

A low-angle, upward-looking photograph of several modern skyscrapers. The buildings feature glass facades and white structural elements, creating a sense of height and architectural complexity. The sky is a clear, vibrant blue. A dark green horizontal band is superimposed across the middle of the image, containing the chapter title.

第七章

机器学习简介



基本概念

- 什么是机器学习 (machine learning)?

- ✓ 找到一个函数 (function)

- 图像分类 (image classification)

$$f(\text{image of a cat}) = \text{"cat"}$$



- 语音识别 (speech recognition)

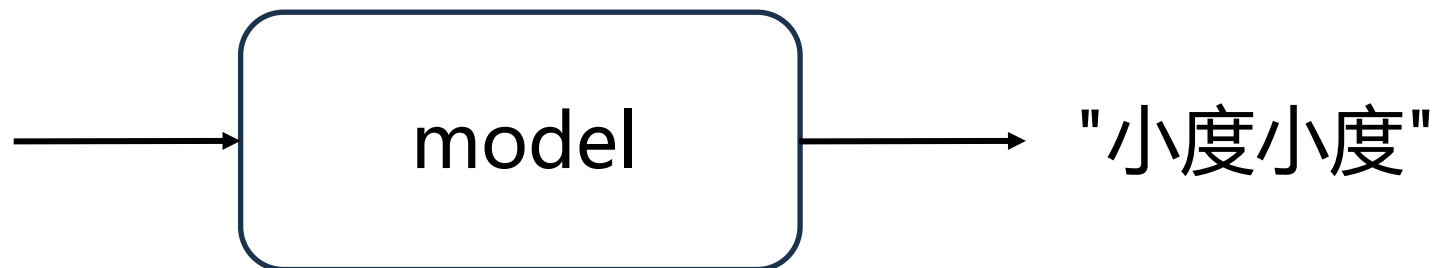
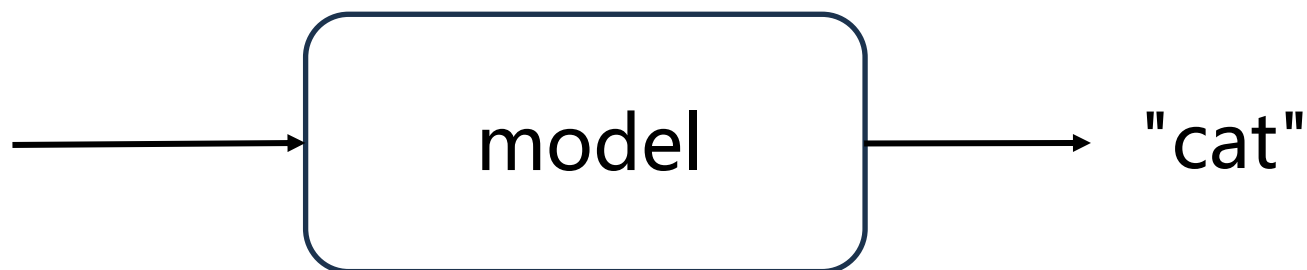
$$f(\text{audio waveform}) = \text{"小度小度"}$$





