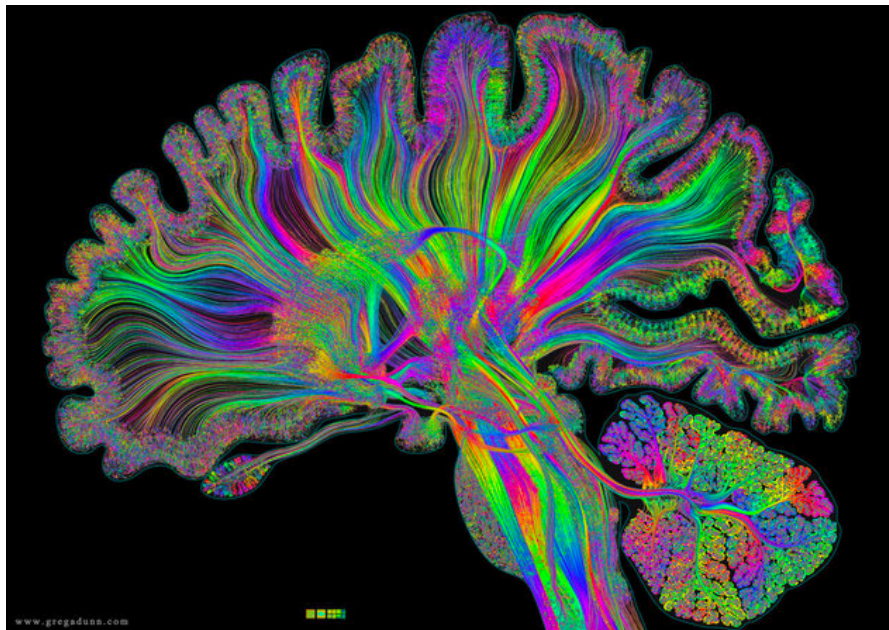
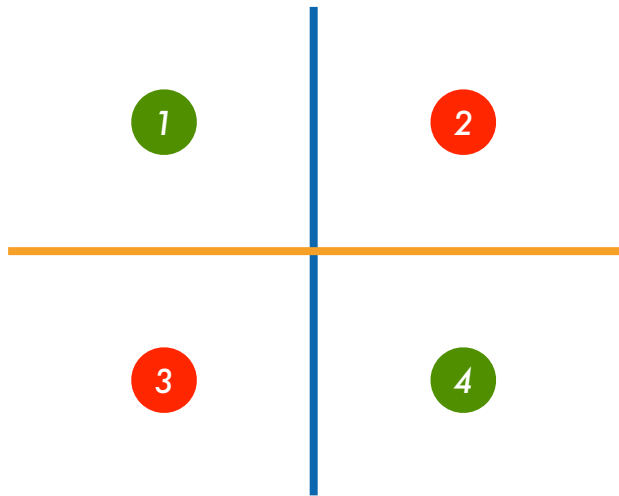




