

# 2-7 Softmax 回归

王中雷

厦门大学王亚南经济研究院和经济学院, 2025

# 内容摘要

1. 模型

2. 前向传播

3. 后向传播

4. 例子

# 介绍

## 1. 二分类模型

- $y \in \{0, 1\}$
- 模型通常用来估计条件概率  $P(y = 1 \mid \mathbf{x})$

## 2. 实际中，类别数往往大于 2，记为 $K$ ： $y \in \{0, 1, \dots, K - 1\}$

- 考虑独热编码 (one-hot representation)
- $\mathbf{y} = (0, \dots, 1, \dots, 0)^T$
- 如果  $y = k$ ,  $\mathbf{y}$  只有  $(k + 1)$  位置的元素为 1，其他位置均为 0

# 模型

## 1. 对于特征向量 $x$

- 我们考虑一个有 $K$  个输出的神经网络模型
- 每一个输出对应于特定类别的“得分”（例如，条件概率）
- 所有“得分”加和不一定等于 1

## 2. 与二分类模型唯一的不同之处在于输出层神经元的个数

- 对于二分类问题，输出层只有一个神经元
- 对于一般的分类问题，输出层往往包含 $d^{[L]} = K$ 个神经元

# 模型

## 1. 回顾

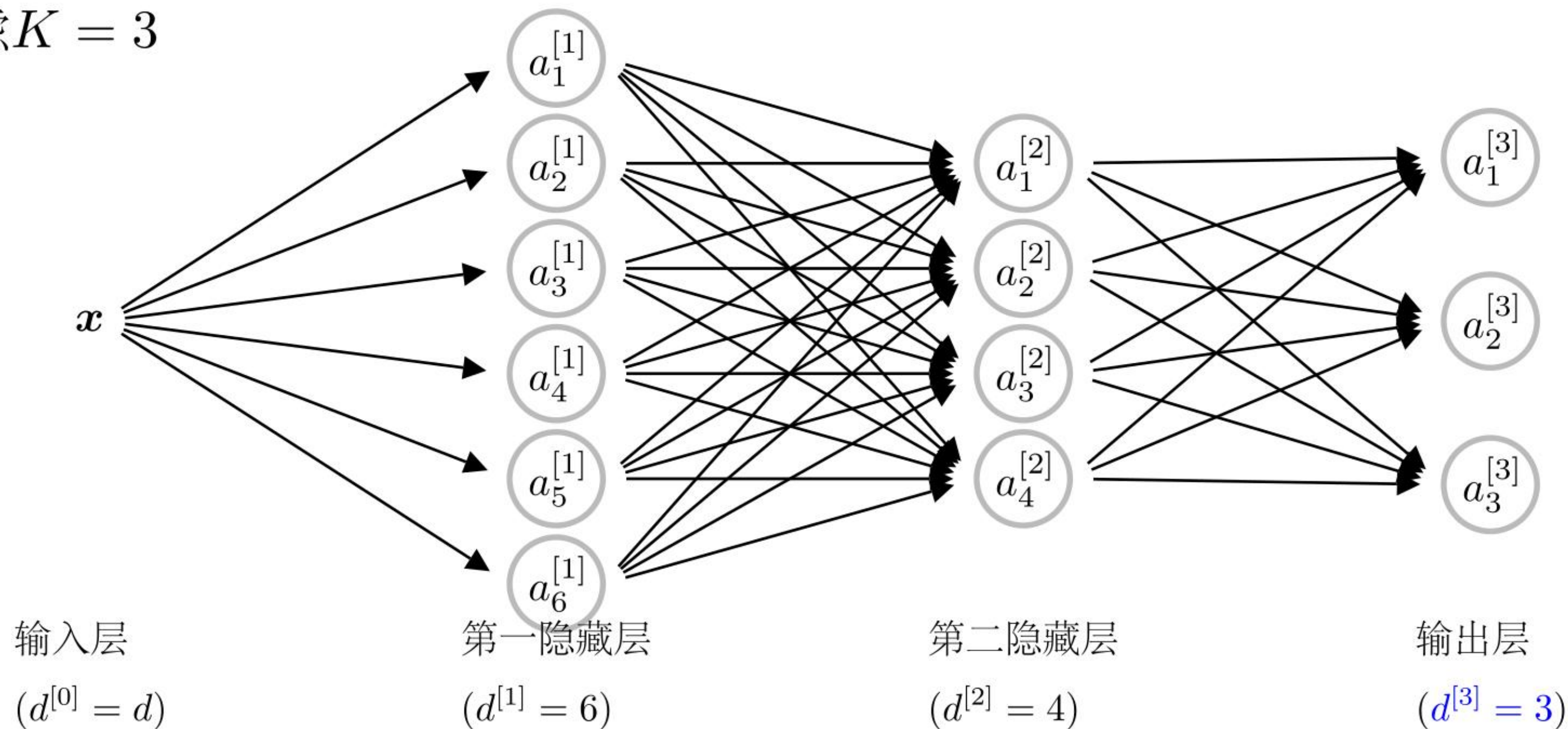
- $L$  : 神经网络模型中的层数
- $d^{[l]}$  : 第 $l$  层神经元个数 ( $l = 0, \dots, L$ )
- $\mathbf{a}^{[l]} = (a_1^{[l]}, \dots, a_{d^{[l]}}^{[l]})^T \in \mathbb{R}^{d^{[l]} \times 1}$
- $\mathbf{W}^{[l]} = (\mathbf{w}_1^{[l]}, \dots, \mathbf{w}_{d^{[l]}}^{[l]})^T \in \mathbb{R}^{d^{[l]} \times d^{[l-1]}}$
- $\mathbf{b}^{[l]} = (b_1^{[l]}, \dots, b_{d^{[l]}}^{[l]})^T \in \mathbb{R}^{d^{[l]} \times 1}$

