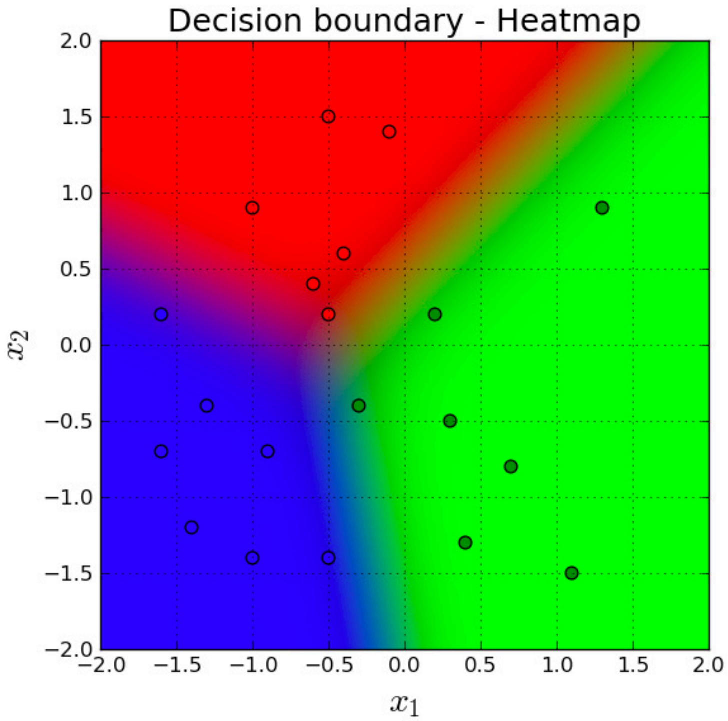
动手学深度学习 v2

李沐・AWS

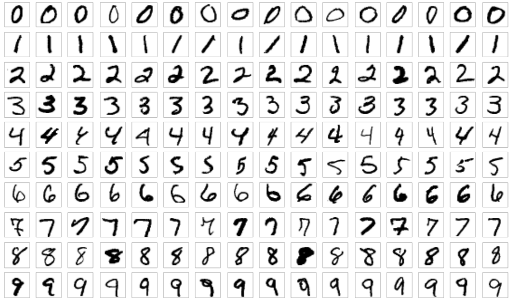
Softmax 回归



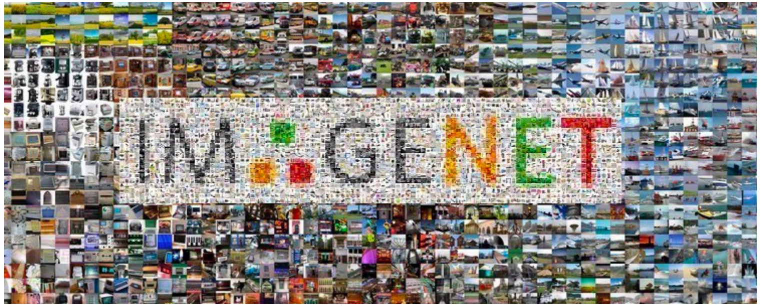
回归 vs 分类

- 回归估计一个连续值
- 分类预测一个离散类别

MNIST: 手写数字识别（10类）

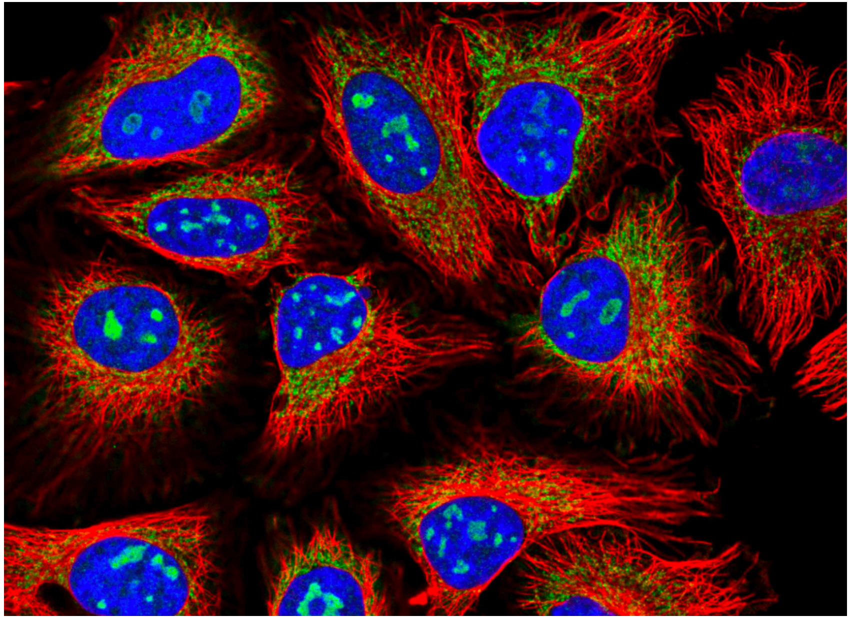


ImageNet: 自然物体分类（1000类）



Kaggle 上的分类问题

将人类蛋白质显微镜图片分成28类



```
0.  Nucleoplasm
1.  Nuclear membrane
2.  Nucleoli
3.  Nucleoli fibrillar
4.  Nuclear speckles
5.  Nuclear bodies
6.  Endoplasmic reticu
7.  Golgi apparatus
8.  Peroxisomes
9.  Endosomes
10.  Lysosomes
11.  Intermediate fila
12.  Actin filaments
13.  Focal adhesion si
14.  Microtubules
15.  Microtubule ends
16.  Cytokinetic bridg
```

