

人工智能导论

Introduction to Artificial Intelligence

第二章 机器学习基础



1

课程安排

课堂情况 30

- ✓ 考勤 15
- ✓ 课堂作业（雨课堂）10 + 课堂互动 5

平时作业 30

- ✓ 会发布3次平时作业，自行选择2次提交 20
- ✓ **一次课堂报告，小组形式 10**

期末考试 40

课堂报告

形式：28组，5-7人一组，每组20min，一周2组

时间：每周第三节课作为课堂分享时间，4-17周

内容：人工智能相关的小知识或热点话题

2

投票 最多可选1项

课堂报告 – 10分

形式: 28组, 5-7人一组, 每组20min, 一周2组

时间: 每周第三节课作为课堂分享时间, 4-17周

内容: 人工智能相关的小知识或热点话题

课堂报告给分方案-投票

A

所有组员加分相同

B

组长加分最高, 其他成员加分相同且低于组长

C

小组按贡献度, 采用5分制自评(5-1分), 人数限制为:
5分最多1人, 4分最多2人, 3分最多3人, 依次类推

3

课程回顾

1. 什么是智能?

4

课程回顾

1. 什么是智能？

智能是一个复杂的概念，它涉及到多个方面和层次。一般来说，**智能是指生物一般性的精神能力，指认识、理解客观事物并运用知识、经验等解决问题的能力。**

- 基本能力：记忆、观察、想象、思考、判断。
- 高级能力：理解、分析、推理、学习、规划和自我改进。

课程回顾

1. 什么是智能？

智能是一个复杂的概念，它涉及到多个方面和层次。一般来说，**智能是指生物一般性的精神能力，指认识、理解客观事物并运用知识、经验等解决问题的能力。**

- 基本能力：记忆、观察、想象、思考、判断。
- 高级能力：理解、分析、推理、学习、规划和自我改进。

2. 什么是人工智能？

课程回顾

1. 什么是智能？

智能是一个复杂的概念，它涉及到多个方面和层次。一般来说，**智能是指生物一般性的精神能力，指认识、理解客观事物并运用知识、经验等解决问题的能力。**

- 基本能力：记忆、观察、想象、思考、判断。
- 高级能力：理解、分析、推理、学习、规划和自我改进。

2. 什么是人工智能？ 人工智能目前还没有统一的定义

课程回顾

1. 什么是智能？

智能是一个复杂的概念，它涉及到多个方面和层次。一般来说，**智能是指生物一般性的精神能力，指认识、理解客观事物并运用知识、经验等解决问题的能力。**

- 基本能力：记忆、观察、想象、思考、判断。
- 高级能力：理解、分析、推理、学习、规划和自我改进。

2. 什么是人工智能？ 人工智能目前还没有统一的定义

- “人工智能之父”、图灵奖得主——约翰·麦卡锡(McCarthy)认为，**人工智能就是要让机器的行为看起来像是人所表现出的智能行为一样**
- 人工智能领域开创者——尼尔森(Nilsson)认为，**人工智能是关于人造物的智能行为，包括知觉、推理、学习、交流和在复杂环境中的行为**

8

课程回顾

3. 人工智能的流派

9

填空题 3分

3. 人工智能的流派

三大学派主义： [填空1] [填空2] [填空3]

10

课程回顾

3. 人工智能的流派

学派	主要思想	代表方法
符号主义	认为人类认知和思维的基本单元是符号，而认知过程就是在符号表示上的一种运算。	决策树、专家系统、知识图谱
联结主义	认为人类的认知过程是由大量简单神经元构成的神经网络中的信息处理过程。	感知机、神经网络
行为主义	认为应该从与外界环境的互动中获取智能，强调从行为的角度来理解智能。	强化学习

11

课程回顾

4. 人工智能的三要素

12

填空题 3分

4. 人工智能的三要素

三要素：

[填空1] [填空2] [填空3]

13

课程回顾

4. 人工智能的三要素



14

课程回顾

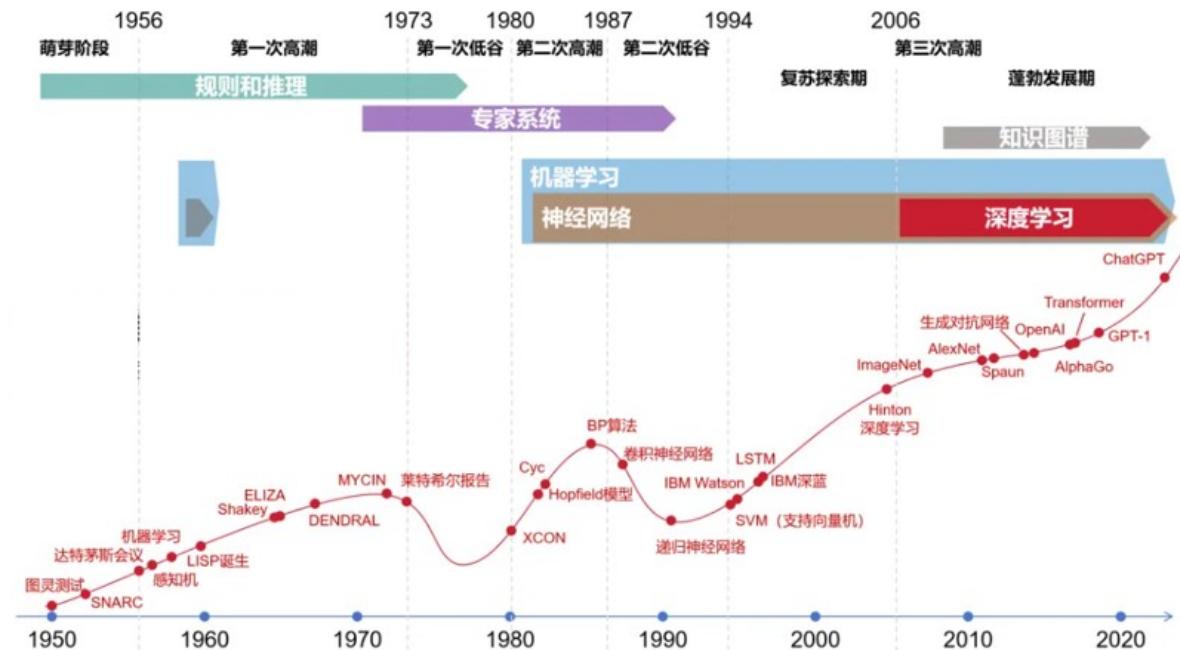
5. 人工智能发展历程

15

课程回顾

5. 人工智能发展历程

三次高潮 · 两次低谷



16

课程回顾

6. 人工智能发展方向

17

课程回顾

6. 人工智能发展方向

机器学习 · 深度学习 · 大模型

人工智能 (Artificial Intelligence)：
符号主义、联结主义、行为主义、统计主义...

机器学习 (Machine Learning)：
从数据中学习模式和规律

深度学习 (Deep Learning)：
使用深度神经网络进行机器学习

Transformer模型

大模型 (Foundation Models)

18

机器学习基础

传统机器学习任务分为两大类：**预测和分类**

- **预测任务**→连续变量（房价、温度……）
- **分类任务**→离散类别（是否为垃圾邮件……）

19

机器学习基础

传统机器学习任务分为两大类：**预测和分类**

- **预测任务**→连续变量（房价、温度……）

1. 线性回归：用线性模型建立特征与目标的关系，预测连续值

- **分类任务**→离散类别（是否为垃圾邮件……）

3. 贝叶斯网络：基于概率理论，利用**贝叶斯定理**来**分类**数据，并假设特征间相互独立以简化计算

4. 支持向量机：通过寻找一个最优超平面来**分类**数据，以**最大化不同类别数据点之间的间隔**

5. K-邻近算法：通过计算待分类样本与训练集中**K个最近邻样本的相似性**来直接**分类**数据

20

机器学习基础

传统机器学习任务分为两大类：**预测和分类**

- **预测任务**→连续变量（房价、温度.....）

1. **线性回归**：用线性模型建立特征与目标的关系，预测连续值

预测和分类的过渡 连续值，再利用**Sigmoid函数**将结

果映射到(0,1)区间，最后划分阈

- **分类任务**→离散类别（是否为垃圾邮件.....）

值后，可用于**二分类任务**

3. **贝叶斯网络**：基于概率理论，利用**贝叶斯定理**来**分类**数据，并假设特征间相互独立以简化计算

4. **支持向量机**：通过寻找一个最优超平面来**分类**数据，以**最大化不同类别数据点之间的间隔**

5. **K-邻近算法**：通过计算待分类样本与训练集中**K个最近邻样本的相似性**来直接**分类**数据

21

Lecture Plan

- 线性回归
- 逻辑回归
- 贝叶斯网络
- 支持向量机SVM
- K-邻近算法KNN



22

Lecture Plan

- 线性回归
- 逻辑回归
- 贝叶斯网络
- 支持向量机SVM
- K-邻近算法KNN



23

机器学习基础

传统机器学习任务分为两大类：**预测和分类**

- **预测任务**→连续变量（房价、温度……）

1. **线性回归**：用**线性模型**建立特征与目标的关系，预测连续值

2. **逻辑回归**：先用**线性模型**预测

连续值，再利用**Sigmoid函数**将结

预测和分类的过渡 果映射到(0,1)区间，最后划分阈

- **分类任务**→离散类别（是否为垃圾邮件……）

值后，可用于**二分类任务**

3. **贝叶斯网络**：基于概率理论，利用**贝叶斯定理**来**分类**数据，并假设特征间相互独立以简化计算

4. **支持向量机**：通过寻找一个最优超平面来**分类**数据，以**最大化不同类别数据点之间的间隔**

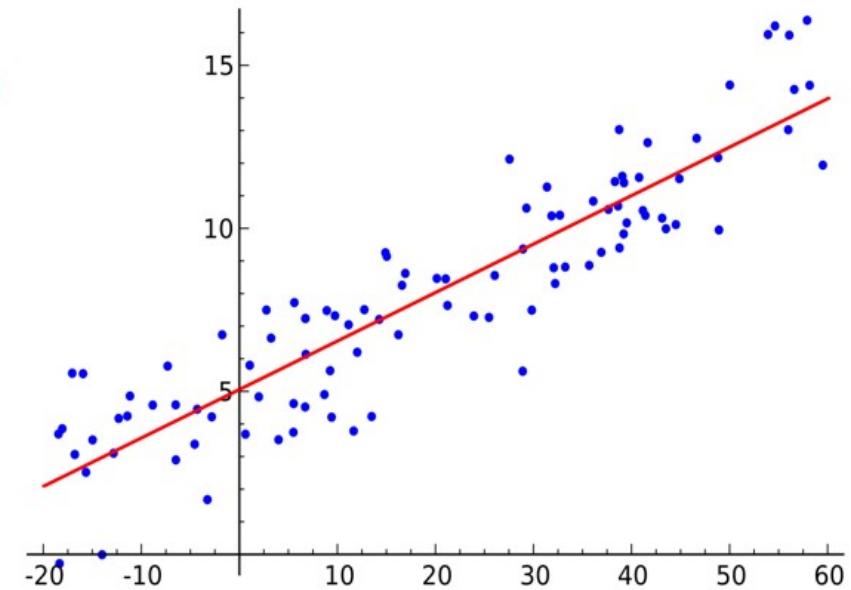
5. **K-邻近算法**：通过计算待分类样本与训练集中**K个最近邻样本的相似性**来直接**分类**数据

