

# 第一章 机器学习

## 1.1 机器学习简介

人工智能是我们想要达到的目标，机器学习则是实现人工智能的手段，深度学习则是机器学习的其中一种。

那么机器学习是什么？机器学习可以看做是从数据中学习一个函数 (function)，对于给定输入得到输出结果。如在语音辨识、图像识别等领域的应用。

机器学习框架如图 1.1 所示，首先包含一系列函数 model 的集合，利用训练数据评价函数的品质，并挑选出最优函数模型。

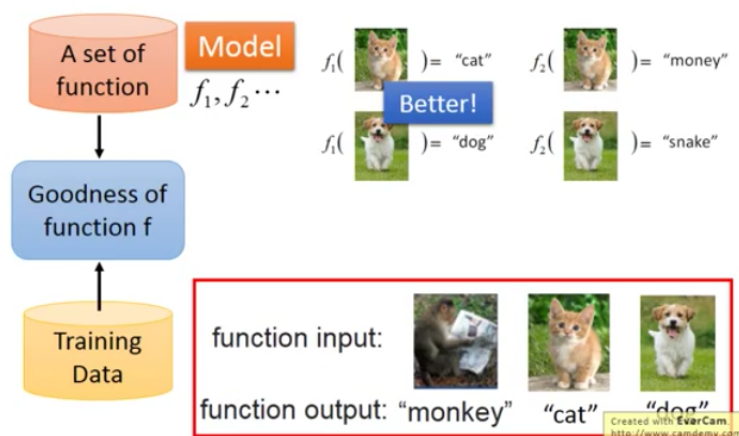


图 1.1: 机器学习框架

详细步骤如图 1.2 所示，可以总结为：

1. 挑选模型
2. 评价函数品质 goodness
3. 挑选最优函数  $f^*$

机器学习的学习图谱如图 1.3 所示，具体描述如下：

1. 监督学习

回归

线性模型

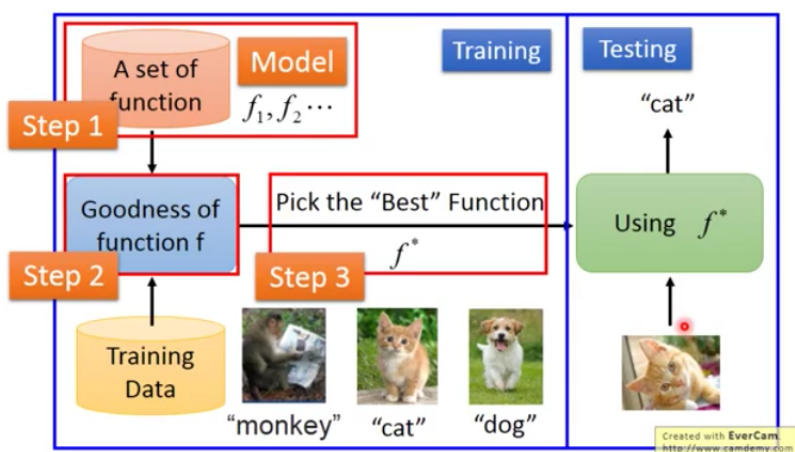


图 1.2: 机器学习三步骤

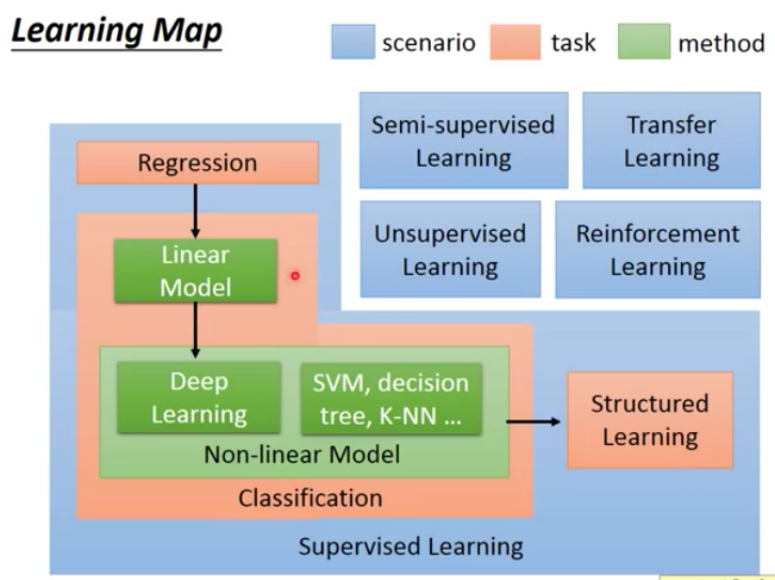


图 1.3: 机器学习图谱

深度学习，非线性模型

其它非线性模型，如 SVM、决策树、knn。

structure learning

2. 无监督学习
3. 半监督学习
4. 迁移学习
5. 强化学习

## 1.2 回归问题

线性模型：  $y = b + \sum w_i x_i$ ，其中  $x_i$  为输入数据的特征， $w_i$  为权重， $b$  为偏置。使用损失函数评价选定模型的好坏。如对于模型  $f$ ，样本  $x^n$ ，对应的输出真值为  $\hat{y}$ ：

