

第05章 前馈神经网络

欧新宇



Softmax 回归

/ Softmax回归简介

/ Softmax回归的基本优化

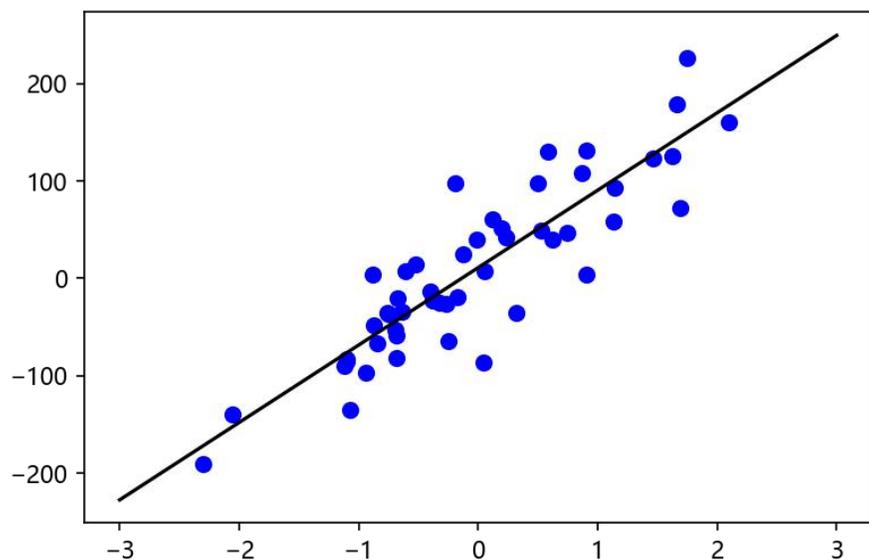




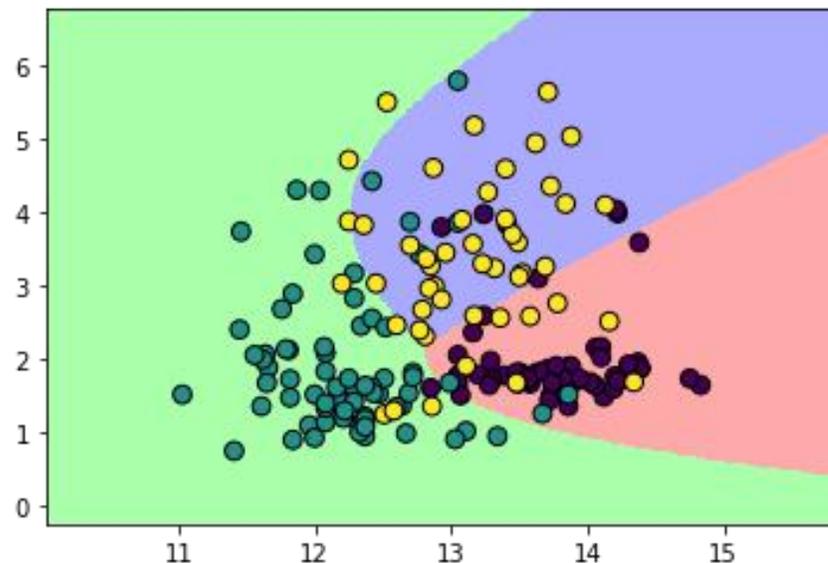
Softmax回归简介

从回归到分类

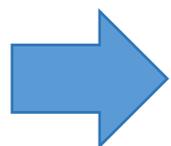
回归估计连续值



分类预测离散类别



多类别分类



Cat ?

0.03

Dog ?

0.05

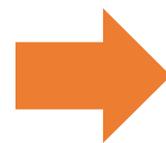
Rabbit?

0.91

Rabbit!

fish?

0.01



Softmax回归简介

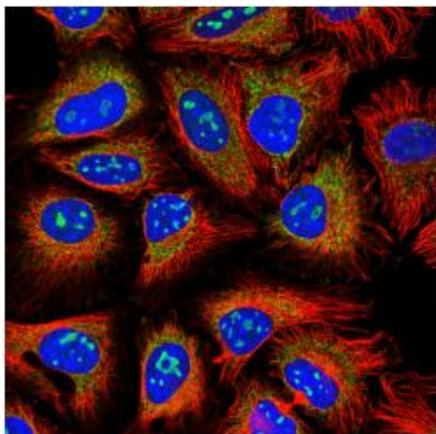
多类别分类



MNIST手写数字识别 (10类)



ImageNet自然物体分类 (1000类)