24

线性回归

线性回归概念

找到一条**最佳拟合直线**，使数据点尽可能靠近这条线



25

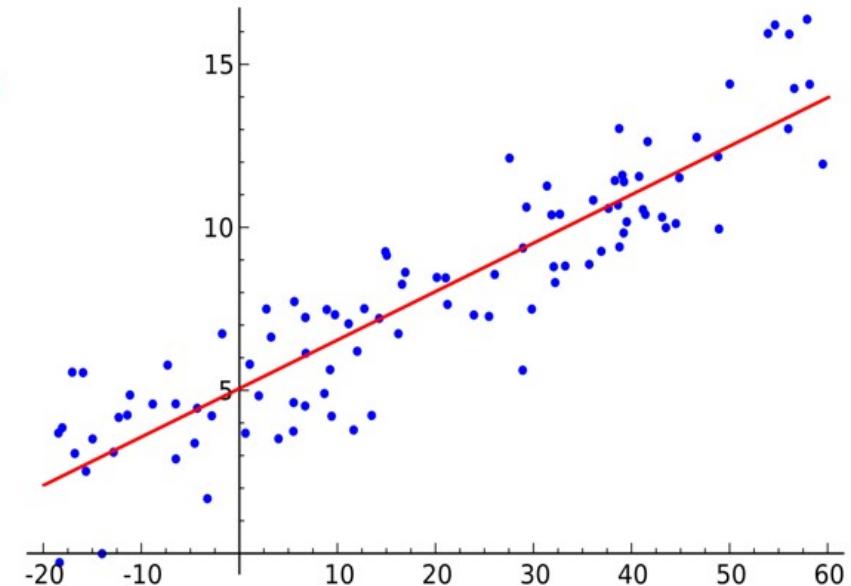
线性回归

线性回归概念

找到一条**最佳拟合直线**, 使数据点尽可能靠近这条线



建立自变量 x 和因变量 y 之间的线性关系



26

线性回归

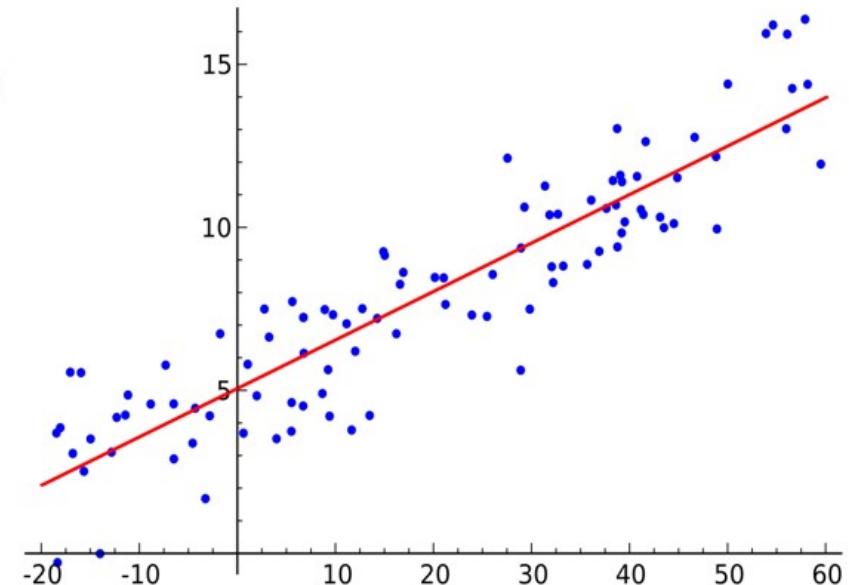
线性回归概念

找到一条**最佳拟合直线**，使数据点尽可能靠近这条线



建立自变量 x 和因变量 y 之间的线性关系

- 当只有一个自变量时，称为**简单线性回归**；
- 当有多个自变量时，称为**多元线性回归**。



27

线性回归

线性回归的基本元素

示例：我们希望根据房屋的面积和房龄来估算房屋价格。

28

线性回归

线性回归的基本元素

示例：我们希望根据房屋的面积和房龄来估算房屋价格。

为了开发一个**能预测房屋价格的线性模型**，我们需要收集一个真实的数据集。这个数据集包括房屋的**面积、房龄和价格**。

线性回归

线性回归的基本元素

示例：我们希望根据房屋的面积和房龄来估算房屋价格。

为了开发一个**能预测房屋价格的线性模型**，我们需要收集一个真实的数据集。这个数据集包括房屋的**面积、房龄和价格**。

在机器学习的**术语**中：

- 该数据集称为 _____。
- 数据（比如一次房屋交易相对应的数据）称为 _____。
- 预测的目标（比如预测房屋价格）称为 _____。
- 预测所依据的自变量（面积和房龄）称为 _____。

30

填空题 4分

线性回归的基本元素

示例：我们希望根据房屋的面积和房龄来估算房屋价格。

为了开发一个**能预测房屋价格的线性模型**，我们需要收集一个真实的数据集。这个数据集包括房屋的**面积、房龄和价格**。

在机器学习的**术语**中：

- 该数据集称为 [填空1] 。
- 数据（比如一次房屋交易相对应的数据）称为 [填空2] 。
- 预测的目标（比如预测房屋价格）称为 [填空3] 。
- 预测所依据的自变量（面积和房龄）称为 [填空4] 。

31

线性回归

线性回归的基本元素

示例：我们希望根据房屋的面积和房龄来估算房屋价格。

为了开发一个**能预测房屋价格的线性模型**，我们需要收集一个真实的数据集。这个数据集包括房屋的**面积、房龄和价格**。

在机器学习的**术语**中：

- 该数据集称为**训练集 (training set)**。
- 数据（比如一次房屋交易相对应的数据）称为**样本 (sample)**。
- 预测的目标（比如预测房屋价格）称为**标签 (label)**。
- 预测所依据的自变量（面积和房龄）称为**特征 (feature)**。

32

线性回归

线性模型

线性假设是指目标/标签（房屋价格）可以表示为特征（面积和房龄）的**加权和**，如下：

$$\text{price} = w_{\text{area}} \cdot \text{area} + w_{\text{age}} \cdot \text{age} + b$$

线性回归

线性模型

线性假设是指目标/标签（房屋价格）可以表示为特征（面积和房龄）的**加权和**，如下：

$$\text{price} = w_{\text{area}} \cdot \text{area} + w_{\text{age}} \cdot \text{age} + b$$

权重 特征

34

线性回归

线性模型

线性假设是指目标/标签（房屋价格）可以表示为特征（面积和房龄）的加权和，如下：

$$\text{price} = w_{\text{area}} \cdot \text{area} + w_{\text{age}} \cdot \text{age} + b$$

权重 特征
↓

决定了每个特征对我们预测值的影响

线性回归

线性模型

线性假设是指目标/标签（房屋价格）可以表示为特征（面积和房龄）的加权和，如下：

$$\text{price} = w_{\text{area}} \cdot \text{area} + w_{\text{age}} \cdot \text{age} + b$$

权重 特征 偏置

线性回归

线性模型

线性假设是指目标/标签（房屋价格）可以表示为特征（面积和房龄）的加权和，如下：

$$\text{price} = w_{\text{area}} \cdot \text{area} + w_{\text{age}} \cdot \text{age} + b$$

权重 特征

偏置

当所有特征都取值为0时，预测值应该为多少

线性回归

线性模型

线性假设是指目标/标签（房屋价格）可以表示为特征（面积和房龄）的**加权和**，如下：

$$\text{price} = w_{\text{area}} \cdot \text{area} + w_{\text{age}} \cdot \text{age} + b$$

权重
↓

决定了每个特征对我们预测值的影响

特征

偏置
↓

当所有特征都取值为0时，预测值应该为多少

给定一个数据集，目标是寻找模型的**权重 w** 和**偏置 b** ，

使模型的预测更符合数据里的真实房屋价格。

38

线性回归

线性模型

在机器学习领域，我们通常使用的是高维数据集，建模时采用**线性代数表示法**会比较方便。

39

线性回归

线性模型

在机器学习领域，我们通常使用的是高维数据集，建模时采用**线性代数表示法**会比较方便。

当我们的输入包含 d 个特征时，我们将预测结果 \hat{y} 表示为：

40

线性回归

线性模型

在机器学习领域，我们通常使用的是高维数据集，建模时采用**线性代数表示法**会比较方便。

当我们的输入包含 d 个特征时，我们将预测结果 \hat{y} 表示为：

$$\hat{y} = w_1x_1 + \dots + w_dx_d + b.$$

41

线性回归

线性模型

在机器学习领域，我们通常使用的是高维数据集，建模时采用**线性代数表示法**会比较方便。

当我们的输入包含 d 个特征时，我们将预测结果 \hat{y} 表示为：

$$\hat{y} = w_1x_1 + \dots + w_dx_d + b.$$

向量形式

- 所有特征： $x \in \mathbb{R}^d$
- 所有权重： $w \in \mathbb{R}^d$

$$\hat{y} = \mathbf{w}^\top \mathbf{x} + b.$$

线性回归

线性模型

在机器学习领域，我们通常使用的是高维数据集，建模时采用**线性代数表示法**会比较方便。

当我们的输入包含 d 个特征时，我们将预测结果 \hat{y} 表示为：

$$\hat{y} = w_1x_1 + \dots + w_dx_d + b.$$

向量形式

- 所有特征： $x \in \mathbb{R}^d$
- 所有权重： $w \in \mathbb{R}^d$

向量 x 对应于**单个数据样本**的特征

$$\hat{y} = w^\top x + b.$$

线性回归

线性模型

在机器学习领域，我们通常使用的是高维数据集，建模时采用**线性代数表示法**会比较方便。

当我们的输入包含 d 个特征时，我们将预测结果 \hat{y} 表示为：

$$\hat{y} = w_1x_1 + \dots + w_dx_d + b.$$

向量形式

- 所有特征： $x \in \mathbb{R}^d$
- 所有权重： $w \in \mathbb{R}^d$

向量 x 对应于**单个数据样本**的特征

$$\hat{y} = w^\top x + b.$$

矩阵形式 $X \in \mathbb{R}^{n \times d}$ 整个数据集的 n 个样本

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{w} + b \quad \text{预测值 } \hat{\mathbf{y}} \in \mathbb{R}^n$$