基本概念

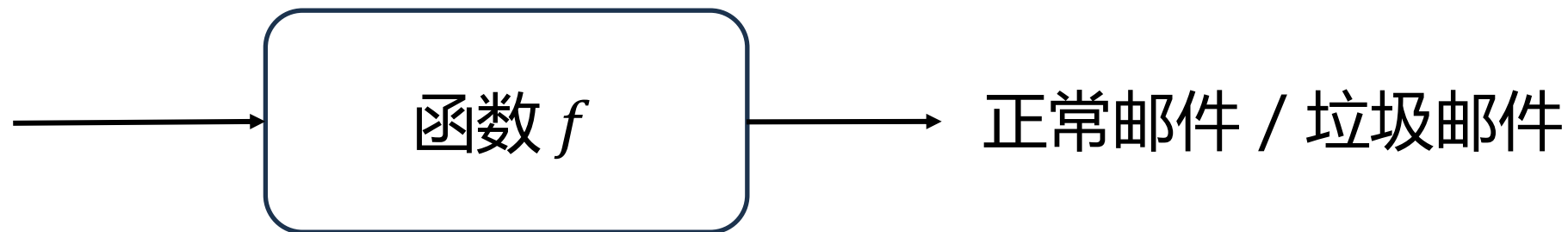
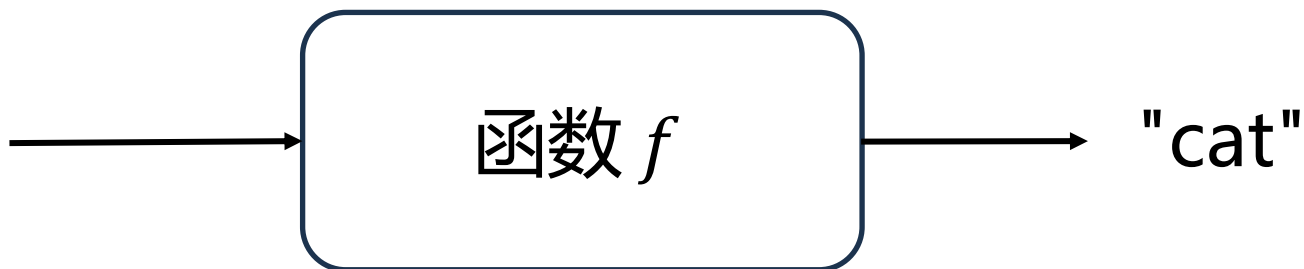
- 函数(function), 模型(model), 算法 (algorithm), 系统(system)





- 函数类型 / 任务 (task) 类型

- ✓ 分类 (classification): 给定若干个类别 (class / category / label), 函数需要输出正确的类别





基本概念



湖北大学
HUBEI UNIVERSITY

- **本课程：**分类函数只能输出**预设的**类别
- **真实世界：**开集识别 (open-set recognition)