0. Nucleoplasm
1. Nuclear membrane
2. Nucleoli
3. Nucleoli fibrillar center
4. Nuclear speckles
5. Nuclear bodies
6. Endoplasmic reticulum
7. Golgi apparatus
8. Peroxisomes
9. Endosomes
10. Lysosomes

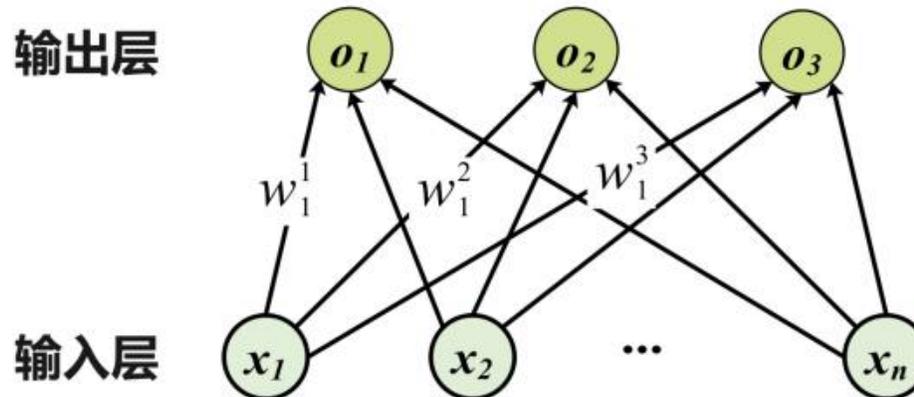
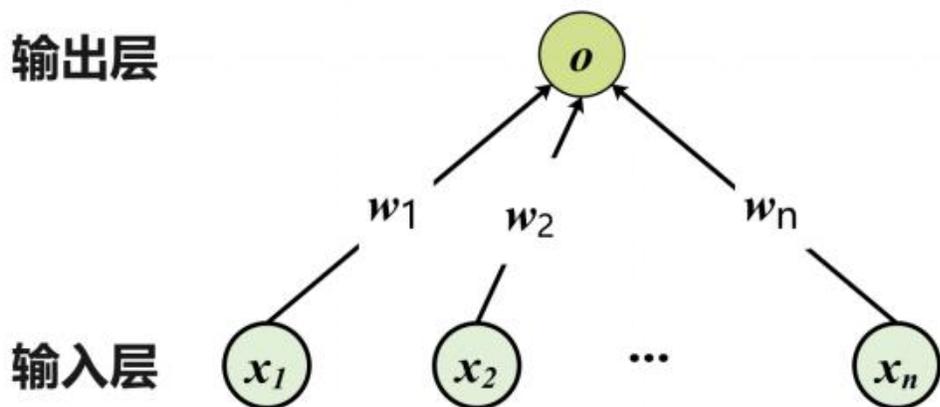
显微镜下的人类蛋白质分类 (28类)



NASA小行星分类

Softmax回归简介

从回归到多类别分类



回归

- ✓ 单连续数值输出
- ✓ 自然区间 \mathbb{R}
- ✓ 预测值与真实值的距离作为损失

多类别分类

- ✓ 通常有**多个**输出
- ✓ 输出 i 是对第 i 个类的置信度预测



Softmax回归的基本优化

Softmax回归的基本优化

均方损失

- ✓ 对类别进行一位有效的Onehot编码

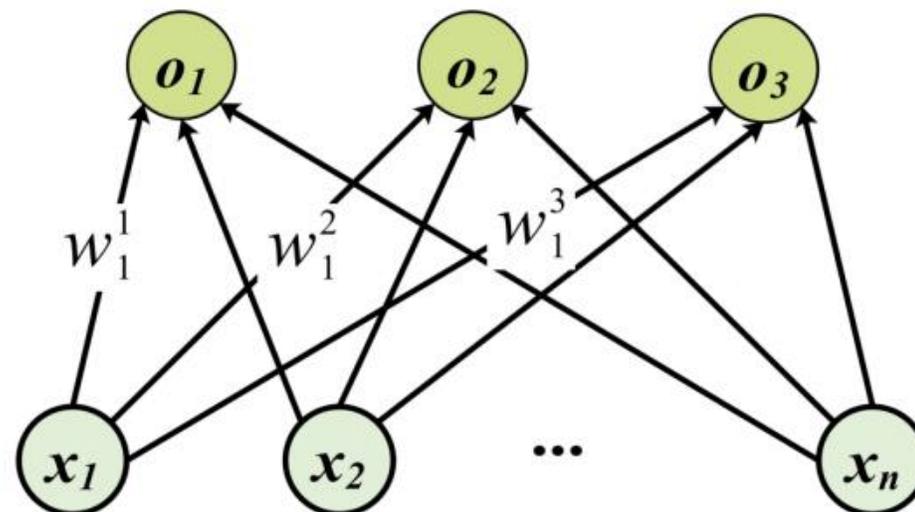
$$y = [y_1, y_2, \dots, y_n]^T$$

$$y_1 = \begin{cases} 1 & , \text{if } i = y \\ 0 & , \text{otherwise} \end{cases}$$