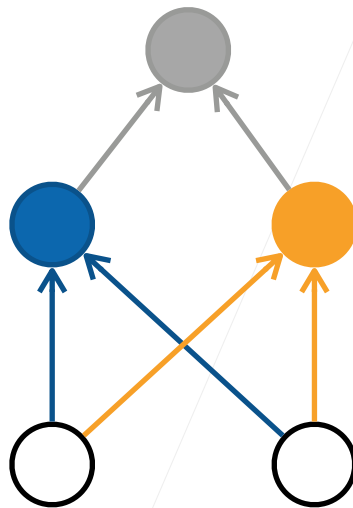
# 多层感知机



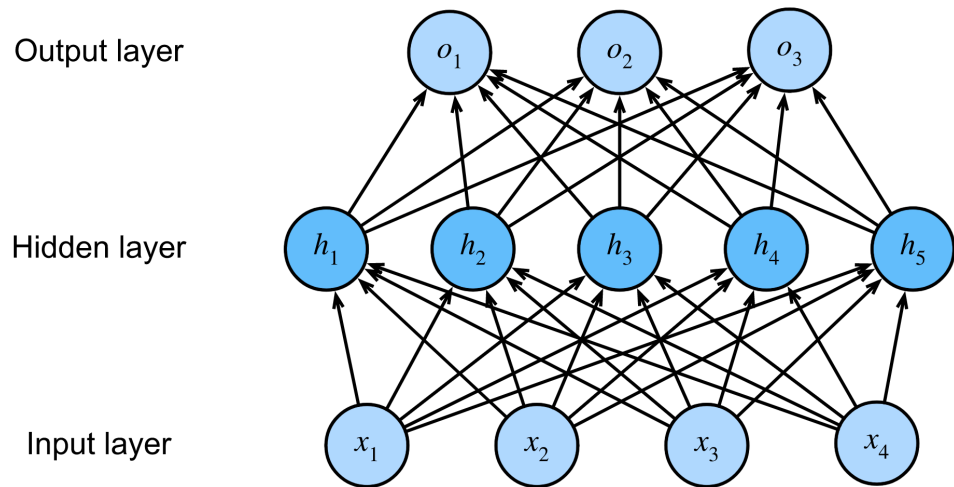
# 学习 XOR



	1	2	3	4
	+	-	+	-
	+	+	-	-
product	+	-	-	+



# 单隐藏层



隐藏层大小是超参数



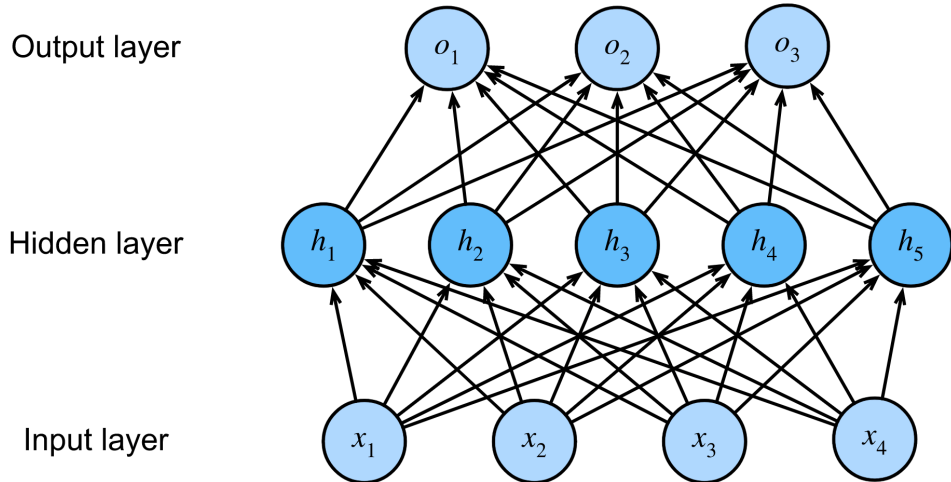
# 单隐藏层 — 单分类

- 输入  $\mathbf{x} \in \mathbb{R}^n$
- 隐藏层  $\mathbf{W}_1 \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b}_1 \in \mathbb{R}^m$
- 输出层  $\mathbf{w}_2 \in \mathbb{R}^m$ ,  $b_2 \in \mathbb{R}$

$$\mathbf{h} = \sigma(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1)$$

$$o = \mathbf{w}_2^T \mathbf{h} + b_2$$

$\sigma$  是按元素的激活函数



# 单隐藏层



- 输入  $\mathbf{x} \in \mathbb{R}^n$
- 隐藏层  $\mathbf{W}_1 \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b}_1 \in \mathbb{R}^m$
- 输出层  $\mathbf{w}_2 \in \mathbb{R}^m$ ,  $b_2 \in \mathbb{R}$

为什么需要非线性  
激活函数?