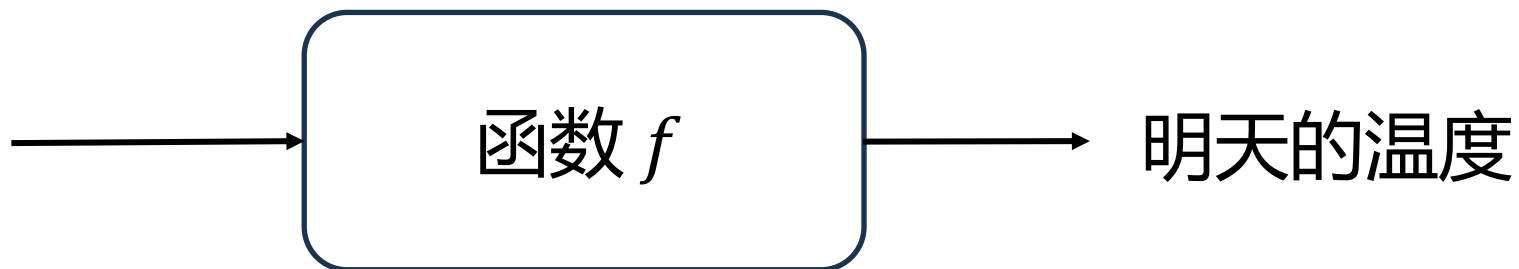
- 狗：95%
- 猫：3%
- 鼠：1%
- 猪：1%



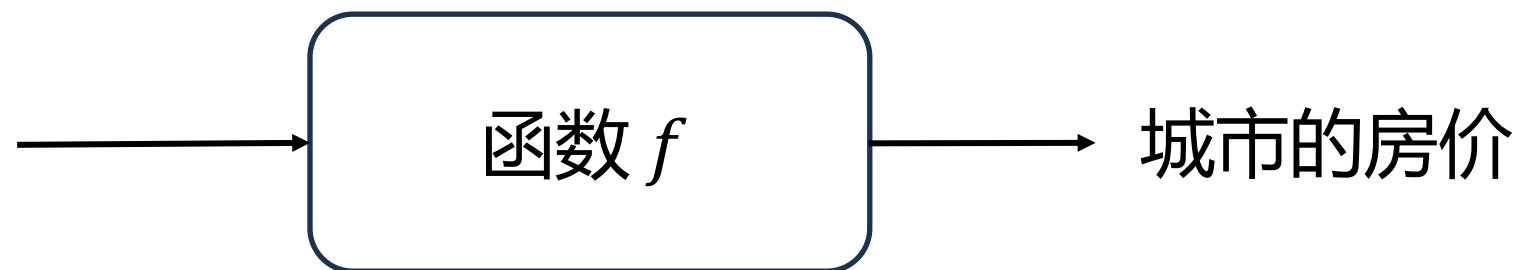
- 函数类型 / 任务 (task) 类型

- ✓ 回归 (regression): 函数会输出一个标量 (scalar)

1. 今天的温度
2. 今天的湿度
3. 今天的降水量

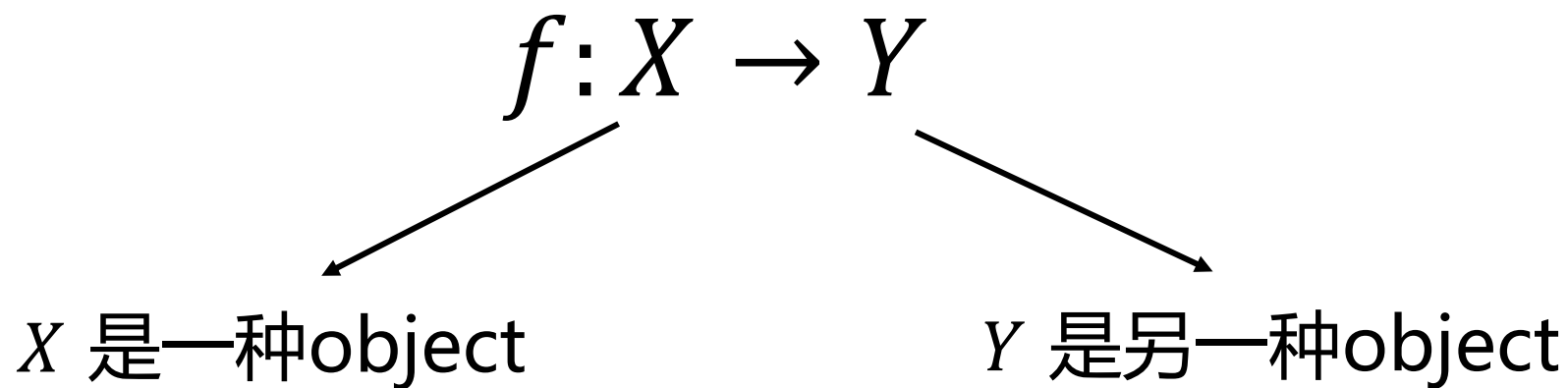


1. 城市的地理位置
2. 城市的就业情况
3. 几线城市





- 函数类型 / 任务 (task) 类型
 - ✓ 结构化学习(structured learning)



- **Object**: 序列(sequence), 列表(list), 树(tree), ...
- 分类和回归任务的输入和输出都是向量(vector)