$$L(f) = \sum_{n=1}^N (\hat{y}^n - f(x_{cp}^n))^2$$

对于线性模型：

$$L(w, b) = \sum_{n=1}^N (\hat{y}^n - (b + w \cdot x_{cp}^n))^2$$

最优化模型：

$$w^*, b^* = \arg \min_{w, b} L(w, b) = \arg \min_{w, b} \sum_{n=1}^N (\hat{y}^n - (b + w \cdot x_{cp}^n))^2$$

为求得最优解，使用梯度下降法进行优化求解。若将  $b$  看做权重  $w$  的一部分，优化模型：

$$w^* = \arg \min_w L(w)$$

权重通过梯度进行迭代：

$$w^1 \leftarrow w^0 - \eta \left. \frac{dL}{dw} \right|_{w=w^0}$$

其中， $\eta$  为学习率。梯度下降实例如图 1.4所示：对参数的偏导：

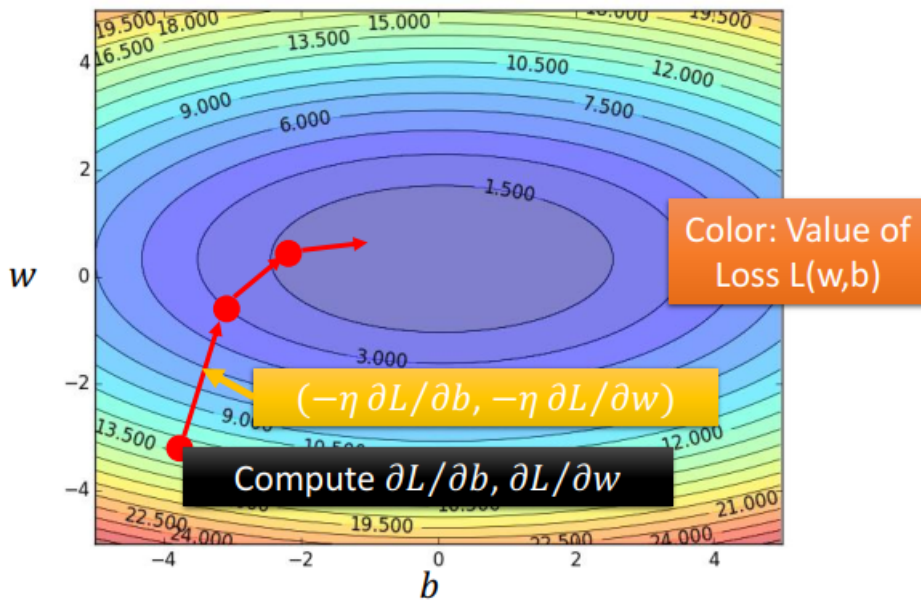


图 1.4: gradient descent

$$\frac{\partial L}{\partial w} = \sum_{n=1}^N 2(\hat{y}^n - (b + w \cdot x^n))(-x^n)$$

$$\frac{\partial L}{\partial b} = - \sum_{n=1}^N 2(\hat{y}^n - (b + w \cdot x^n))$$