- ✓ 使用均方误差损失函数进行训练

- ✓ 最大值最优预测

$$\hat{y} = \underset{i}{\operatorname{argmax}} o_i$$



Softmax回归的基本优化

无校验比例 vs 校验比例

无校验比例

- ✓ 对类别进行Onehot有效编码
- ✓ 最大值最优预测

$$\hat{y} = \operatorname{argmax}_i o_i$$

- ✓ 需要更置信地识别正确类

$$o_y - o_i \geq \Delta(y, i)$$

校验比例

- ✓ 输出匹配概率 (非负, 和为1)

$$\hat{y} = \operatorname{softmax}(\mathbf{o})$$

$$\hat{y} = \frac{\exp(o_i)}{\sum_k \exp(o_k)}$$

- ✓ 概率 \hat{y} 和 y 的区别作为损失:

$$\text{loss} = |\hat{y} - y|$$

Softmax回归的基本优化

Softmax和交叉熵损失

- 交叉熵常用来衡量两个概率的区别：

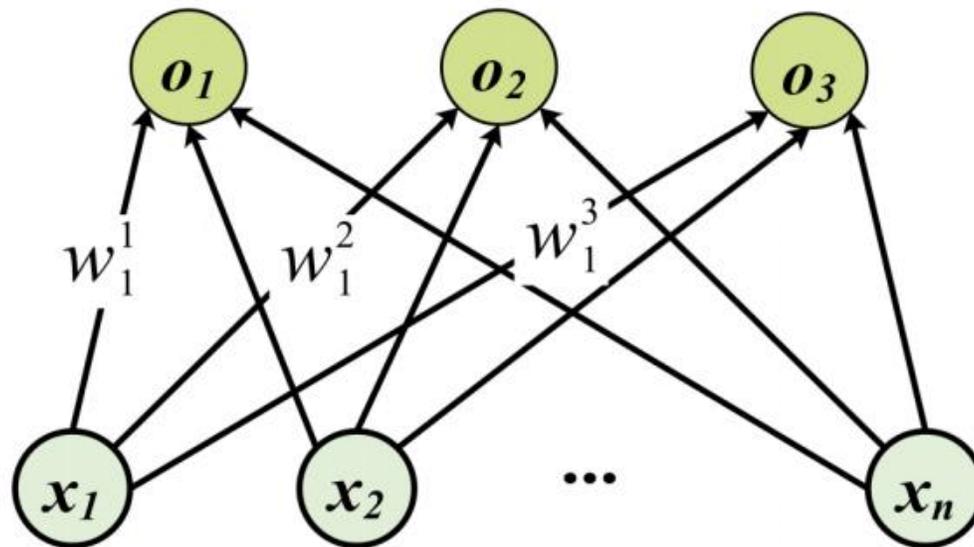
$$H(p, q) = - \sum_i p_i \log(q_i)$$

- 基于交叉熵的损失

$$l(\mathbf{y}, \hat{\mathbf{y}}) = - \sum_i y_i \log(\hat{y}_i) = - \log(\hat{\mathbf{y}})$$

- 交叉熵的梯度是真实概率和预测概率的区别

$$\partial_{o_i} l(\mathbf{y}, \hat{\mathbf{y}}) = \text{softmax}(o)_i - y_i$$



小 结

- **Softmax回归**是典型的**多分类模型**
- 使用Softmax算子可以得到**每个类**的预测置信度
- Softmax的**总损失**等于**所有类别的损失之和**
- 使用**交叉熵**来衡量**预测和标签**的区别



邓肯·卢斯 (1925–2012) 美国科学院院士，
认知心理学家、数学家、社会学家。
Softmax函数发明者。

读万卷书 行万里路 只为最好的修炼

QQ: 14777591 (宇宙骑士)

Email: ouxinyu@alumni.hust.edu.cn

Website: <http://ouxinyu.cn>

Tel: 18687840023