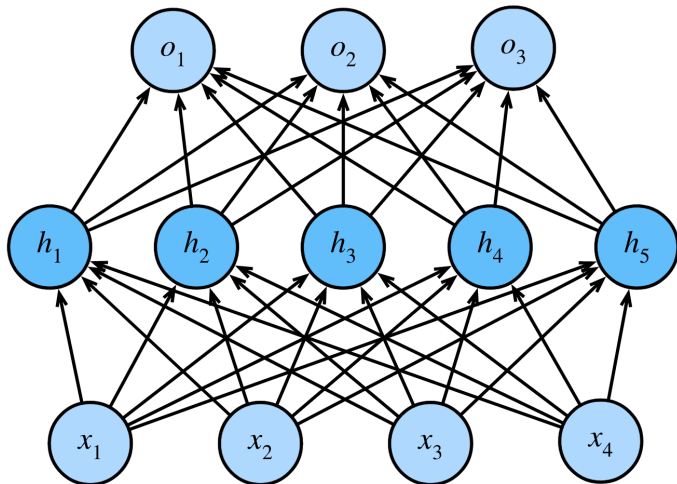
$$\mathbf{h} = \sigma(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1)$$

$$o = \mathbf{w}_2^T \mathbf{h} + b_2$$

$\sigma$  是按元素的激活函数

Hidden layer

Input layer



# 单隐藏层



- 输入  $\mathbf{x} \in \mathbb{R}^n$

为什么需要非线性  
激活函数?

- 隐藏层  $\mathbf{W}_1 \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b}_1 \in \mathbb{R}^m$