44

线性回归

线性模型

在机器学习领域，我们通常使用的是高维数据集，建模时采用**线性代数表示法**会比较方便。

当我们的输入包含 d 个特征时，我们将预测结果 \hat{y} 表示为：

$$\hat{y} = w_1x_1 + \dots + w_dx_d + b.$$

向量形式

- 所有特征： $x \in \mathbb{R}^d$
- 所有权重： $w \in \mathbb{R}^d$

向量 x 对应于**单个数据样本**的特征

$$\hat{y} = w^\top x + b.$$

矩阵形式

$X \in \mathbb{R}^{n \times d}$ 整个数据集的 n 个样本

每一行是一个样本，每一列是一种特征

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{w} + b \quad \text{预测值 } \hat{\mathbf{y}} \in \mathbb{R}^n$$

45

Lecture Plan

- 线性回归
- 逻辑回归
- 贝叶斯网络
- 支持向量机SVM
- K-邻近算法KNN



46

机器学习基础

传统机器学习任务分为两大类：**预测和分类**

- **预测任务**→连续变量（房价、温度……）

1. 线性回归：用线性模型建立特征与目标的关系，预测连续值

- 2. **逻辑回归**：先用线性模型**预测**

连续值，再利用**Sigmoid函数**将结果映射到(0,1)区间，最后划分阈

预测和分类的过渡

值后，可用于**二分类任务**

- **分类任务**→离散类别（是否为垃圾邮件……）

3. 贝叶斯网络：基于概率理论，利用**贝叶斯定理**来**分类**数据，并假设特征间相互独立以简化计算

4. 支持向量机：通过寻找一个最优超平面来**分类**数据，以**最大化不同类别数据点之间的间隔**

5. K-邻近算法：通过计算待分类样本与训练集中**K个最近邻样本的相似性**来直接**分类**数据

47

逻辑回归

在日常生活中，需要涉及预测**离散类别**：

- 客户是否会购买某个商品
- 借款人是否会违约
-

48

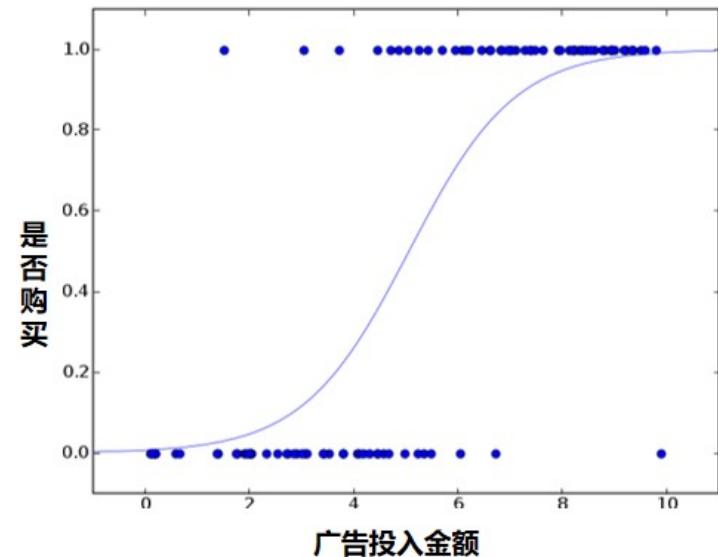
逻辑回归

在日常生活中，需要涉及预测**离散类别**：

- 客户是否会购买某个商品
- 借款人是否会违约
-

假设我们有这样一组数据：

给不同的用户投放不同金额的广告，记录他们购买广告商品的行为， 1 代表购买， 0 代表未购买。



49

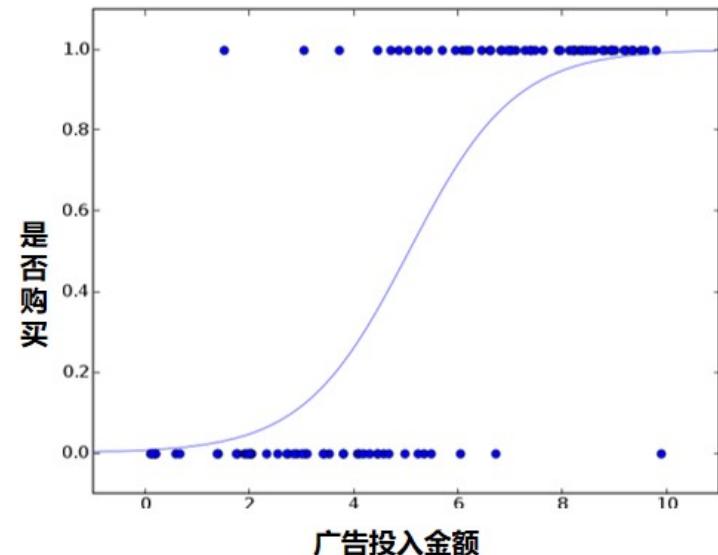
逻辑回归

在日常生活中，需要涉及预测**离散类别**：

- 客户是否会购买某个商品
- 借款人是否会违约
-

假设我们有这样一组数据：

给不同的用户投放不同金额的广告，记录他们购买广告商品的行为， 1 代表购买， 0 代表未购买。



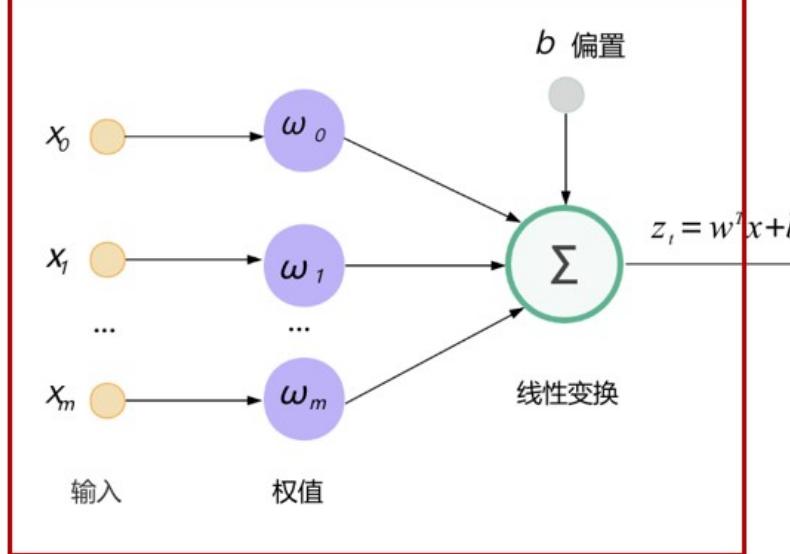
如何通过**线性模型**完成**离散类别的预测**？

50

逻辑回归

逻辑回归

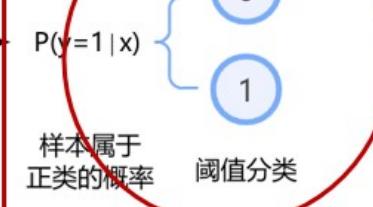
线性模型



Sigmoid函数变换

$$\sigma(\cdot)$$
$$\sigma(x) = \frac{1}{1+e^{-x}}$$

离散标签



输入特征向量



线性组合



Sigmoid激活

概率输出 (0~1)

51

逻辑回归

逻辑回归公式

$$h(x) = w_0 + w_1 x^{(1)} + w_2 x^{(2)} + \dots + w_n x^{(n)}$$

可以设 $x^{(0)} = 1$ ，则

标量形式： $h(x) = w_0 x^{(0)} + w_1 x^{(1)} + w_2 x^{(2)} + \dots + w_n x^{(n)} = \sum_{k=0}^n w_k x^{(k)}$

向量形式： $h(x) = w^T x$

$$z = h(x) = w^T x \quad z \in (-\infty, +\infty)$$

52

逻辑回归

逻辑回归公式

$$h(x) = w_0 + w_1 x^{(1)} + w_2 x^{(2)} + \dots + w_n x^{(n)}$$

可以设 $x^{(0)} = 1$ ，则

标量形式： $h(x) = w_0 x^{(0)} + w_1 x^{(1)} + w_2 x^{(2)} + \dots + w_n x^{(n)} = \sum_{k=0}^n w_k x^{(k)}$

向量形式： $h(x) = w^T x$

$$z = h(x) = w^T x \quad z \in (-\infty, +\infty)$$

我们希望输出的值代表事件发生的概率，即 $z \in [0,1]$

对输出 z 作用一个函数 $\sigma(z)$ ，将 z 压缩至 $[0,1]$ 区间

53

逻辑回归

逻辑回归公式

$$h(x) = w_0 + w_1 x^{(1)} + w_2 x^{(2)} + \dots + w_n x^{(n)}$$

可以设 $x^{(0)} = 1$ ，则

标量形式： $h(x) = w_0 x^{(0)} + w_1 x^{(1)} + w_2 x^{(2)} + \dots + w_n x^{(n)} = \sum_{k=0}^n w_k x^{(k)}$

向量形式： $h(x) = w^T x$

$$z = h(x) = w^T x \quad z \in (-\infty, +\infty)$$

我们希望输出的值代表事件发生的概率，即 $z \in [0, 1]$

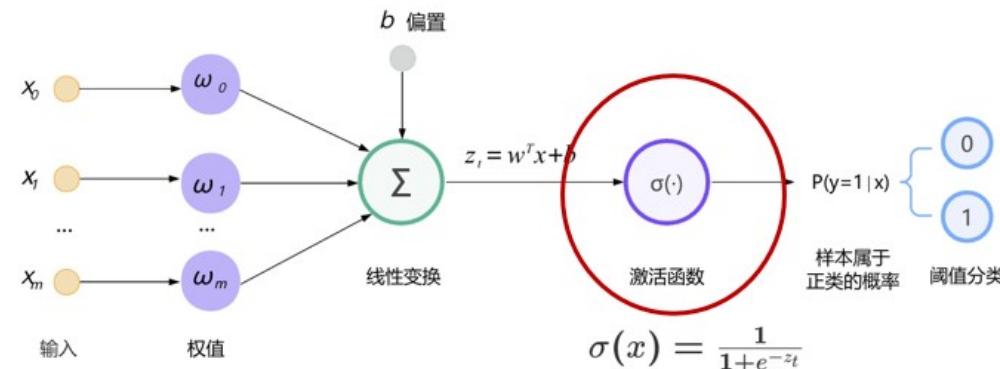
对输出 z 作用一个函数 $\sigma(z)$ ，将 z 压缩至 $[0, 1]$ 区间