基本概念

- 结构化学习的应用场景



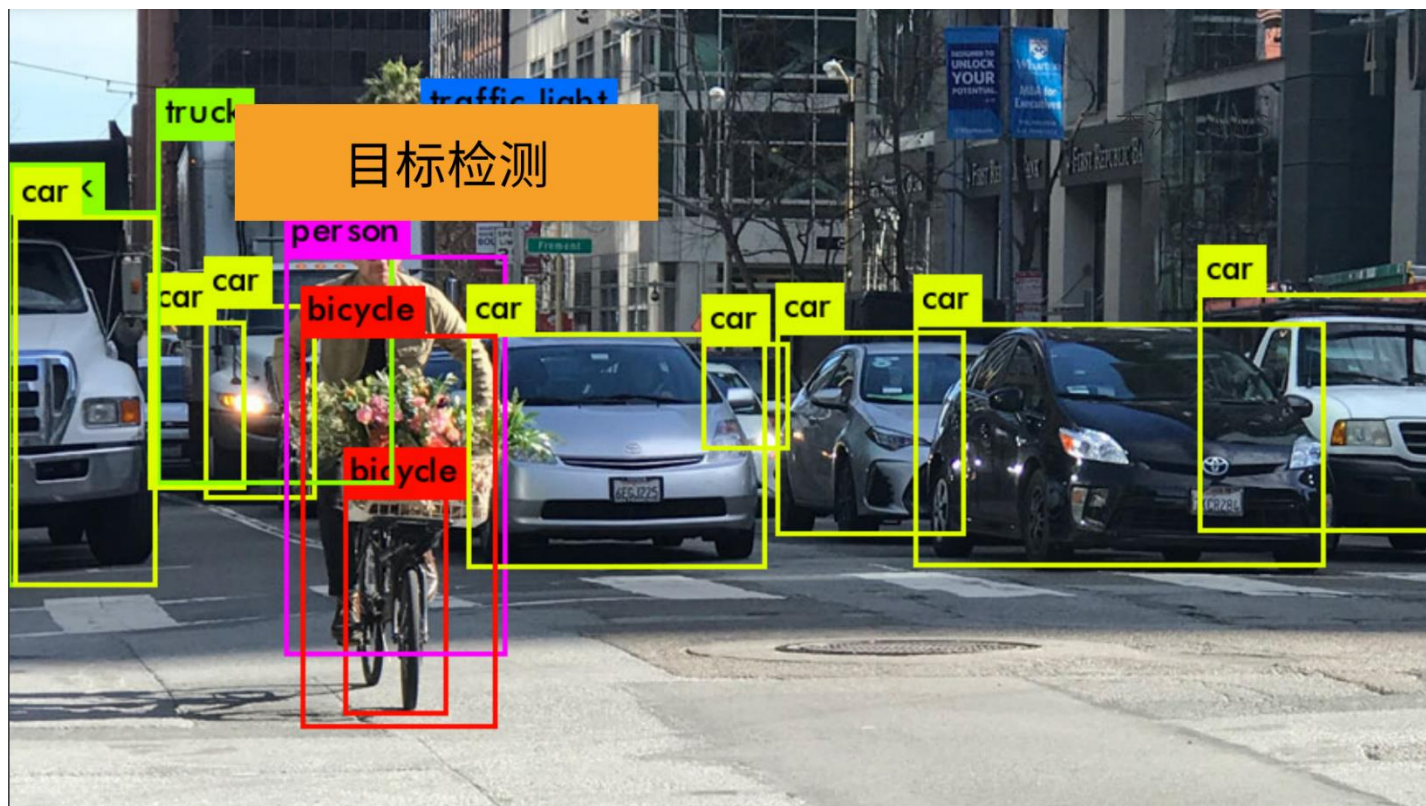
湖北大学
HUBEI UNIVERSITY

新闻文本
(sequence)

