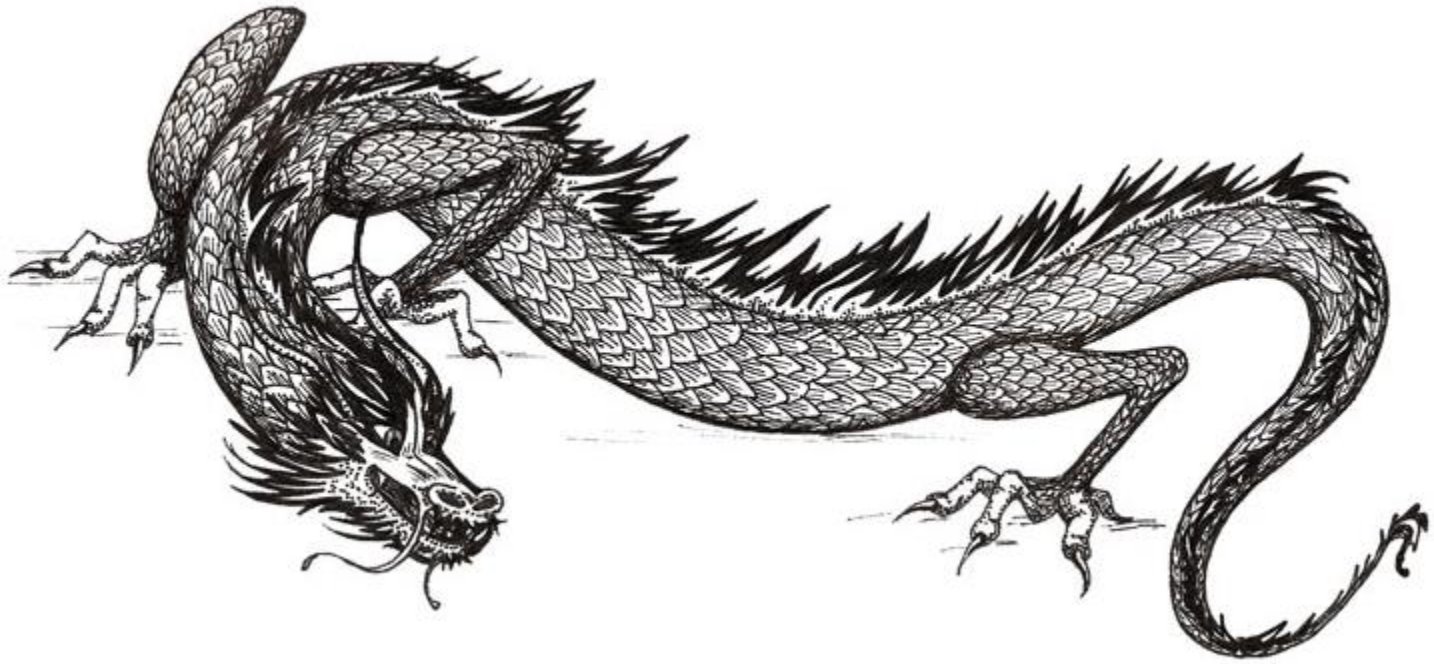


链式法则

主讲：龙良曲

Derivative Rules

Rules	Function	Derivative
Multiplication by constant	cf	cf'
Power Rule	x^n	nx^{n-1}
Sum Rule	$f + g$	$f' + g'$
Difference Rule	$f - g$	$f' - g'$
Product Rule	fg	$f\,g' + f'\,g$
Quotient Rule	f/g	$(f'\,g - g'\,f)/g^2$
Reciprocal Rule	$1/f$	$-f'/f^2$
Chain Rule (as "Composition of Functions")	$f \circ g$	$(f' \circ g) \times g'$
Chain Rule (using ′)	$f(g(x))$	$f'(g(x))g'(x)$
Chain Rule (using $\frac{d}{dx}$)	$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$	