Sigmoid函数

54

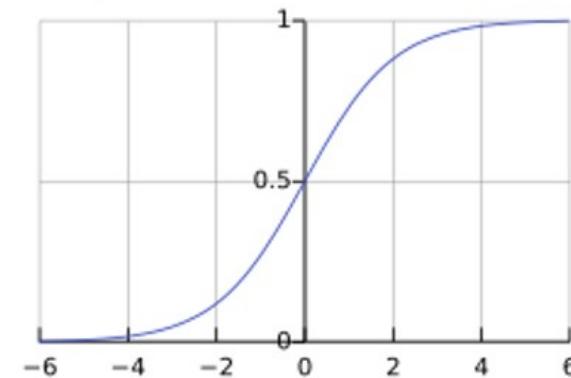
逻辑回归

Sigmoid函数



$\sigma(z)$ 代表一个常用连续S形函数 (Sigmoid function) 或逻辑函数 (Logistic function)

$$\sigma(z) = g(z) = \frac{1}{1+e^{-z}} \quad z=w^T x$$



当 $\sigma(z)$ 大于等于0.5时，预测为1；当 $\sigma(z)$ 小于0.5时，预测为0

55

Lecture Plan

- 线性回归
- 逻辑回归
- 贝叶斯网络
- 支持向量机SVM
- K-邻近算法KNN



56

机器学习基础

传统机器学习任务分为两大类：**预测和分类**

- 预测任务→连续变量（房价、温度……）

1. 线性回归：用线性模型建立特征与目标的关系，预测连续值

- 2. 逻辑回归：先用线性模型预测

连续值，再利用**Sigmoid函数**将结

预测和分类的过渡 果映射到(0,1)区间，最后划分阈

- 分类任务→离散类别（是否为垃圾邮件……）

值后，可用于**二分类任务**

3. **贝叶斯网络**：基于概率理论，利用**贝叶斯定理**来**分类**数据，并假设特征间相互独立以简化计算

4. **支持向量机**：通过寻找一个最优超平面来**分类**数据，以**最大化不同类别数据点之间的间隔**

5. **K-邻近算法**：通过计算待分类样本与训练集中**K个最近邻样本的相似性**来直接**分类**数据

57

贝叶斯网络

贝叶斯



托马斯·贝叶斯 (Thomas Bayes, 约1701年—1761年4月7日) , 18世纪英国数学家。1742年成为英国皇家学会会员。

贝叶斯以其在**概率论领域**的研究闻名于世，他提出的贝叶斯定理对于现代概率论和数理统计的发展有重要的影响。

贝叶斯网络

贝叶斯定理

条件概率 (后验概率) : 事件A在事件B已经发生条件下的发生概率。

$$P(A|B)$$

联合概率: 事件A和B共同发生的概率。

$$P(A \cap B)$$

边缘概率 (先验概率) : 所有可能情况下，某个事件发生的概率。

边缘化：在联合概率中，对于特定事件，把其他变量/事件通过合并为全概率的形式进行消去（对离散随机变量求和得全概率，对连续随机变量积分得全概率）。

$$P(A)$$

$$P(A) = \sum_b P(A, b)$$

离散随机变量求和

$$P(A) = \int P(A, b) db$$

连续随机变量积分

59

贝叶斯网络

贝叶斯定理

条件概率

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

60

贝叶斯网络

贝叶斯定理

条件概率

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

61

贝叶斯网络

贝叶斯定理

条件概率

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



贝叶斯公式

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

62

贝叶斯网络

贝叶斯定理

贝叶斯公式

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

局限：只能处理**两个随机变量**的条件关系，表达能力有限

63

贝叶斯网络

多变量场景

- 疾病诊断 (症状、检验结果、遗传因素.....)
- 风险预测 (市场、政策、用户行为.....)

现实问题往往是**多因素**影响的

64

贝叶斯网络

多变量场景

- 疾病诊断 (症状、检验结果、遗传因素.....)
- 风险预测 (市场、政策、用户行为.....)

现实问题往往是**多因素**影响的

如果使用贝叶斯公式处理 → 联合分布十分复杂，**参数爆炸**

贝叶斯网络

多变量场景

- 疾病诊断 (症状、检验结果、遗传因素.....)
- 风险预测 (市场、政策、用户行为.....)

现实问题往往是**多因素影响的**

如果使用贝叶斯公式处理 → 联合分布十分复杂，**参数爆炸**

- 考虑三变量 A 、 B 、 C ，需要计算联合分布 $P(A, B, C)$ ，涉及 2^3 种组合→参数量为 2^3
- 考虑四变量 A 、 B 、 C 、 D ，需要计算联合分布 $P(A, B, C, D)$ ，涉及 2^4 种组合→参数量为 2^4
-

66

贝叶斯网络

多变量场景

- 疾病诊断 (症状、检验结果、遗传因素.....)
- 风险预测 (市场、政策、用户行为.....)

现实问题往往是**多因素影响的**

如果使用贝叶斯公式处理 → 联合分布十分复杂，**参数爆炸**

- 考虑三变量 A 、 B 、 C ，需要计算联合分布 $P(A, B, C)$ ，涉及 2^3 种组合→参数量为 2^3
- 考虑四变量 A 、 B 、 C 、 D ，需要计算联合分布 $P(A, B, C, D)$ ，涉及 2^4 种组合→参数量为 2^4
-

猜想：利用条件独立性，减少参数

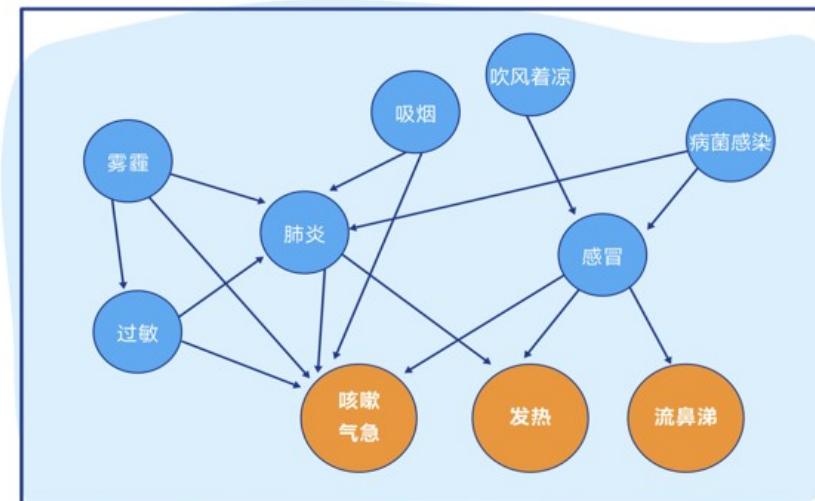
67

贝叶斯网络

贝叶斯网络基础概念

贝叶斯网络，又称信念网络、有向无环图模型，是一种概率图模型。

- 随机变量→节点
- 因果/依赖条件关系→有向边
- 条件独立→无连接



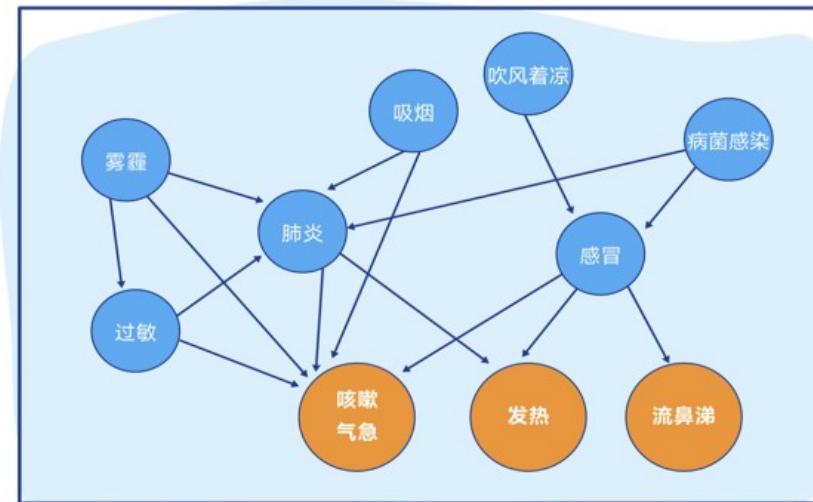
68

贝叶斯网络

贝叶斯网络基础概念

贝叶斯网络，又称信念网络、有向无环图模型，是一种概率图模型。

- 随机变量→节点
- 因果/依赖条件关系→有向边
- 条件独立→无连接



特点：每个节点只依赖于其父节点

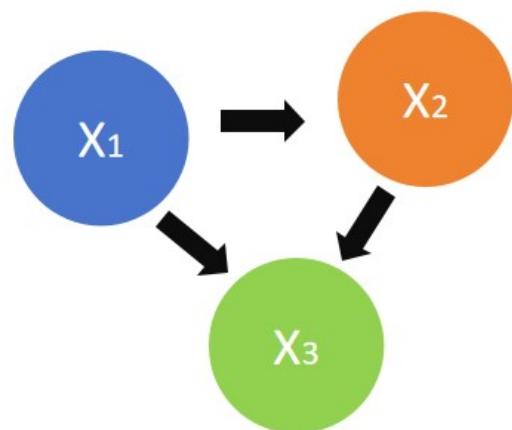
69

贝叶斯网络

贝叶斯网络基础概念

贝叶斯网络结构图如下：

以 X_i 表示第*i*个节点，而 X_i 的“因”以 P_i 表示， X_i 的“果”以 C_i 表示；



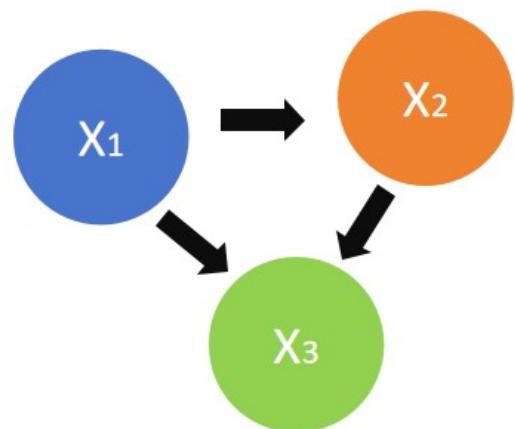
$$P_1 = ? \quad C_1 = ?$$

70

填空题 6分

贝叶斯网络基础概念

贝叶斯网络结构图如下：



以 X_i 表示第*i*个节点，而 X_i 的“因”以 P_i 表示， X_i 的“果”以 C_i 表示；

$$P_1 = ? \quad C_1 = ?$$

$$P_1 = [\text{填空1}] \quad C_1 = [\text{填空2}]$$

$$P_2 = [\text{填空3}] \quad C_2 = [\text{填空4}]$$

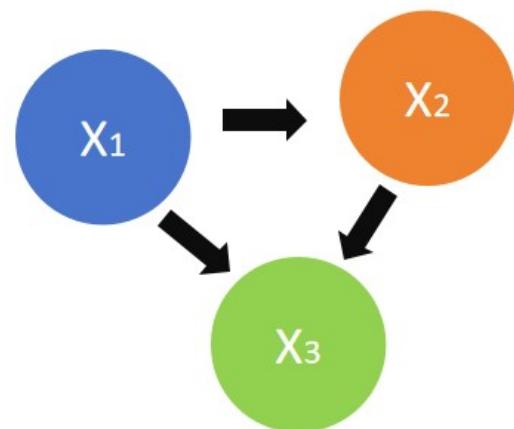
$$P_3 = [\text{填空5}] \quad C_3 = [\text{填空6}]$$

