



Python机器学习从入门到实践



汇报人：XX



2024-01-11

目录

- **机器学习概述**
- **Python机器学习基础**
- **监督学习算法**
- **无监督学习算法**
- **神经网络与深度学习**
- **Python机器学习实践案例**

01

机器学习概述



机器学习的定义



机器学习是一种人工智能（AI）技术，它使计算机系统能够从数据中学习并改进其性能，而无需进行明确的编程。

机器学习通过训练模型来识别数据中的模式，并使用这些模式来对新数据进行预测或分类。





机器学习的分类



监督学习

训练数据带有标签，模型通过学习输入与输出之间的映射关系来进行预测。



无监督学习

训练数据没有标签，模型通过发现数据中的内在结构和模式来进行学习。



半监督学习

部分训练数据带有标签，模型利用有标签和无标签数据进行学习。



强化学习

模型通过与环境的交互来学习，目标是最大化累积奖励。



机器学习的发展历程

1950年代

机器学习的概念初步形成，基于神经网络的感知机模型被提出。

1990年代

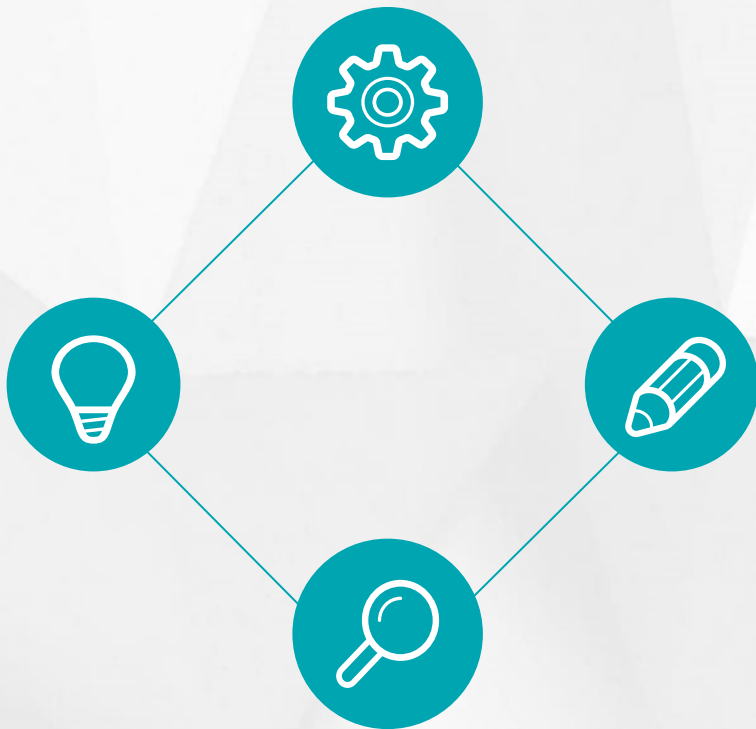
支持向量机（SVM）等核方法成为研究热点，机器学习在文本分类、图像识别等领域取得显著成果。

1980年代

决策树、K近邻等算法相继出现，机器学习开始应用于实际问题。

2000年代至今

深度学习崛起，神经网络模型在图像、语音、自然语言处理等领域取得突破性进展，机器学习应用广泛扩展。



02

Python机器学习基础



Python语言在机器学习中的应用

01

简洁易懂的语法

Python语言采用简洁易懂的语法，使得机器学习算法的实现更加直观和易于理解。

02

丰富的库支持

Python拥有大量的机器学习库，如scikit-learn、TensorFlow、Keras等，这些库提供了丰富的算法和工具，方便用户快速构建和训练机器学习模型。