向量化

朴素贝叶斯
分类器

新闻类别



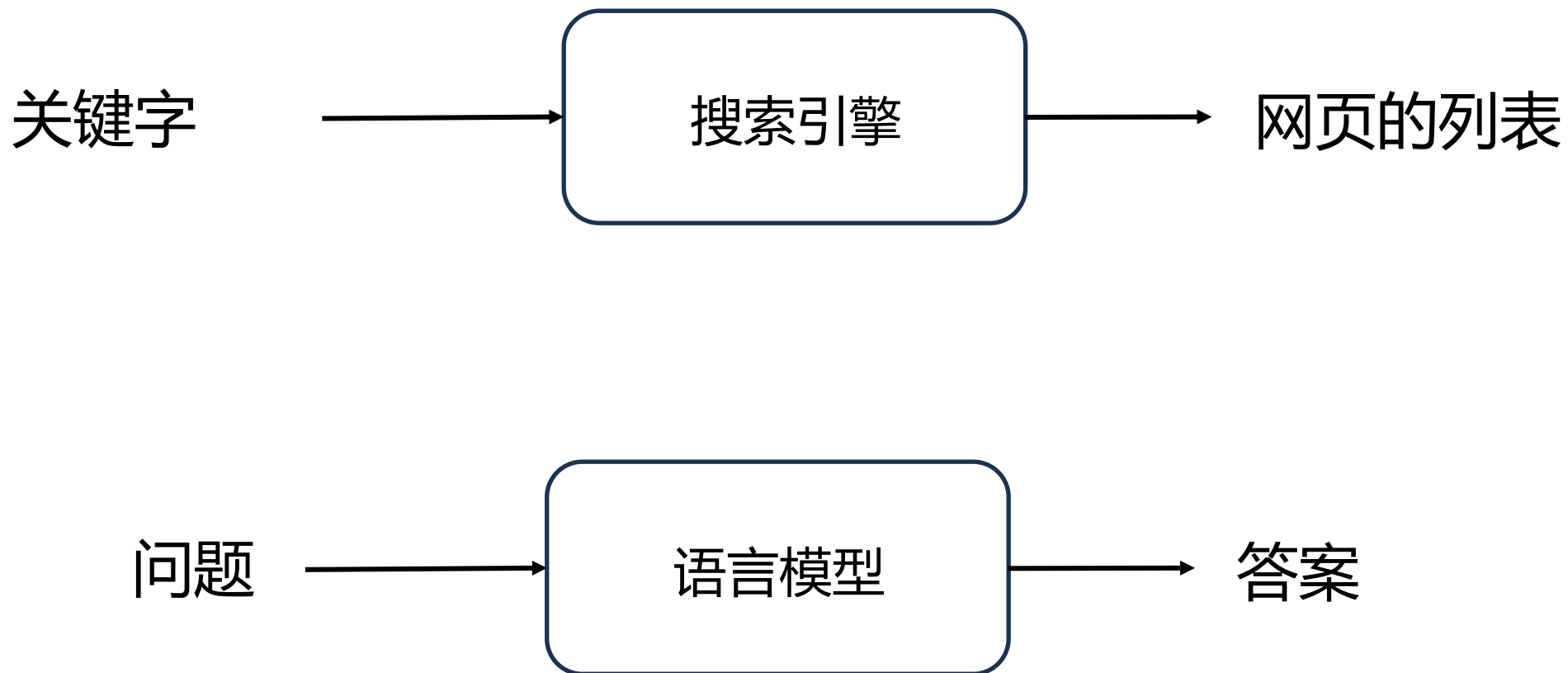
目标检测 (Object Detection)

X: 图片 → Y: bounding box



基本概念

- 结构化学习的应用场景






机器学习的步骤

- **步骤一**：选择 / 设计带有**未知参数 (parameter)** 的函数

今天的温度 = $f(\text{昨天的温度})$

模型 (model) $y = w x_1 + b$ 基于**领域知识 (domain knowledge)** 选择一个函数

今天的温度 昨天的温度 特征 (feature)



w, b 是未知参数，需要从数据中学习获得 (**训练, training**)

权重 (weight) 偏置 (bias)



机器学习的步骤



湖北大学
HUBEI UNIVERSITY

- 步骤二：定义损失函数 (loss function)

损失函数的值反应的是模型的预测结果和真实结果是否接近

$$loss = L(w, b)$$

$$y = wx_1 + b$$

$$y = 1 \times x_1 + (-1)$$

参数初始化

04/15 周一	04/16 周二	04/17 周三	04/18 周四	04/19 周五	04/20 周六
多云	小雨	中雨	多云	中雨	阴
无持续风向	无持续风向	无持续风向	无持续风向	无持续风向	西南风
微风	微风	微风	微风	微风	微风
30℃	28℃	23℃	25℃	24℃	25℃

$$1 \times 30 + (-1) = 29 = y \quad \hat{y}$$

$$1 \times 24 + (-1) = 23 = y \quad \hat{y}$$

$$\text{损失函数值 } e = |y - \hat{y}| = 1$$

$$e = |y - \hat{y}| = 2$$



机器学习的步骤



湖北大学
HUBEI UNIVERSITY

- **步骤二：** 定义损失函数 (loss function)

$$L = \frac{1}{N} \sum_{i=1}^N e_i$$

$$e = |y - \hat{y}|$$

平均绝对误差 (mean absolute error, MAE)

$$e = (y - \hat{y})^2$$

均方误差 (mean square error, MSE)



