

符号学习





课程信息

- Stephen Muggleton (s.muggleton@gmail.com)
- 葛存菁 (gecunjing@nju.edu.cn)
- 时间：周二、四的7、8节课
- 地点：仙II-505
- 课件：<https://gecunjing.github.io/IP/>
- 考核：期末考试（闭卷）



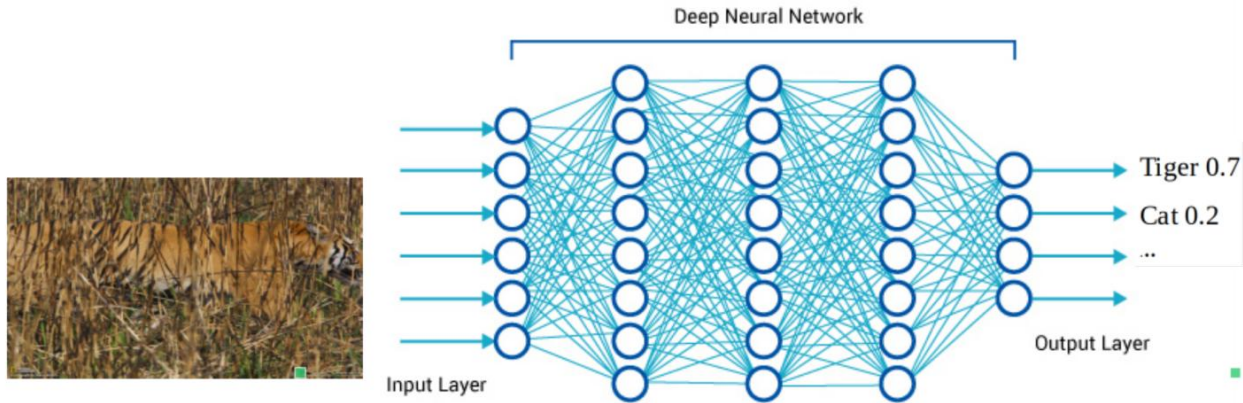


机器学习

- 问题抽象（监督学习为例）
 - 学习目标函数 F
 - 输入：给定标签的样本集合 E
 - 输出：假设 H ， H 为 F 的近似，用 H 进行预测
- 方法
 - 决策树
 - 线性模型
 - 神经网络
 -
 - 归纳逻辑程序（Inductive Logic Programming, ILP）

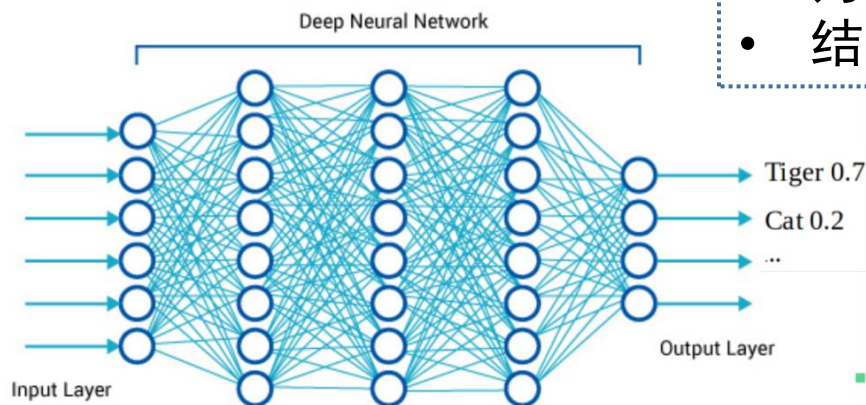
可信与可解释难题

- 例子：多目标识别



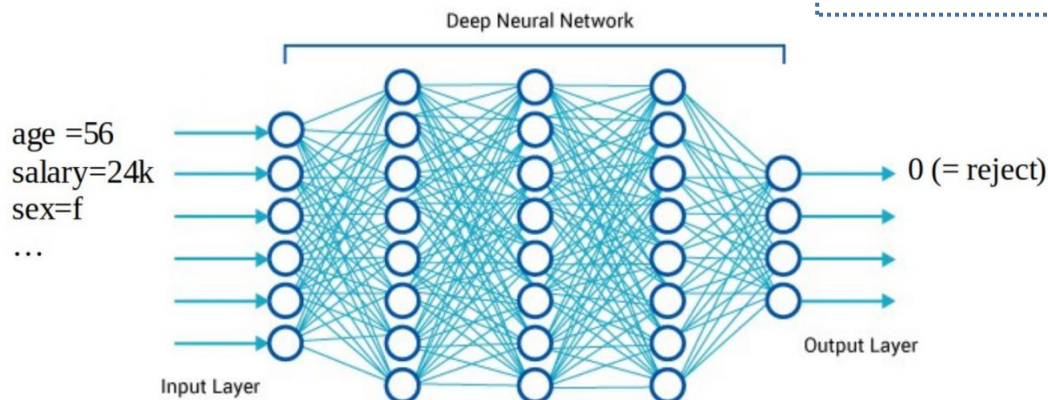
可信与可解释难题

- 例子：多目标识别



- 为什么？
- 结果如何算得的？

- 例子：信贷应用



- 大模型“幻觉”问题

- 结果的**可信度**
- 计算的**透明度**

可信可解释的人工智能

- 安全攸关领域
 - 航空、航天、轨道交通、核工业等
- 自动化芯片设计
 - 99.99999%正确也不够
 - 正确性验证的成本很高
 - 弱可解释性导致定位和修复困难
- 公平性决策
 - 法案判决、房屋分配、人员聘用等
-

A yellow starburst shape with a red outline, containing the text "承担责任?".

承担责任?

Symbolic AI

- 优点

- 可信：结果有保证、甚至有证明
- 可解释：符号语言的公式
- 利用少量数据
-

- 缺点

- 计算开销高
- 概念符号化
- 人机交互
- 非理性/无逻辑决策
- 依赖专家知识
-

预备知识