- 输出层  $\mathbf{w}_2 \in \mathbb{R}^m$ ,  $b_2 \in \mathbb{R}$

$$\mathbf{h} = \mathbf{W}_1 \mathbf{x} + \mathbf{b}_1$$

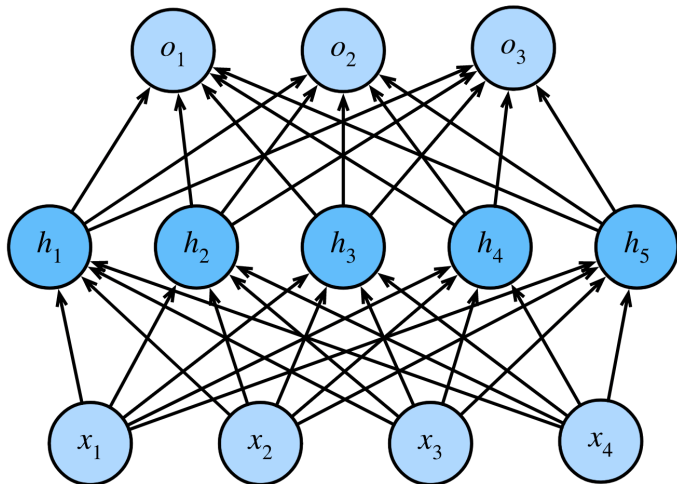
$$o = \mathbf{w}_2^T \mathbf{h} + b_2$$

$$\text{hence } o = \mathbf{w}_2^T \mathbf{W}_1 \mathbf{x} + b'$$

仍然是线性

Hidden layer

Input layer

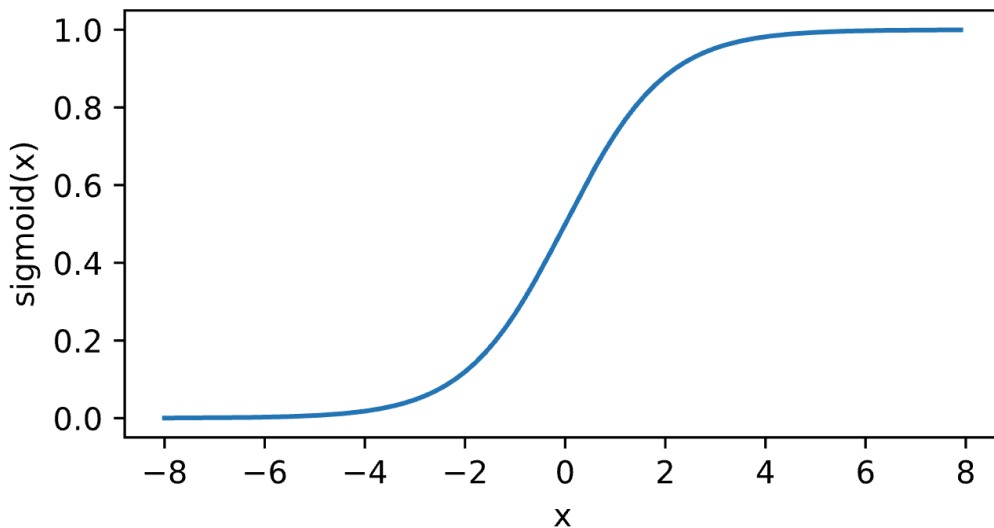




# Sigmoid 激活函数

将输入投影到  $(0, 1)$ , 是一个软的  $\sigma(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$