2. 对于二分类问题, 我们有 $d^{[L]} = 1$

3. 对于一般分类问题, 我们有 $d^{[L]} = K$

# 模型

1. 考虑  $K = 3$



# 前向传播

1. 令  $\mathbf{A}^{[0]} = \mathbf{X}$

2. 对于  $l = 1, \dots, L$ ,

$$\begin{aligned}\mathbf{Z}^{[l]} &= (\mathbf{b}^{[l]})^T + \mathbf{A}^{[l-1]} (\mathbf{W}^{[l]})^T \\ \mathbf{A}^{[l]} &= \sigma^{[l]}(\mathbf{Z}^{[l]})\end{aligned}$$

- $\sigma^{[l]}(z)$ : 第  $l$  层对应的激活函数
- Python 中的广播机制用于计算激活值

3. 对于最后一层（输出层），

$$\mathbf{W}^{[L]} \in \mathbb{R}^{K \times d_{L-1}}, \quad \mathbf{b} \in \mathbb{R}^{K \times 1}$$

# 前向传播

1. Softmax 回归对应的代价函数为

$$\mathcal{J}(\boldsymbol{\theta}) = -n^{-1} \sum_{i=1}^n \sum_{k=1}^K y_{ik} \log a_{ik}^{[L]}$$

- $\boldsymbol{\theta}$  : 模型参数
- $\mathbf{y}_i = (y_{i1}, \dots, y_{iK})^T$  : 对应于第 $i$  个样本点的独热编码
- $a_{ik}^{[L]}$  : 对第 $i$  个训练样本, 其标签为 $k$  的条件概率的估计

2. 输出层对应的 $\mathbf{A}^{[L]}$  的维度

- $\mathbf{A}^{[L]} \in \mathbb{R}^{n \times 1}$  : 二分类问题
- $\mathbf{A}^{[L]} \in \mathbb{R}^{n \times K}$  : 一般的 softmax 回归问题

# 后向传播

1.  $\mathbf{dA}^{[L]}$  可通过代价函数得到

2. 假设  $\mathbf{dA}^{[l]}$  是已知的 ( $l = L, \dots, 2$ )

$$\mathbf{dZ}^{[l]} = \mathbf{dA}^{[l]} \circ \sigma^{[l]'}(\mathbf{Z}^{[l]})$$

$$\mathbf{dW}^{[l]} = (\mathbf{dZ}^{[l]})^T \mathbf{dA}^{[l-1]}$$

$$\mathbf{db}^{[l]} = (\mathbf{dZ}^{[l]})^T \mathbf{1}$$

$$\mathbf{dA}^{[l-1]} = \mathbf{dZ}^{[l]} \mathbf{W}^{[l]}$$

3. 对于一般 softmax 回归问题, 我们只需要讨论如何得到  $\mathbf{dA}^{[L]}$

# 后向传播

1. 代价函数为

$$\mathcal{J}(\boldsymbol{\theta}) = -n^{-1} \sum_{i=1}^n \sum_{k=1}^K y_{ik} \log a_{ik}^{[L]}$$

2. 因此,  $\mathrm{d}\mathbf{A}^{[L]}$  的第  $(i, k)$  位置的元素为

$$\frac{\partial \mathcal{J}}{\partial a_{ik}^{[L]}} = -\frac{y_{ik}}{a_{ik}^{[L]}}$$

3. 完成!

# 例子

测试图片



神经网络模型（两个隐藏层）

估计结果