71

贝叶斯网络

贝叶斯网络基础概念

贝叶斯网络结构图如下：



以 X_i 表示第*i*个节点，而 X_i 的“因”以 P_i 表示， X_i 的“果”以 C_i 表示；

$$P_1 = ? \quad C_1 = ?$$

$$P_1 = \emptyset; \quad C_1 = \{X_2, X_3\}$$

$$P_2 = \{X_1\}; \quad C_2 = \{X_3\}$$

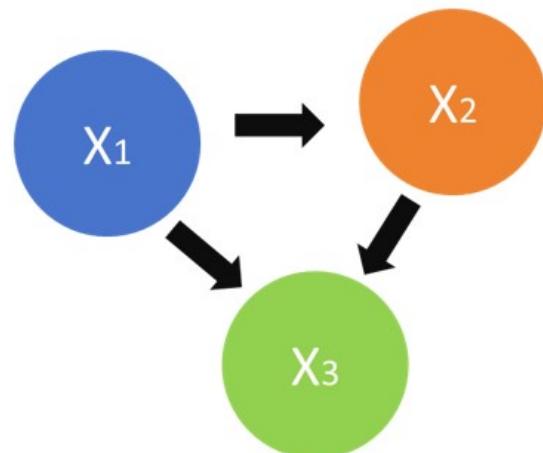
$$P_3 = \{X_1, X_2\}; \quad C_3 = \emptyset$$

72

贝叶斯网络

贝叶斯网络公式

令 $G = (I, E)$ 表示一个有向无环图，其中 I 代表图中所有节点集合，而 E 代表有向连接边集合，且令 $X = (X_i) \quad i \in I$ 为有向无环图中的某一节点 i 所代表的随机变量。



若图中**所有节点 x 的联合概率**可以表示成：

$$p(x) = \prod_{i \in I} p(x_i | x_{\text{pa}(i)})$$

\downarrow
 $\text{pa}(i)$ 表示节点 i 的“因”

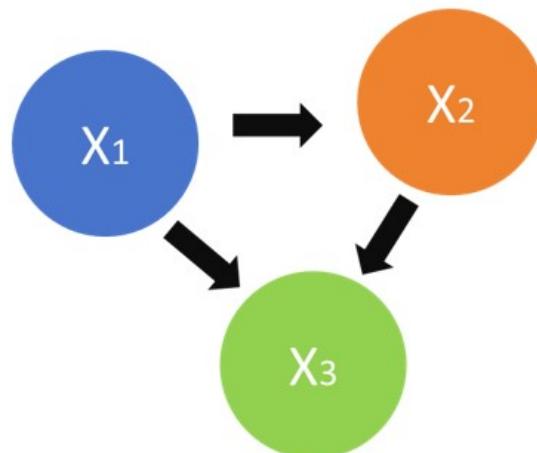
则有向无环图 G 为**贝叶斯网络**。

73

贝叶斯网络

贝叶斯网络公式

令 $G = (I, E)$ 表示一个有向无环图，其中 I 代表图中所有节点集合，而 E 代表有向连接边集合，
且令 $X = (X_i) \ i \in I$ 为有向无环图中的某一节点 i 所代表的随机变量。



根据左图的贝叶斯网络， x_1, x_2, x_3 的联合概率为？

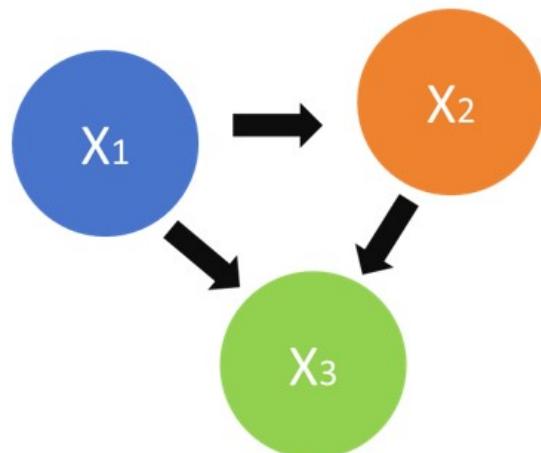
$$p(x_1, x_2, x_3) = ?$$

74

主观题 10分

贝叶斯网络公式

令 $G = (I, E)$ 表示一个有向无环图，其中 I 代表图中所有节点集合，而 E 代表有向连接边集合，
且令 $X = (X_i) \ i \in I$ 为有向无环图中的某一节点 i 所代表的随机变量。



根据左图的贝叶斯网络， x_1, x_2, x_3 的联合概率为？

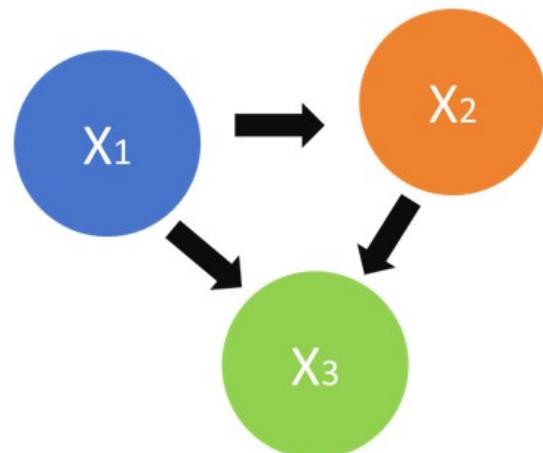
$$p(x_1, x_2, x_3) = ? \quad \text{请上传图片回答}$$

75

贝叶斯网络

贝叶斯网络公式

令 $G = (I, E)$ 表示一个有向无环图，其中 I 代表图中所有节点集合，而 E 代表有向连接边集合，且令 $X = (X_i) \ i \in I$ 为有向无环图中的某一节点 i 所代表的随机变量。



根据左图的贝叶斯网络， x_1, x_2, x_3 的联合概率为？

$$p(x_1, x_2, x_3) = p(x_3|x_2, x_1)p(x_2|x_1)p(x_1)$$

Lecture Plan

- 线性回归
- 逻辑回归
- 贝叶斯网络
- 支持向量机SVM
- K-邻近算法KNN



77

机器学习基础

传统机器学习任务分为两大类：**预测和分类**

- **预测任务**→连续变量（房价、温度……）

1. 线性回归：用线性模型建立特征与目标的关系，预测连续值

- 2. 逻辑回归：先用线性模型预测

连续值，再利用**Sigmoid函数**将结

预测和分类的过渡 果映射到(0,1)区间，最后划分阈

- **分类任务**→离散类别（是否为垃圾邮件……）

值后，可用于**二分类任务**

3. 贝叶斯网络：基于概率理论，利用**贝叶斯定理**来分类数据，并假设特征间相互独立以简化计算

4. **支持向量机**：通过寻找一个最优超平面来**分类**数据，以**最大化不同类别数据点之间的间隔**

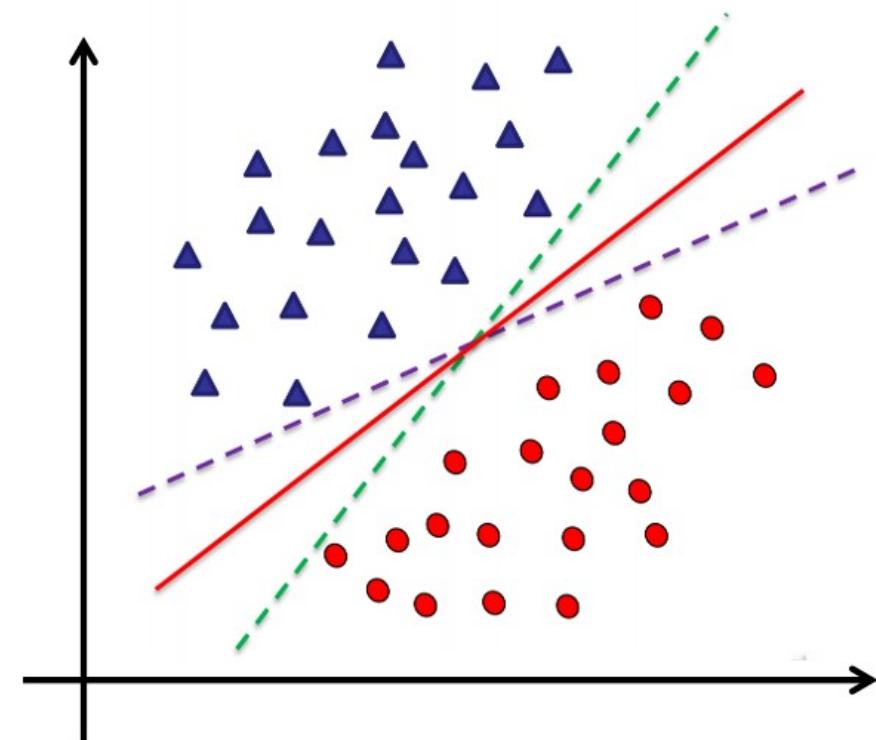
5. K-邻近算法：通过计算待分类样本与训练集中**K个最近邻样本的相似性**来直接**分类**数据