Basic Rule

- $f + g$

- $f - g$

Product rule

- $(fg)' = f'g + fg'$

- $x^{4'} = (x^2 * x^2)' = 2x * x^2 + x^2 * 2x = 4x^3$

Quotient Rule

- $(\frac{f}{g})' = \frac{f'g - fg'}{g^2}$
- e.g. Softmax

$$p_i = \frac{e^{a_i}}{\sum_{k=1}^{N} e^{a_k}}$$

$$\frac{\partial p_i}{\partial a_j} = \frac{\partial \frac{e^{a_i}}{\sum_{k=1}^{N} e^{a_k}}}{\partial a_j}$$

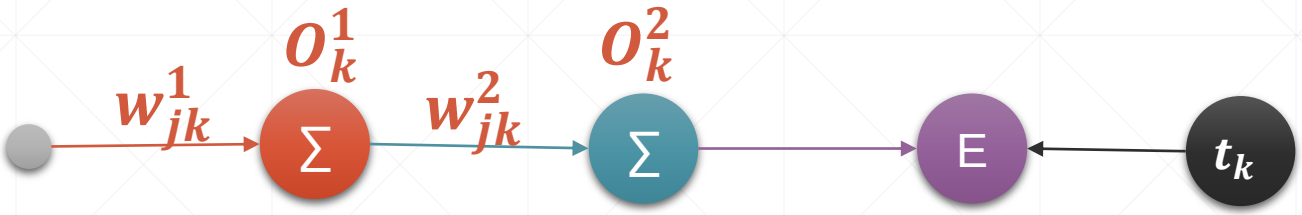
$$\begin{aligned}
\frac{\partial \frac{e^{a_i}}{\sum_{k=1}^{N} e^{a_k}}}{\partial a_j} &= \frac{e^{a_i} \sum_{k=1}^{N} e^{a_k} - e^{a_j} e^{a_i}}{\left(\sum_{k=1}^{N} e^{a_k}\right)^2} \\
&= \frac{e^{a_i} \left(\sum_{k=1}^{N} e^{a_k} - e^{a_j}\right)}{\left(\sum_{k=1}^{N} e^{a_k}\right)^2} \\
&= \frac{e^{a_j}}{\sum_{k=1}^{N} e^{a_k}} \times \frac{\left(\sum_{k=1}^{N} e^{a_k} - e^{a_j}\right)}{\sum_{k=1}^{N} e^{a_k}} \\
&= p_i(1 - p_j)
\end{aligned}$$

Chain rule

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

- $y_2 = y_1 w_2 + b_2$
- $y_1 = x w_1 + b_1$
- $\frac{\partial y_2}{\partial w_1} = \frac{\partial f(y_1)}{\partial w_1} = \frac{\partial f(y_1)}{\partial y_1}\frac{\partial y_1}{\partial w_1} = w_2 * x$
  - $y_2 = (x w_1 + b_1) * \mathrm{w}_2 + \mathrm{b}_2$

Chain rule



- $\frac{\partial E}{\partial w_{jk}^1} = \frac{\partial E}{\partial O_k^1}\frac{\partial O_k^1}{\partial x} = \frac{\partial E}{\partial O_k^2}\frac{\partial O_k^2}{\partial O_k^1}\frac{\partial O_k^1}{\partial x}$

Chain rule

```
x = tf.constant(1.)
w1 = tf.constant(2.)
b1 = tf.constant(1.)
w2 = tf.constant(2.)
b2 = tf.constant(1.)

with tf.GradientTape(persistent=True) as tape:
    tape.watch([w1, b1, w2, b2])

    y1 = x * w1 + b1
    y2 = y1 * w2 + b2

dy2_dy1 = tape.gradient(y2, [y1])[0]
dy1_dw1 = tape.gradient(y1, [w1])[0]
dy2_dw1 = tape.gradient(y2, [w1])[0]
tf.Tensor(2.0, shape=(), dtype=float32)
tf.Tensor(2.0, shape=(), dtype=float32)
```

下一课时

多层感知机梯度

Thank You.