为分子布局符号，反过来的版本叫分母布局符号

3.3 x,y 均为向量

$$\partial\mathbf{y}/\partial\mathbf{x}$$

$$\mathbf{x}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}\qquad\mathbf{y}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_m\end{bmatrix}$$

$$\frac{\partial\mathbf{y}}{\partial\mathbf{x}}=\begin{bmatrix}\frac{\partial y_1}{\partial\mathbf{x}}\\\frac{\partial y_2}{\partial\mathbf{x}}\\\vdots\\\frac{\partial y_m}{\partial\mathbf{x}}\end{bmatrix}=\begin{bmatrix}\frac{\partial y_1}{\partial x_1},\frac{\partial y_1}{\partial x_2},\ldots,\frac{\partial y_1}{\partial x_n}\\\frac{\partial y_2}{\partial x_1},\frac{\partial y_2}{\partial x_2},\ldots,\frac{\partial y_2}{\partial x_n}\\\vdots\\\frac{\partial y_m}{\partial x_1},\frac{\partial y_m}{\partial x_2},\ldots,\frac{\partial y_m}{\partial x_n}\end{bmatrix}$$

- 样例:

$\mathbf{y}$	$\mathbf{a}$	$\mathbf{x}$	$\mathbf{Ax}$	$\mathbf{x}^T\mathbf{A}$
$\frac{\partial\mathbf{y}}{\partial\mathbf{x}}$	$\mathbf{0}$	$\mathbf{I}$	$\mathbf{A}$	$\mathbf{A}^T$

$\mathbf{y}$	$a\mathbf{u}$	$\mathbf{Au}$	$\mathbf{u}+\mathbf{v}$
$\frac{\partial\mathbf{y}}{\partial\mathbf{x}}$	$a\frac{\partial\mathbf{u}}{\partial\mathbf{x}}$	$\mathbf{A}\frac{\partial\mathbf{u}}{\partial\mathbf{x}}$	$\frac{\partial\mathbf{u}}{\partial\mathbf{x}}+\frac{\partial\mathbf{v}}{\partial\mathbf{x}}$

$\mathbf{x}\in\mathbb{R}^n,\quad\mathbf{y}\in\mathbb{R}^m,\quad\frac{\partial\mathbf{y}}{\partial\mathbf{x}}\in\mathbb{R}^{m\times n}$

a, $\mathbf{a}$ and $\mathbf{A}$ are not functions of $\mathbf{x}$

$\mathbf{0}$ and $\mathbf{I}$ are matrices

3.4 拓展到矩阵

		标量	向量	矩阵
		x (1,)	**x** (n,1)	**X** (n, k)
标量	y (1,)	$\frac{\partial y}{\partial x}$ (1,)	$\frac{\partial y}{\partial \boldsymbol{x}}$ $(1,n)$	$\frac{\partial y}{\partial \boldsymbol{X}}$ (k,n)
向量	**y** (m,1)	$\frac{\partial \boldsymbol{y}}{\partial x}(m,1)$	$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}(m,n)$	$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{X}}(m,k,n)$
矩阵	**Y** (m, l)	$\frac{\partial \boldsymbol{Y}}{\partial x}(m,l)$	$\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{x}}(m,l,n)$	$\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}}(m,l,k,n)$

自动求导

1. 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

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

1.1 例 1

假设

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

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

计算

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

分解

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

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

1.2 例 2

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

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

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

2. 自动求导

- 自动求导计算一个函数在指定值上的导数
- 它有区别于

- 符号求导

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

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

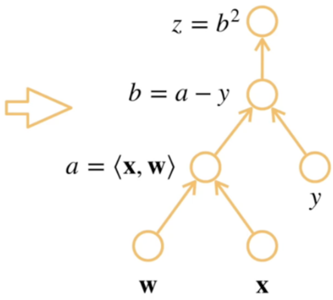
- 数值求导

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

2.1 计算图

- 将代码分解成操作子
- 将计算表示成一个无环图

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



- 显式构造
  Tensorflow/Theano/MXNet

```python
from mxnet import sym

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

- 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
```

3. 自动求导的两种模式

- **链式法则**

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u_n} \frac{\partial u_n}{\partial u_{n-1}} \cdots \frac{\partial u_2}{\partial u_1} \frac{\partial u_1}{\partial x}$$

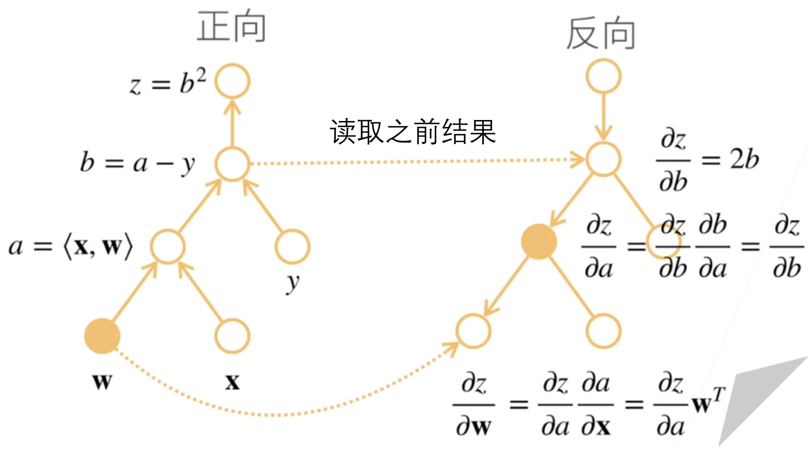
- **正向累积**

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

- **反向累积/反向传递**

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

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



- **反向累积总结**
  a) 构造计算图
  b) 前向: 执行图，存储中间结果
  c) 反向: 从相反方向执行图
  去除不需要的枝

- **反向累积复杂度**
  a) 计算复杂度: O(n)，n 表示操作子个数
  通常正向和反向的代价类似
  b) 内存复杂度: O(n)，因为需要存储正向的所有中间结果
  c) 相比之下，正向累积:
     i. O(n) 计算复杂度用来计算一个变量的梯度
     ii. O(1) 内存复杂度