78

支持向量机SVM

支持向量机概述

应该选择哪一条线（决策超平面）
作为两类样本的分界线？



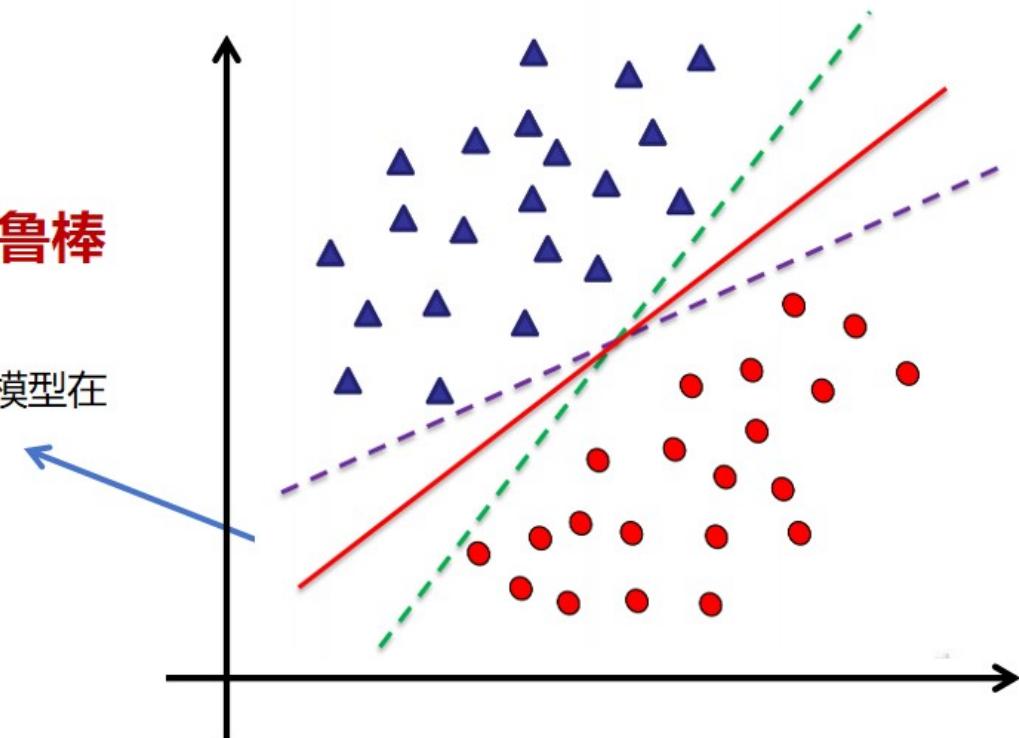
79

支持向量机SVM

支持向量机概述

红色的决策线对输入扰动更加鲁棒

选择更加鲁棒的决策超平面，有助于提高模型在
噪声数据或分布偏移场景下的泛化能力



80

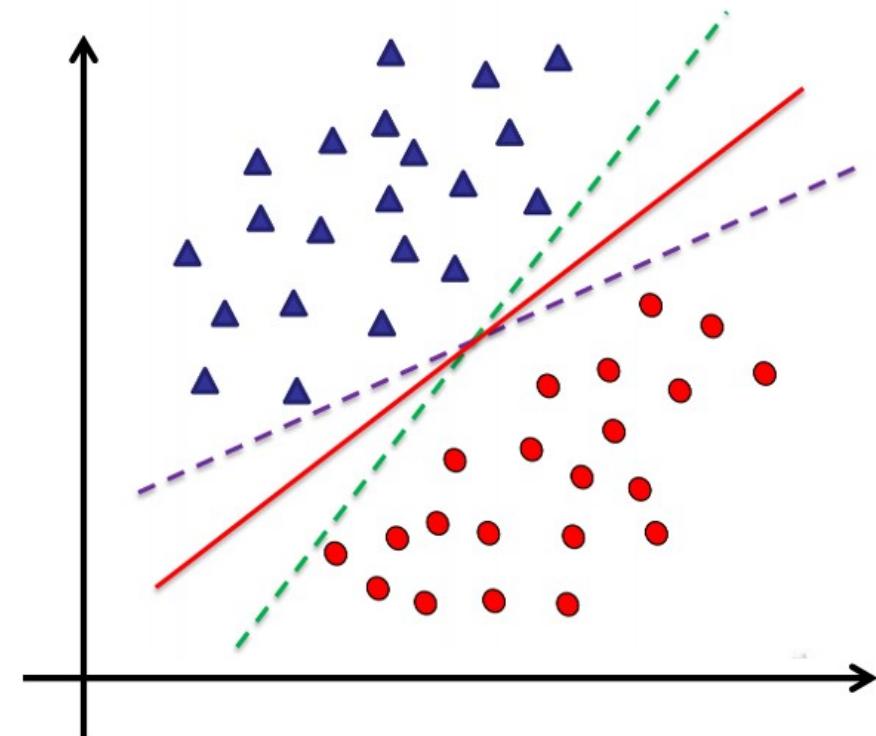
支持向量机SVM

支持向量机概述

支持向量机 (Support Vector Machine, SVM)

对数据进行二分类

SVM核心思想是寻找一个分类超平面，
使得样本点与超平面的**距离最大化**



81

支持向量机SVM

支持向量机概述

任意超平面可以用下面这个线性方程来描述：

$$w^T x + b = 0$$

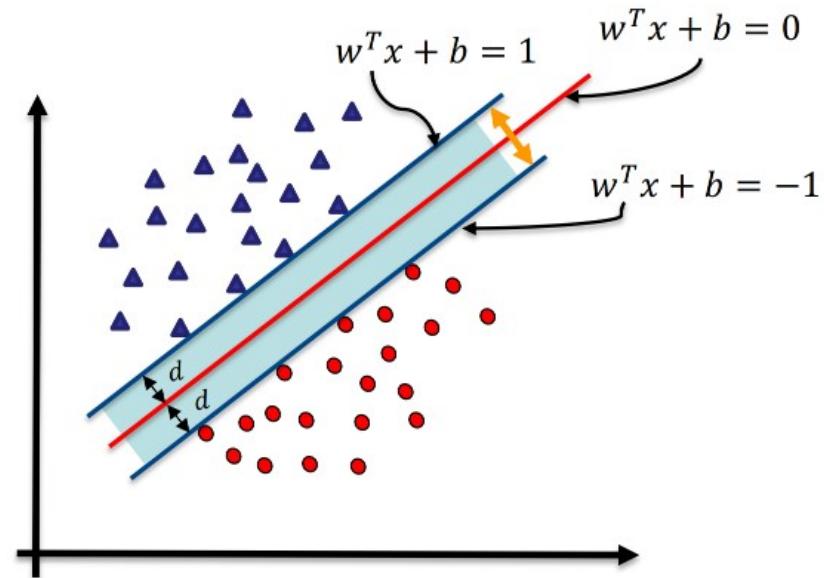
二维空间点 (x, y) 到直线 $Ax + By + C = 0$ 的距离是：

$$\frac{|Ax + By + C|}{\sqrt{A^2 + B^2}}$$

扩展到 n 维空间后，

点 $x = x^{(1)}, x^{(2)}, \dots, x^{(n)}$ 到超平面 $w^T x + b = 0$ 的距离为： $\frac{|w^T x + b|}{\|w\|}$

其中， $\|w\| = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$



支持向量机SVM

支持向量机概述

任意超平面可以用下面这个线性方程来描述：

$$w^T x + b = 0$$

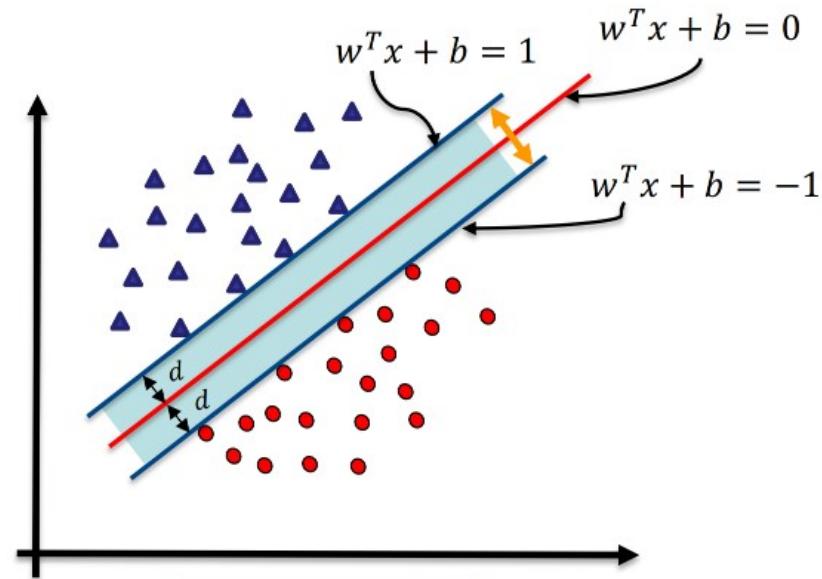
二维空间点 (x, y) 到直线 $Ax + By + C = 0$ 的距离是：

$$\frac{|Ax + By + C|}{\sqrt{A^2 + B^2}}$$

扩展到 n 维空间后，

点 $x = x^{(1)}, x^{(2)}, \dots, x^{(n)}$ 到超平面 $w^T x + b = 0$ 的距离为： $\frac{|w^T x + b|}{\|w\|}$

