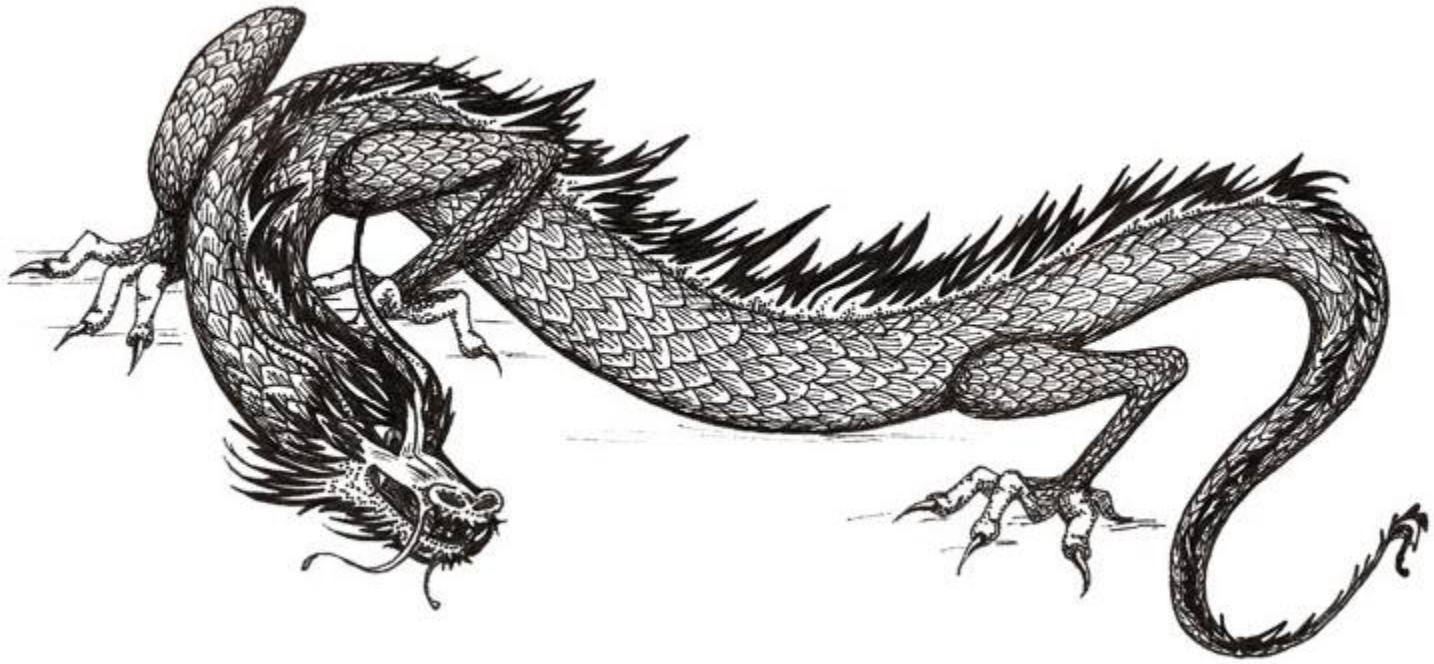


激活函数及其梯度

主讲：龙良曲

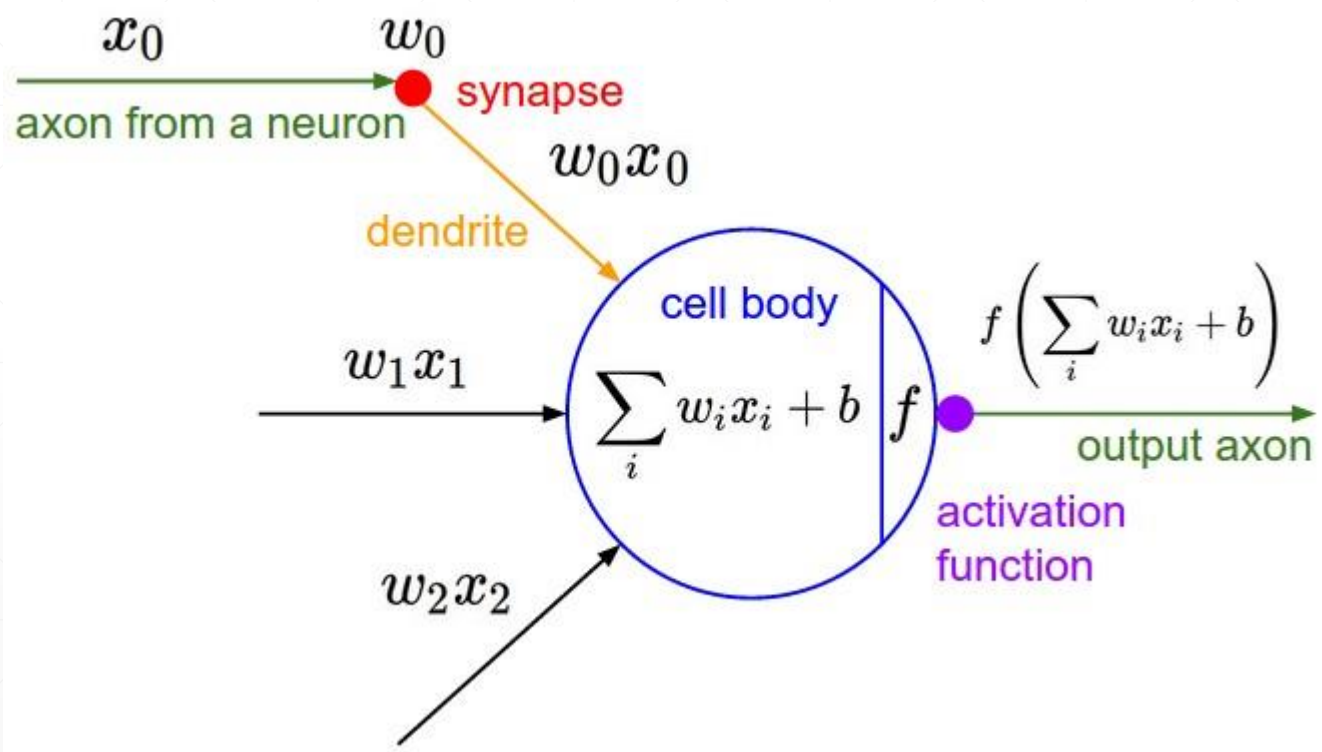
Outline

- sigmoid
- tanh
- relu

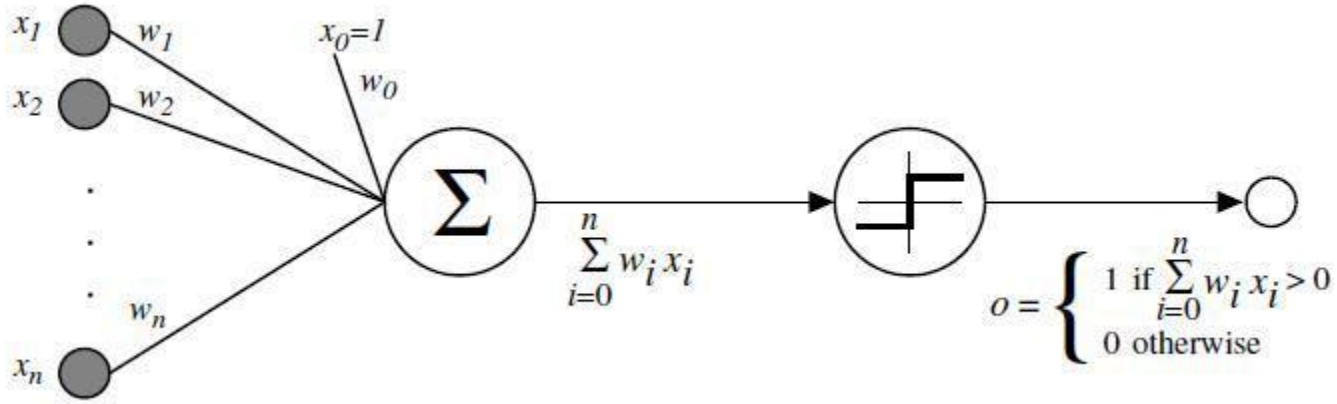