03

强大的数据处理能力

Python具备强大的数据处理能力，可以轻松处理各种类型的数据，包括文本、图像、音频等，为机器学习提供了良好的数据基础。



Python机器学习常用库

scikit-learn

scikit-learn是一个功能强大的机器学习库，提供了各种分类、回归、聚类等算法，以及数据预处理、模型评估等工具。

Keras

Keras是一个基于TensorFlow的高级深度学习框架，提供简洁易懂的API，方便用户快速构建和训练深度学习模型。



TensorFlow

TensorFlow是一个开源的深度学习框架，支持各种深度学习模型的构建和训练，包括卷积神经网络、循环神经网络等。

PyTorch

PyTorch是另一个流行的深度学习框架，提供动态计算图和GPU加速等功能，适用于各种复杂的深度学习应用。



Python机器学习数据处理

数据清洗

Python提供了pandas等库进行数据清洗，包括缺失值处理、异常值处理、重复值处理等。

特征工程

特征工程是机器学习中非常重要的一步，Python提供了各种特征提取、特征选择、特征转换等方法，如文本特征提取、图像特征提取等。

数据标准化/归一化

数据标准化/归一化是机器学习中常用的数据预处理方法，可以消除不同特征之间的量纲差异，提高模型的训练效果。Python提供了各种标准化/归一化方法，如Min-Max归一化、Z-score标准化等。