https://www.kaggle.com/c/human-protein-atlas-image-classification

Kaggle 上的分类问题

将恶意软件分成9个类别

https://www.kaggle.com/c/malware-classification

Kaggle 上的分类问题

将恶意的 Wikipedia 评论分成 7 类

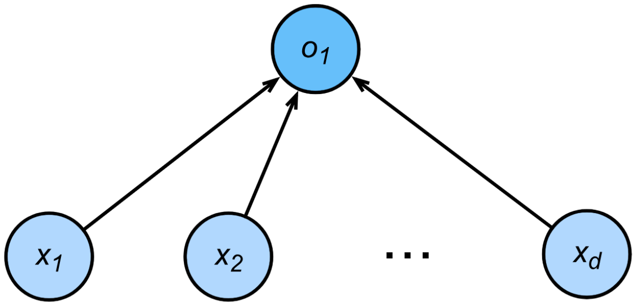
comment_text	toxic	severe_toxic	obsc
Explanation\nWhy the edits made under my usern...	0	0	0
D'aww! He matches this background colour I'm s...	0	0	0
Hey man, I'm really not trying to edit war. It...	0	0	0
"\nMore\nI can't make any real suggestions on ...	0	0	0
You, sir, are my hero. Any chance you remember...	0	0	0

https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge

从回归到多类分类

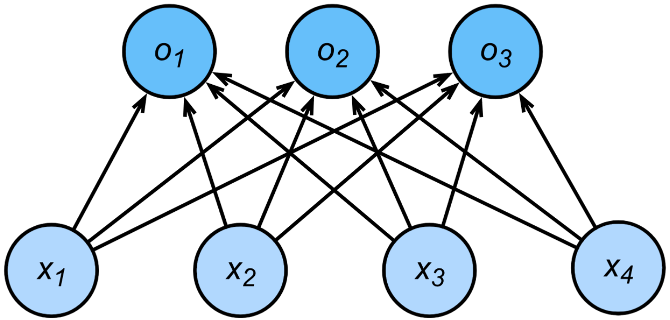
回归

- 单连续数值输出
- 自然区间 $\mathbb{R}$
- 跟真实值的区别作为损失



分类

- 通常多个输出
- 输出 i 是预测为第 i 类的置信度



从回归到多类分类 — 均方损失

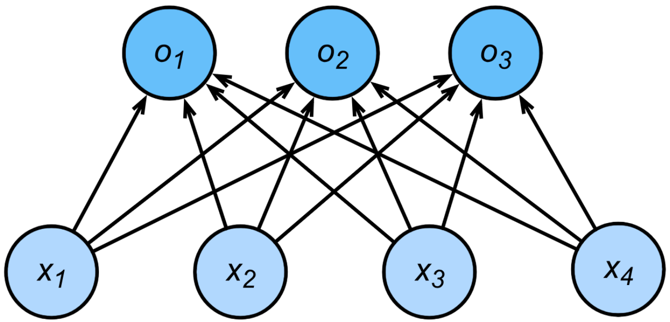
- 对类别进行一位有效编码

$$\mathbf{y} = [y_1, y_2, \ldots, y_n]^\top$$

$$y_i = \begin{cases} 1 \text{ if } i = y \\ 0 \text{ otherwise} \end{cases}$$

- 使用均方损失训练
- 最大值为预测

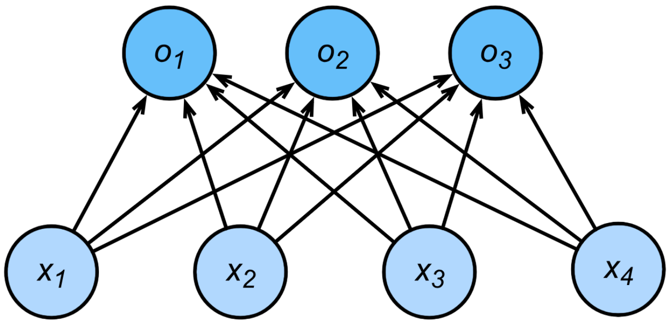
$$\hat{y} = \underset{i}{\operatorname{argmax}}\, o_i$$



从回归到多类分类 — 无校验比例

- 对类别进行一位有效编码
- 最大值最为预测

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



- 需要更置信的识别正确类（大余量）

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

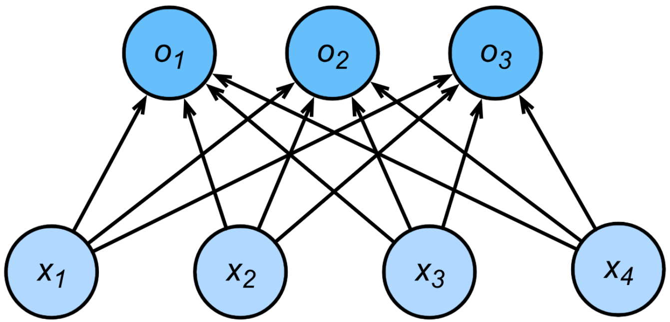
从回归到多类分类 — 校验比例

- 输出匹配概率（非负，和为 1）

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

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

- 概率 $\mathbf{y}$ 和 $\hat{\mathbf{y}}$ 的区别作为损失



Softmax 和交叉熵损失

- 交叉熵常用来衡量两个概率的区别 $H(\mathbf{p}, \mathbf{q}) = \sum_i -p_i \log(q_i)$
- 将它作为损失

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

- 其梯度是真实概率和预测概率的区别

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

总结

- Softmax 回归是一个多类分类模型
- 使用 Softmax 操作子得到每个类的预测置信度
- 使用交叉熵来来衡量预测和标号的区别