Activation Function

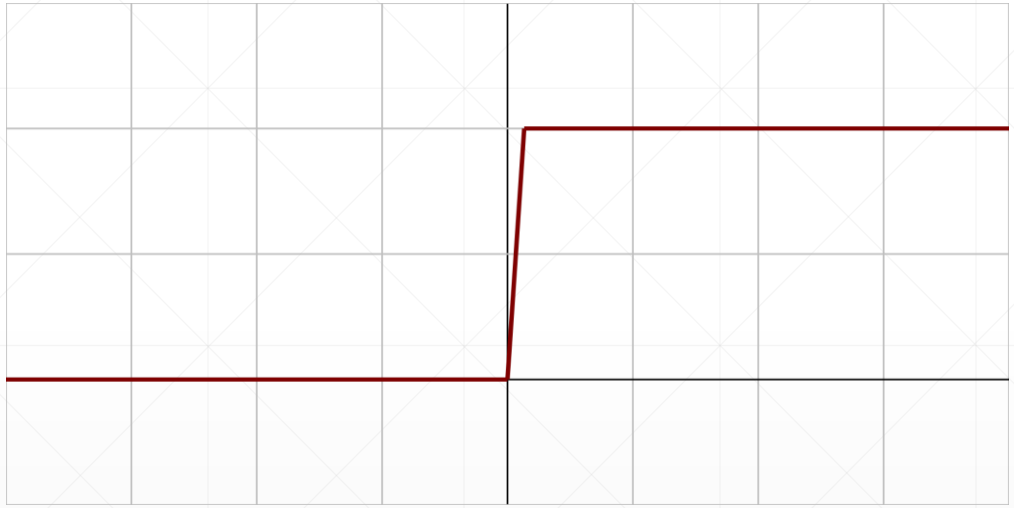
PITTS WITH LETTVIN: Pitts with Jerome Lettvin and one subject of their experiments on visual perception (1959).