其中， $\|w\| = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$



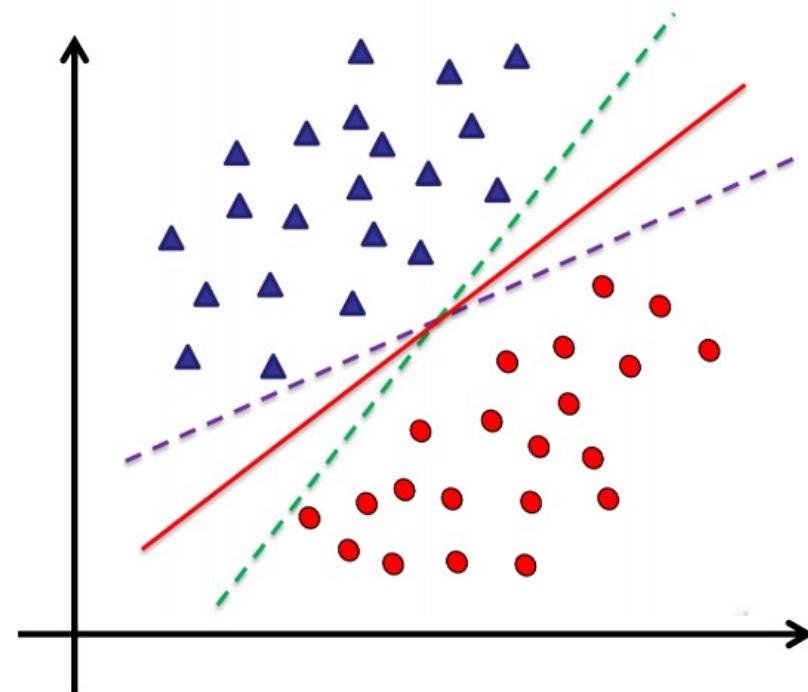
最大化所有点的距离和

83

支持向量机SVM

支持向量机概述

SVM大多用于线性可分数据，需要寻找一个最优超平面，不仅能正确分类样本，还使得离超平面最近的点的**间隔最大**

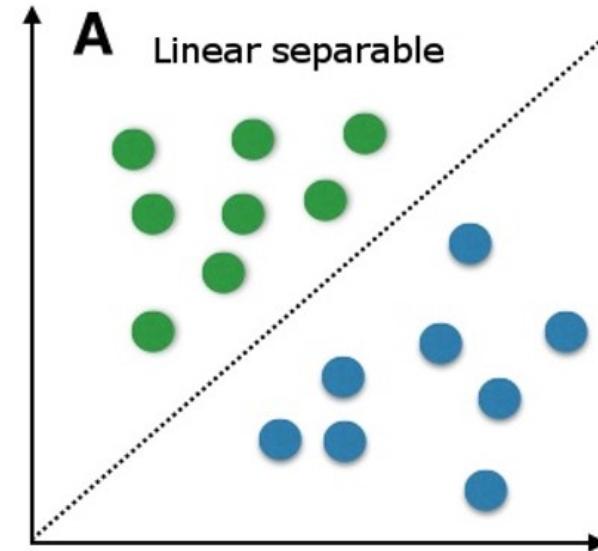


84

支持向量机SVM

线性可分

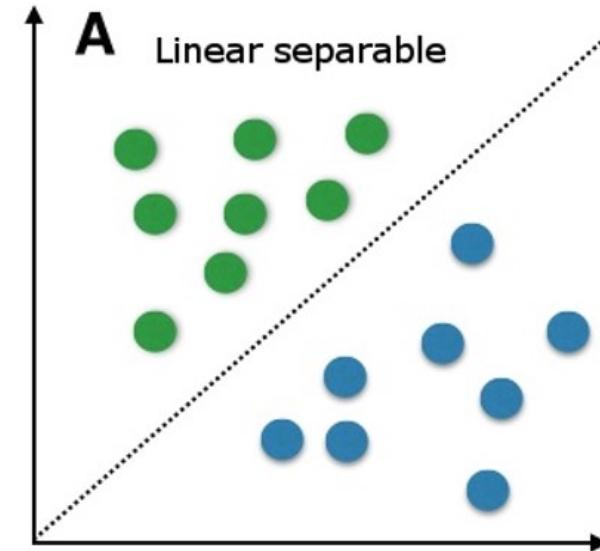
用一个**线性函数**将两类样本分开



支持向量机SVM

线性可分

用一个**线性函数**将两类样本分开

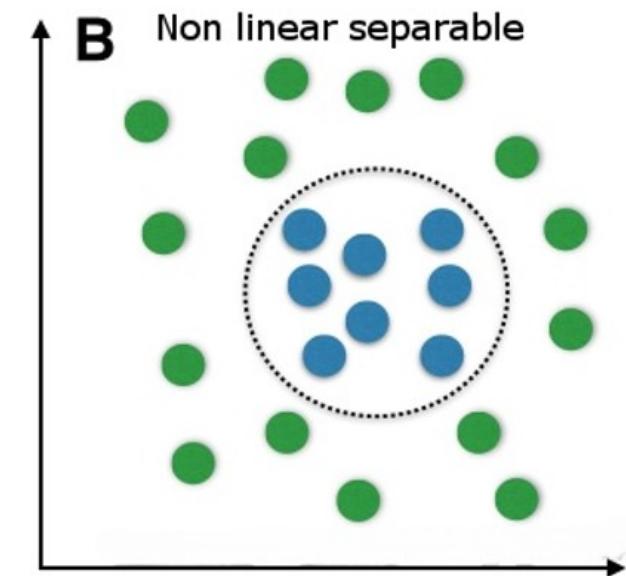


二维空间中的直线，三维空间中的平面以及高维空间中的超平面.....