机器学习的步骤



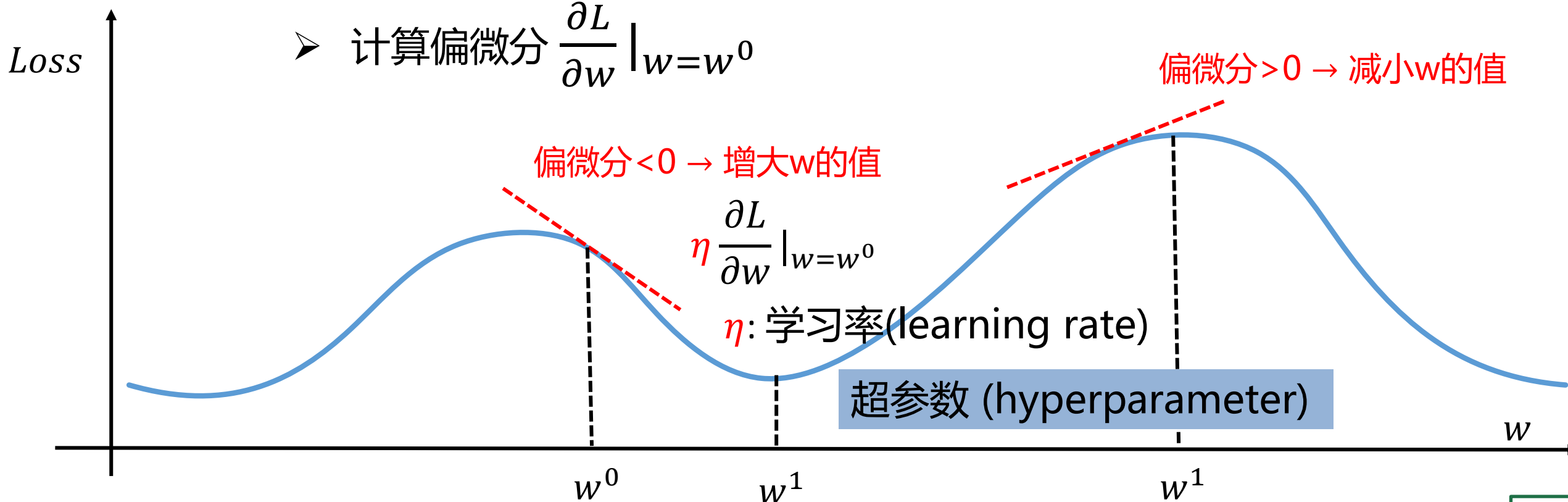
湖北大学
HUBEI UNIVERSITY

- 步骤三：优化 (optimization) $w^*, b^* = \operatorname{argmin}_{w, b} L$

- ✓ 梯度下降法 (gradient descent)

- 随机初始化 w^0

- 计算偏微分 $\frac{\partial L}{\partial w} \big|_{w=w^0}$





- 超参数(hyperparameter)

- ✓ 自己设定, 不需要通过学习获得
- ✓ 做研究时, 需要进行超参分析

5.3.4. Hyperparameter Analysis

The DGS framework refers to two hyperparameters: the size of the edge classifier, and compression ratio. We study the influence of these hyperparameters on edge query task using the dataset *subelj_jung* and *slashdot*, and the results are shown in Figures 8 and 9.