$$\text{sigmoid}(x) = \frac{1}{1 + \exp(-x)}$$

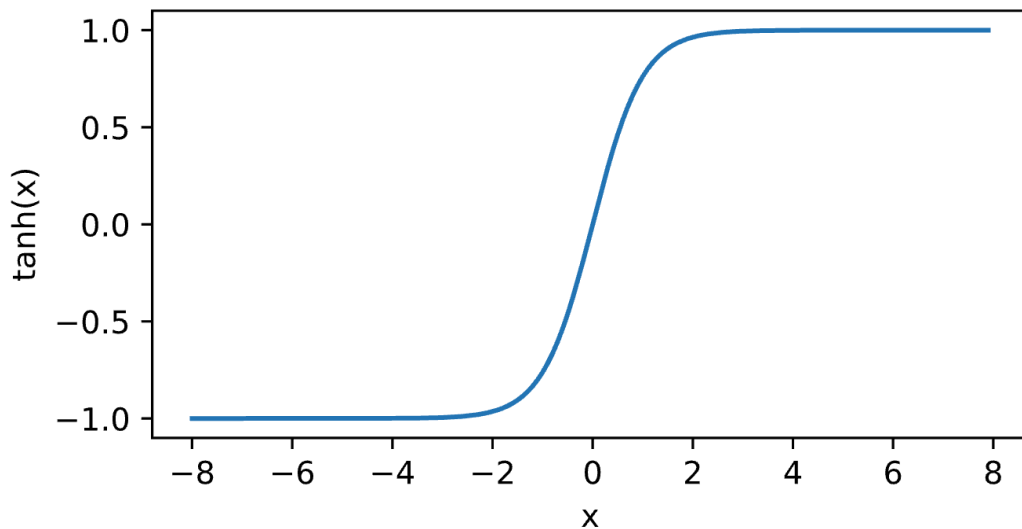


# Tanh 激活函数



将输入投影到  $(-1, 1)$

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



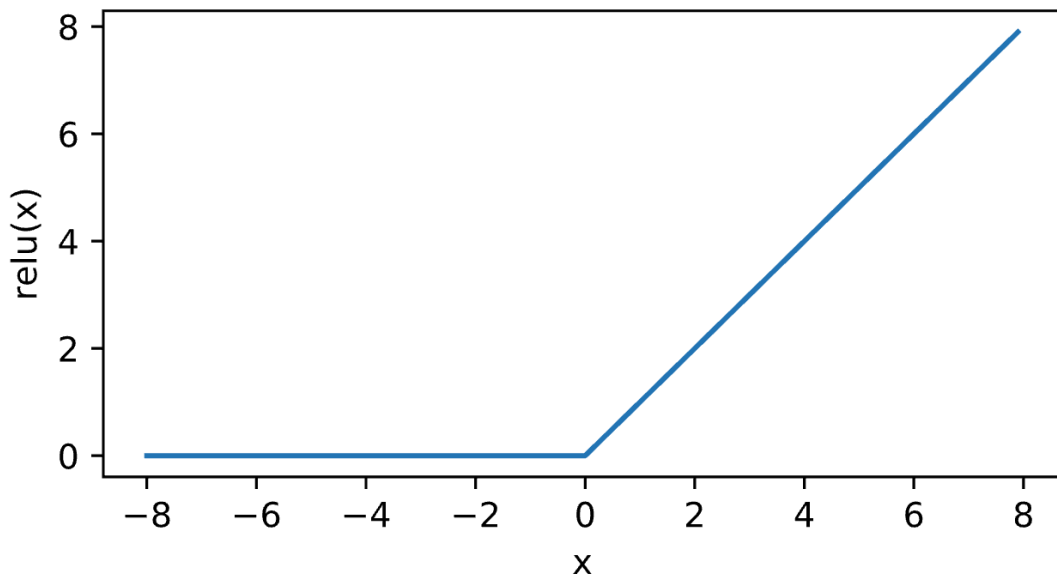


# ReLU 激活函数



ReLU: rectified linear unit

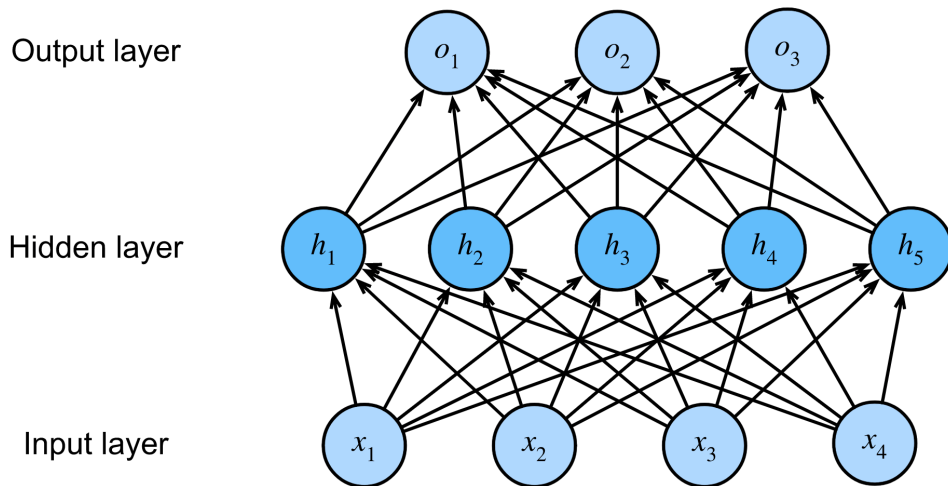
$$\text{ReLU}(x) = \max(x, 0)$$



# 多类分类



$$y_1, y_2, \dots, y_k = \text{softmax}(o_1, o_2, \dots, o_k)$$



# 多类分类



- 输入  $\mathbf{x} \in \mathbb{R}^n$
- 隐藏层  $\mathbf{W}_1 \in \mathbb{R}^{m \times n}, \mathbf{b}_1 \in \mathbb{R}^m$
- 输出层  $\mathbf{W}_2 \in \mathbb{R}^{m \times k}, \mathbf{b}_2 \in \mathbb{R}^k$

$$\mathbf{h} = \sigma(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1)$$

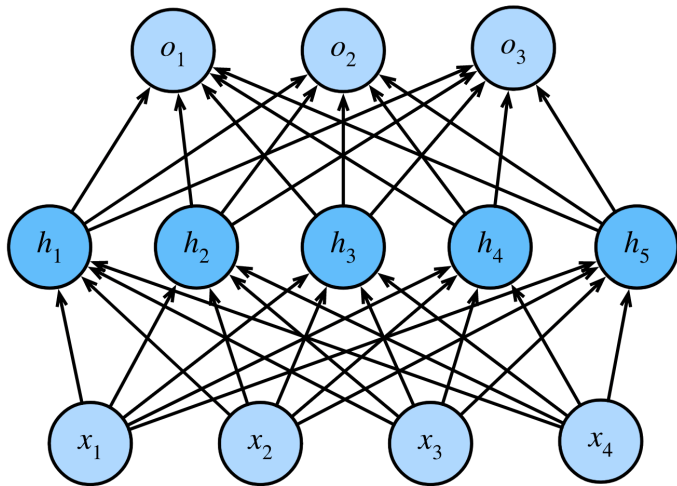
$$\mathbf{o} = \mathbf{W}_2^T \mathbf{h} + \mathbf{b}_2$$