Wikipedia



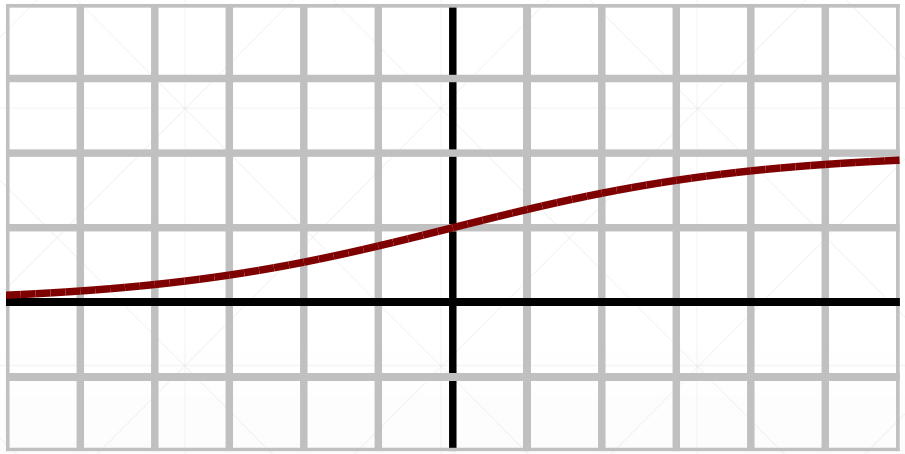
Derivative


x_1
x_2
x_n
w_1
w_2
w_n
$x_0=1$
w_0
Σ
$\sum_{i=0}^{n} w_i x_i$
$o = \begin{cases} 1 \text{ if } \sum_{i=0}^{n} w_i x_i > 0 \\ 0 \text{ otherwise} \end{cases}$



Sigmoid / Logistic

$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$$



Derivative

$$
\begin{aligned}
\frac{d}{dx}\sigma(x) &= \frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right) \\
&= \frac{e^{-x}}{(1+e^{-x})^2} \\
&= \frac{(1+e^{-x})-1}{(1+e^{-x})^2} \\
&= \frac{1+e^{-x}}{(1+e^{-x})^2} - \left(\frac{1}{1+e^{-x}}\right)^2 \\
&= \sigma(x) - \sigma(x)^2 \\
\sigma' &= \sigma(1-\sigma)
\end{aligned}
$$

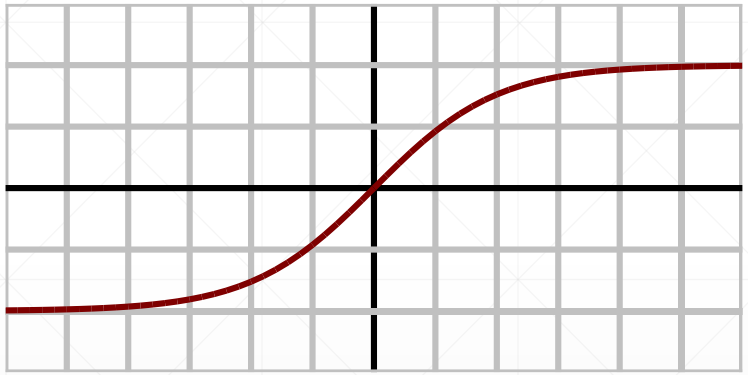
tf.sigmoid

```
a = tf.linspace(-10., 10., 10)

with tf.GradientTape() as tape:
    tape.watch(a)
    y = tf.sigmoid(a)

grads = tape.gradient(y, [a])

x: [-10.         -7.7777777  -5.5555553  -3.333333   -1.1111107   1.1111116
   3.333334    5.5555563   7.7777786  10.        ]
y: [4.5388937e-05 4.1878223e-04 3.8510561e-03 3.4445226e-02 2.4766389e-01
 7.5233626e-01 9.6555483e-01 9.9614894e-01 9.9958128e-01 9.9995458e-01]
grad: [4.5386874e-05 4.1860685e-04 3.8362255e-03 3.3258751e-02 1.8632649e-01
 1.8632641e-01 3.3258699e-02 3.8362255e-03 4.1854731e-04 4.5416677e-05]
```