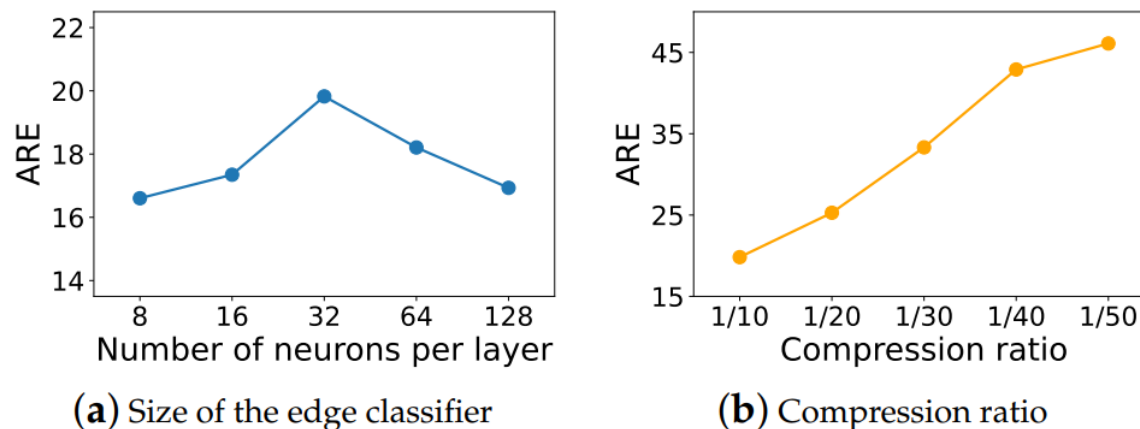


Figure 8. Hyperparameter analysis on edge query task using dataset *subelj_jung*.

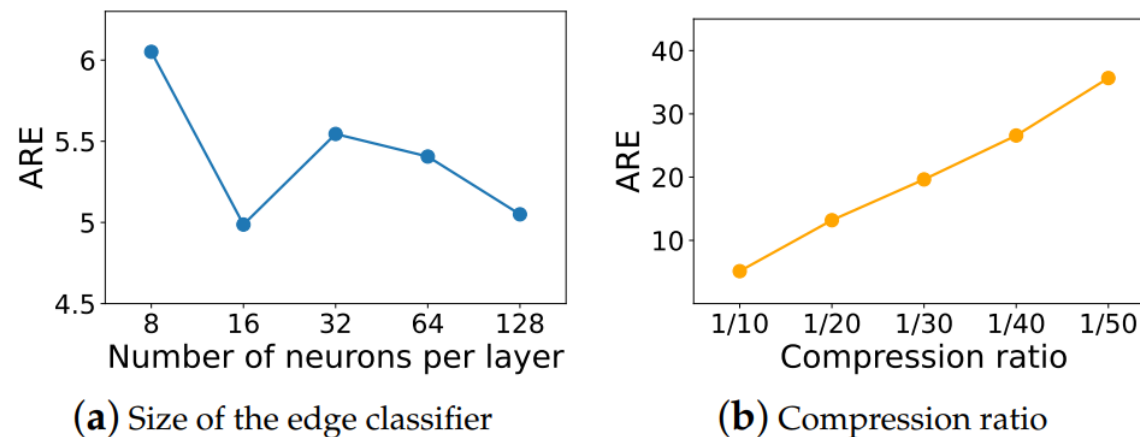


Figure 9. Hyperparameter analysis on edge query task using dataset *slashdot*.