$$\mathbf{y} = \text{softmax}(\mathbf{o})$$

Output layer

Hidden layer

Input layer



# 多隐藏层

$$\mathbf{h}_1 = \sigma(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1)$$

$$\mathbf{h}_2 = \sigma(\mathbf{W}_2 \mathbf{h}_1 + \mathbf{b}_2)$$

$$\mathbf{h}_3 = \sigma(\mathbf{W}_3 \mathbf{h}_2 + \mathbf{b}_3)$$

$$\mathbf{o} = \mathbf{W}_4 \mathbf{h}_3 + \mathbf{b}_4$$

## 超参数

- 隐藏层数
- 每层隐藏层的大小

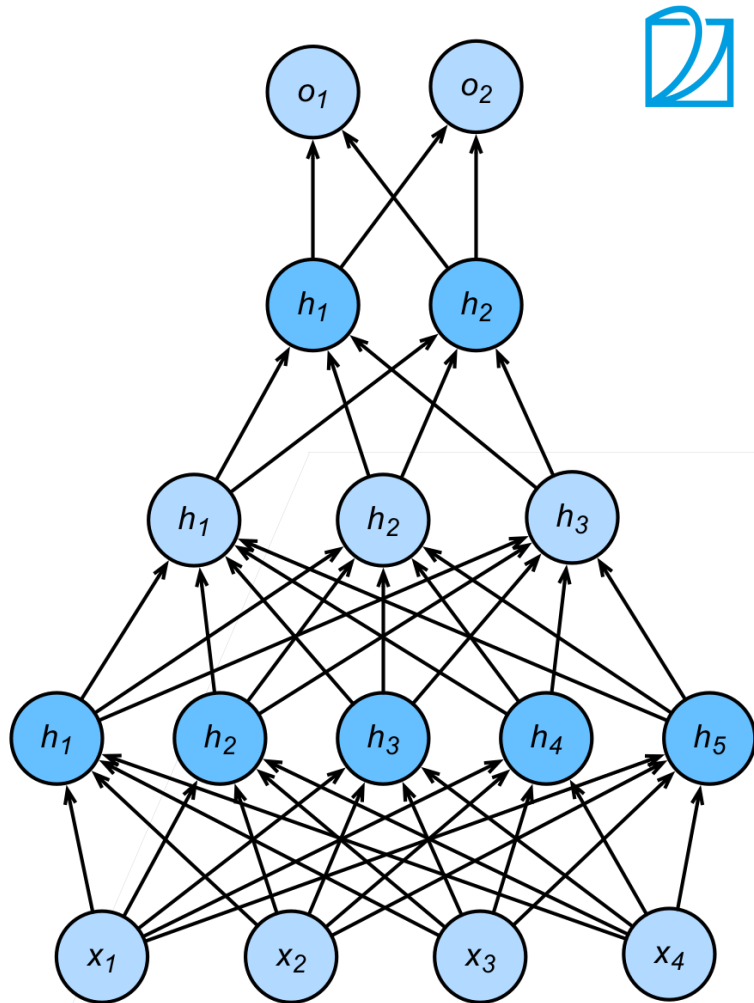
Output layer

Hidden layer

Hidden layer

Hidden layer

Input layer



# 总结



- 多层感知机使用隐藏层和激活函数来得到非线性模型
- 常用激活函数是Sigmoid, Tanh, ReLU
- 使用 Softmax 来处理多类分类
- 超参数为隐藏层数, 和各个隐藏层大小