Tanh

$$f(x) = \tanh(x) = \frac{(e^{x} - e^{-x})}{(e^{x} + e^{-x})}$$

$$= 2sigmoid(2x) - 1$$



Derivative

$$\frac{d}{dx}\tanh(x) = \frac{(e^x + e^{-x})(e^x + e^{-x}) - (e^x - e^{-x})(e^x - e^{-x})}{(e^x + e^{-x})^2}$$

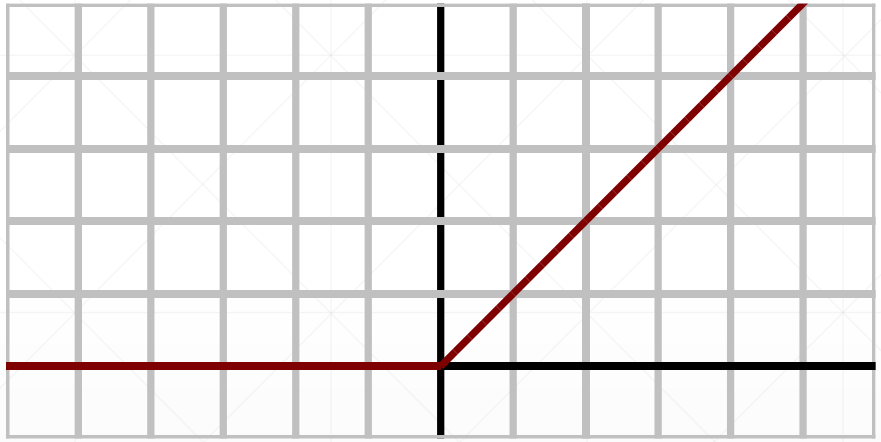
$$= 1 - \frac{(e^x - e^{-x})^2}{(e^x + e^{-x})^2} = 1 - \tanh^2(x)$$

tf.tanh

```
In [5]: a=tf.linspace(-5.,5.,10)

In [6]: tf.tanh(a)
Out[6]:
<tf.Tensor: id=10, shape=(10,), dtype=float32, numpy=
array([-0.9999092 , -0.9991625 , -0.99229795, -0.9311096 , -0.50467217,
        0.5046725 ,  0.93110967,  0.99229795,  0.9991625 ,  0.9999092 ],
      dtype=float32)>
```

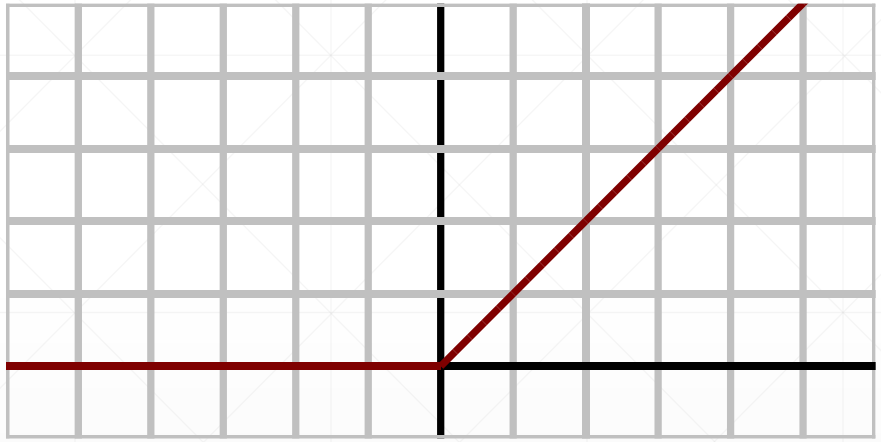
Rectified Linear Unit

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$



Derivative

$$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$$



tf.nn.relu

```
In [11]: a=tf.linspace(-1.,1.,10)

In [12]: tf.nn.relu(a
<tf.Tensor: id=24, shape=(10,), dtype=float32, numpy=
array([0.        , 0.        , 0.        , 0.        , 0.        ,
       0.11111116, 0.33333337, 0.5555556 , 0.7777778 , 1.        ],
      dtype=float32)>

In [13]: tf.nn.leaky_relu(a)
<tf.Tensor: id=26, shape=(10,), dtype=float32, numpy=
array([-0.2       , -0.15555556, -0.11111112, -0.06666666, -0.02222222,
        0.11111116,  0.33333337,  0.5555556 ,  0.7777778 ,  1.        ],
      dtype=float32)>
```

下一课时

损失函数及其梯度

Thank You.