机器学习的步骤



湖北大学
HUBEI UNIVERSITY

- 步骤三：优化 (optimization) $w^*, b^* = \operatorname{argmin}_{w, b} L$

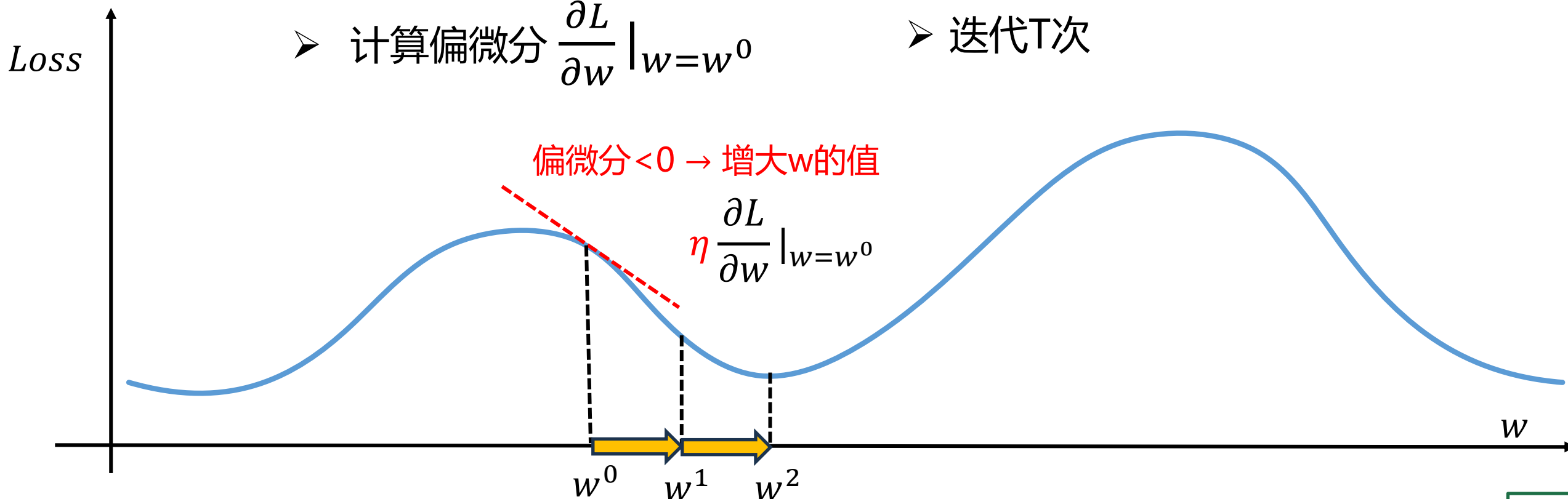
- ✓ 梯度下降法 (gradient descent)

- 随机初始化 w^0

- $w^1 \leftarrow w^0 - \eta \frac{\partial L}{\partial w} \big|_{w=w^0}$

- 计算偏微分 $\frac{\partial L}{\partial w} \big|_{w=w^0}$

- 迭代T次





机器学习的步骤



湖北大学
HUBEI UNIVERSITY

✓ 梯度下降法 (gradient descent)

$$w^*, b^* = \operatorname{argmin}_{w, b} L$$

- 随机初始化 w^0, b^0
- 计算偏微分

$$\frac{\partial L}{\partial w} \Big|_{w=w^0, b=b^0}$$

$$\frac{\partial L}{\partial b} \Big|_{w=w^0, b=b^0}$$

$$w^1 \leftarrow w^0 - \eta \frac{\partial L}{\partial w} \Big|_{w=w^0, b=b^0}$$

$$b^1 \leftarrow b^0 - \eta \frac{\partial L}{\partial b} \Big|_{w=w^0, b=b^0}$$

直接调用深度学习库PyTorch中的方法进行计算 (实验六)

- 迭代T次, 更新w和b



机器学习的步骤



湖北大学
HUBEI UNIVERSITY



训练 (training)

$$y = \mathbf{w}x_1 + b$$

$$y = \mathbf{w}_1x_1 + \mathbf{w}_2x_2 + \mathbf{w}_3x_3 + \cdots + b$$

一般写成向量相乘的形式：

$$y = \mathbf{w}^T \cdot \mathbf{x} + b$$

线性模型 (linear model)

线性回归 (linear regression)



机器学习的步骤



湖北大学
HUBEI UNIVERSITY

$$y = w_1x_1 + w_2x_2 + w_3x_3 + \cdots + b$$

一般写成向量相乘的形式：

$$y = \mathbf{w}^T \cdot \mathbf{x} + b$$

训练集：

$$\mathbf{x}_1 = [x_{11}, x_{12}, x_{13}, \dots]^T$$

$$\mathbf{x}_2 = [x_{21}, x_{22}, x_{23}, \dots]^T$$

$$\mathbf{x}_3 = [x_{31}, x_{32}, x_{33}, \dots]^T$$

...

一般会把整个训练集写成矩阵的形式：

$$\mathbf{y}^T = \mathbf{w}^T \mathbf{X} + b$$

结束语



谢谢!