03

监督学习算法



线性回归

01

线性回归模型

通过最小化预测值与真实值之间的均方误差，学习得到最优的线性模型参数。

02

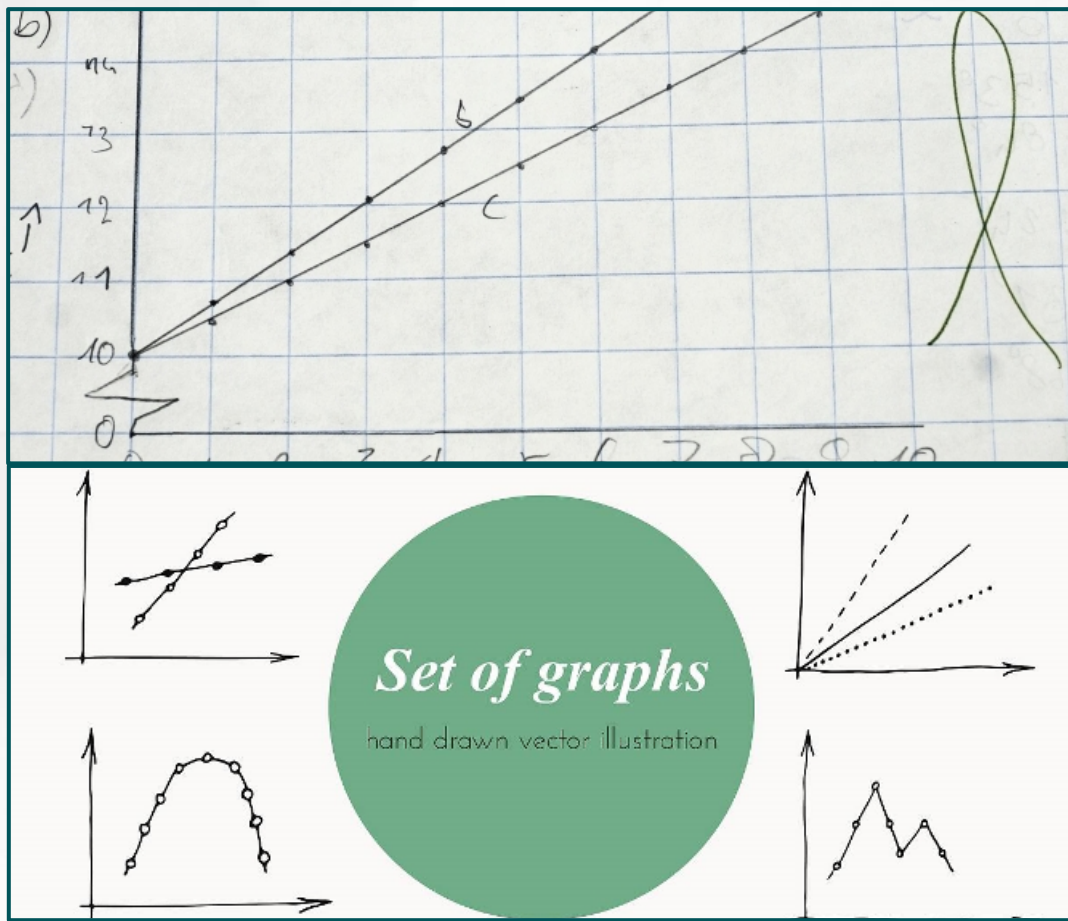
特征选择

选择与输出变量相关性强、且彼此之间相关性弱的特征，以提高模型的预测性能。

03

正则化

通过引入正则化项，防止模型过拟合，提高模型的泛化能力。





逻辑回归

01

逻辑回归模型

使用sigmoid函数将线性模型的输出映射到 $[0,1]$ 区间，表示样本属于正类的概率。

02

损失函数

采用交叉熵损失函数，衡量模型预测概率分布与真实概率分布之间的差异。

03

多分类问题

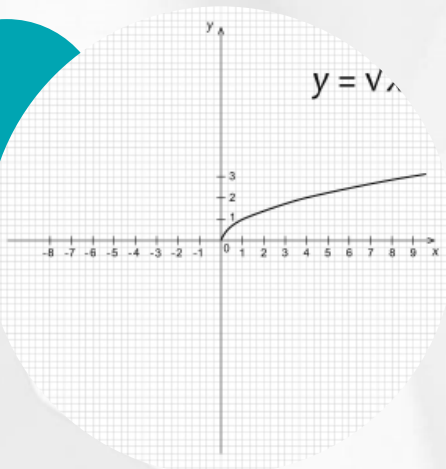
通过构建多个二分类逻辑回归模型，或使用softmax函数将输出映射到多个类别上，实现多分类问题的求解。





支持向量机

01

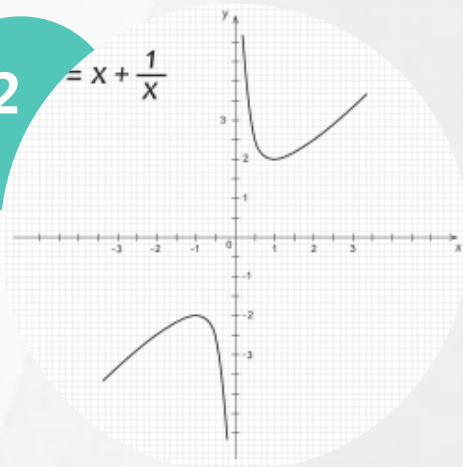


支持向量机模型



通过最大化间隔，寻找最优超平面将不同类别的样本分开。

02



核函数



引入核函数，将样本映射到高维空间，使得原本线性不可分的问题变得线性可分。

03



软间隔



允许一些样本点不满足约束条件，以缓解过拟合问题。



决策树与随机森林



决策树

通过递归地选择最优特征进行划分，构建一棵树状结构，实现对样本的分类或回归。



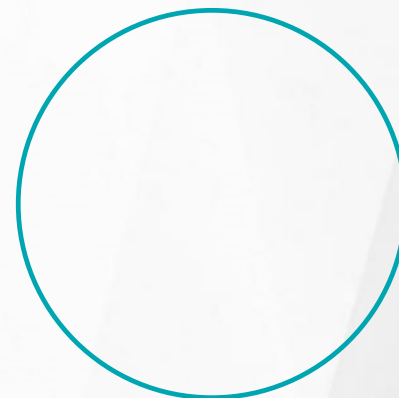
特征选择

采用信息增益、增益率、基尼指数等指标评估特征的重要性，选择最优特征进行划分。



剪枝

通过去除一些子树或叶节点，降低决策树的复杂度，防止过拟合。



随机森林

构建多个决策树，对它们的预测结果进行集成，以提高模型的预测精度和稳定性。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：
<https://d.book118.com/295341240233011222>