支持向量机SVM

线性不可分

部分样本用**线性分类面**划分时会存在分类错误



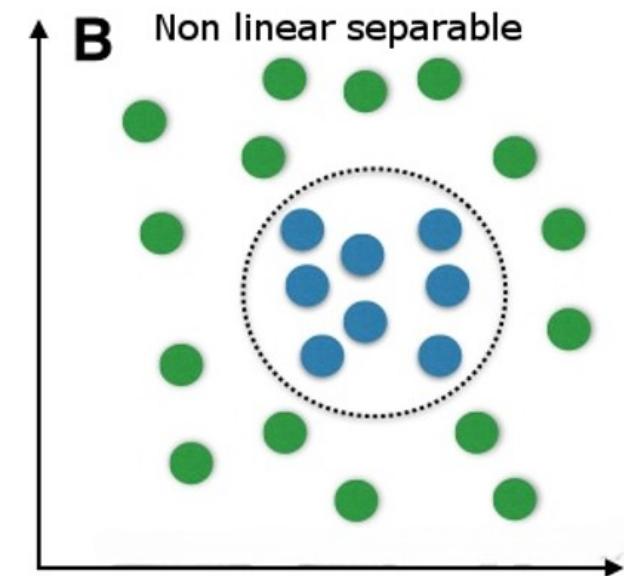
87

支持向量机SVM

线性不可分

部分样本用**线性分类面**划分时会存在分类错误

例如，经典的异或XOR问题



支持向量机SVM

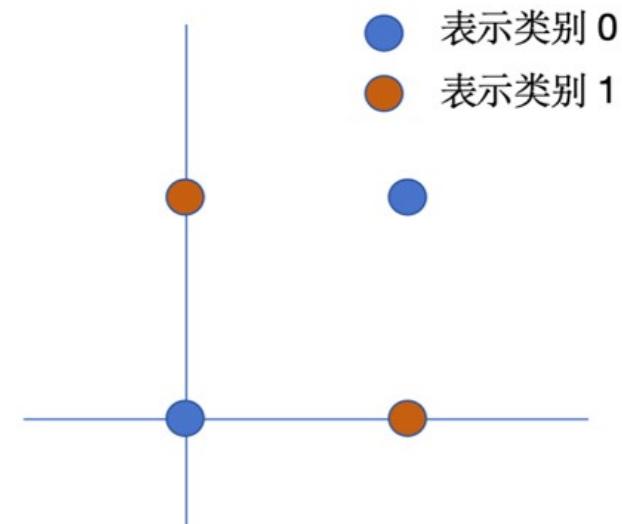
异或XOR线性不可分

逻辑运算中，XOR（异或）遵循以下规则：

当两个输入中恰好有一个为真（1）时，输出为真（1）；
否则输出为假（0）。

XOR的输入和输出可以视为如下坐标点：

(0, 0) 输出 0	(0, 1) 输出 1
(1, 0) 输出 1	(1, 1) 输出 0



89

支持向量机SVM

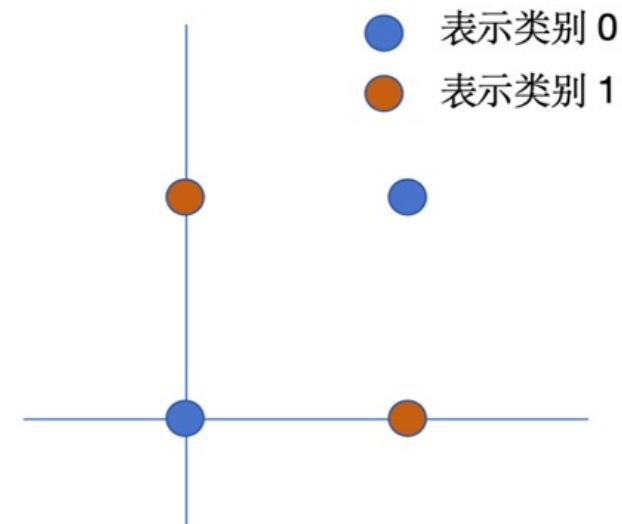
异或XOR线性不可分

逻辑运算中，XOR（异或）遵循以下规则：

当两个输入中恰好有一个为真（1）时，输出为真（1）；
否则输出为假（0）。

XOR的输入和输出可以视为如下坐标点：

(0, 0) 输出 0 (0, 1) 输出 1
(1, 0) 输出 1 (1, 1) 输出 0



在二维平面上，无法通过直线进行划分

90

支持向量机SVM

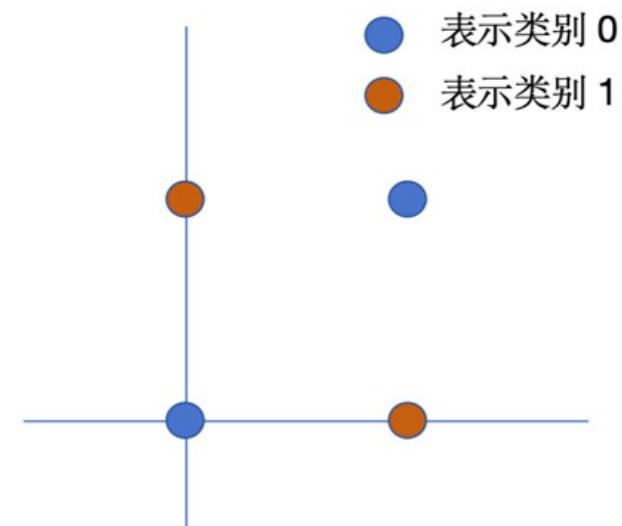
异或XOR线性不可分

逻辑运算中，XOR（异或）遵循以下规则：

当两个输入中恰好有一个为真（1）时，输出为真（1）；
否则输出为假（0）。

XOR的输入和输出可以视为如下坐标点：

(0, 0) 输出 0 (0, 1) 输出 1
(1, 0) 输出 1 (1, 1) 输出 0



在二维平面上，无法通过直线进行划分

考虑通过**核函数**或后续课程提及的**多层感知机**解决

91

Lecture Plan

- 线性回归
- 逻辑回归
- 贝叶斯网络
- 支持向量机SVM
- K-邻近算法KNN



92

机器学习基础

传统机器学习任务分为两大类：**预测和分类**

- **预测任务**→连续变量（房价、温度……）

1. 线性回归：用线性模型建立特征与目标的关系，预测连续值

- 2. 逻辑回归：先用线性模型预测

连续值，再利用**Sigmoid函数**将结

预测和分类的过渡 果映射到(0,1)区间，最后划分阈

- **分类任务**→离散类别（是否为垃圾邮件……）

值后，可用于**二分类任务**

3. 贝叶斯网络：基于概率理论，利用**贝叶斯定理**来分类数据，并假设特征间相互独立以简化计算

4. 支持向量机：通过寻找一个最优超平面来**分类**数据，以**最大化不同类别数据点之间的间隔**

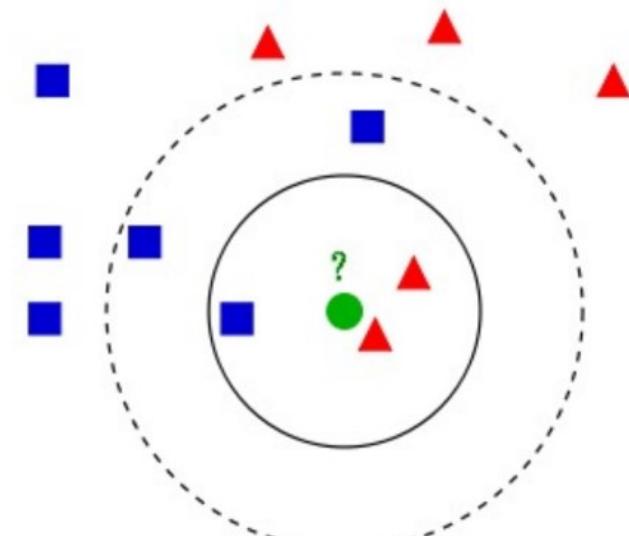
5. **K-邻近算法**：通过计算待分类样本与训练集中**K个最近邻样本的相似性**来直接**分类**数据

93

K-近邻算法 KNN

K-近邻算法

两类数据，一类为蓝色的方块，一类为红色的三角形



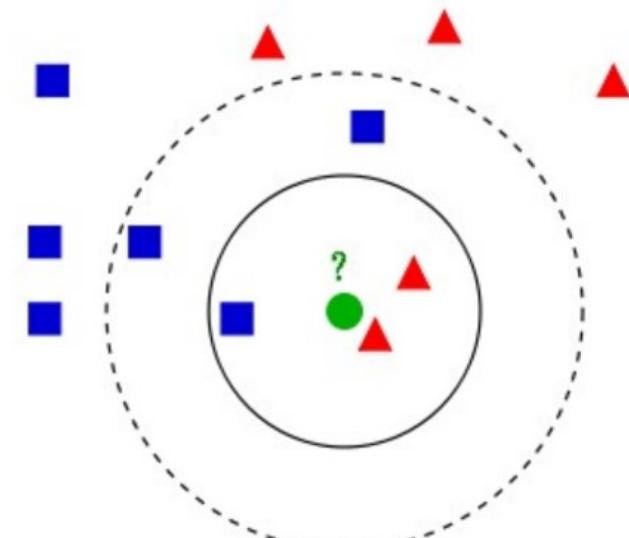
94

K-近邻算法 KNN

K-近邻算法

两类数据，一类为蓝色的方块，一类为红色的三角形

问：绿色的圆属于哪一类



95

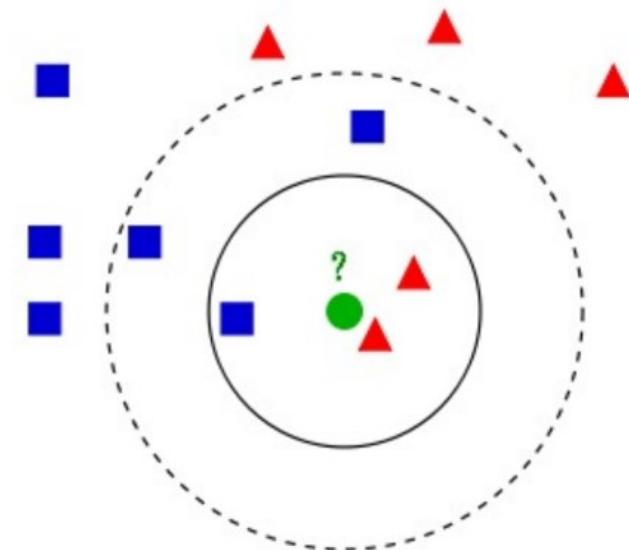
K-近邻算法 KNN

K-近邻算法

两类数据，一类为蓝色的**方块**，一类为红色的**三角形**

问：绿色的**圆**属于哪一类

KNN (K-nearest neighbor)核心思想：从训练集中找出与
绿色圆**最近**的K个样本，绿色圆属于**数量最多的那一类**。



96

K-近邻算法 KNN

K-近邻算法

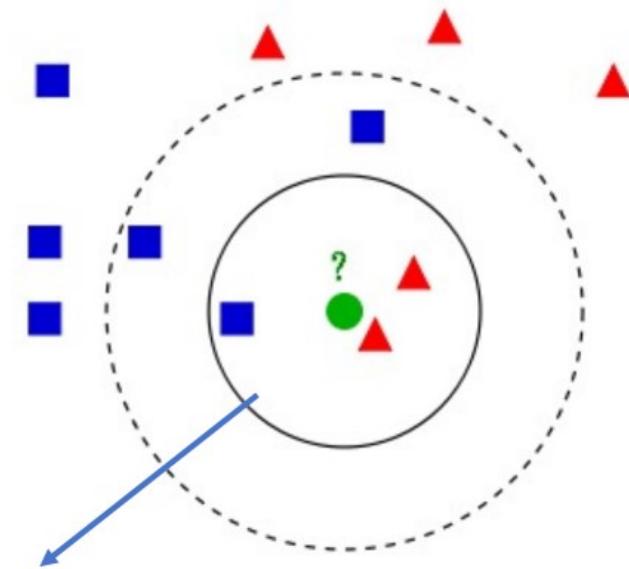
两类数据，一类为蓝色的方块，一类为红色的三角形

问：绿色的圆属于哪一类

KNN (K-nearest neighbor)核心思想：从训练集中找出与

绿色圆最近的K个样本，绿色圆属于数量最多的那一类。

离绿色圆最近的3个样本中，有2个为红色三角形，有1个为蓝色方块



97

K-近邻算法 KNN

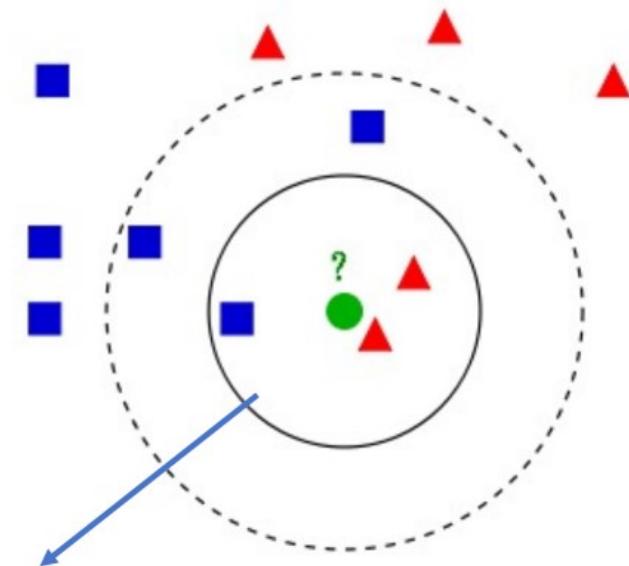
K-近邻算法

两类数据，一类为蓝色的方块，一类为红色的三角形

问：绿色的圆属于哪一类

KNN (K-nearest neighbor)核心思想：从训练集中找出与

绿色圆最近的K个样本，绿色圆属于数量最多那一类。



离绿色圆最近的3个样本中，有2个为红色三角形，有1个为蓝色方块

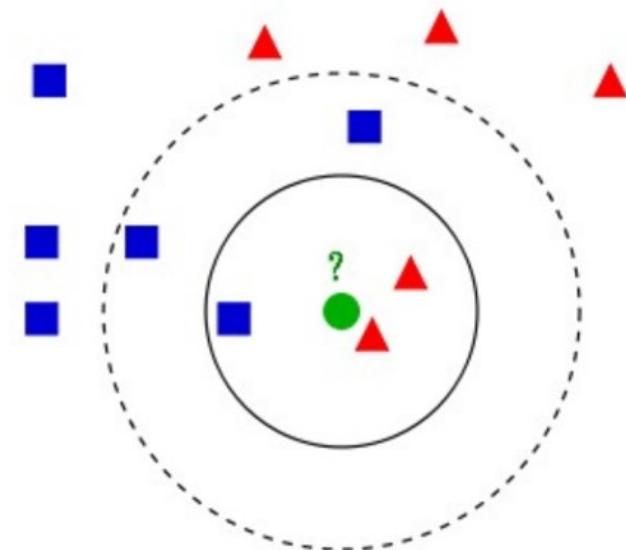
因此，绿色的圆属于红色的三角形

98

K-近邻算法 KNN

K-近邻算法流程

1. 计算测试样本与训练集中每个样本的距离
2. 按照距离的远近排序
3. 选取与当前测试样本最近的K个训练样本，作为该测试样本的邻居
4. 统计这K个样本的类别频次
5. K个样本里频次最高的类别，即为测试样本的类别

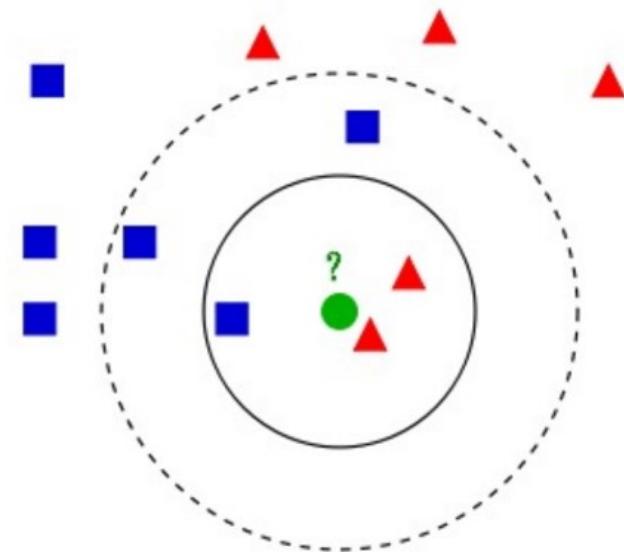


99

K-近邻算法 KNN

K-近邻算法 三要素

- **距离度量**: 用于衡量样本之间的相似性
- **K 值选择**: 选取最近的 K 个邻居
- **决策规则**: 决定如何根据邻居的结果做出预测



100

K-近邻算法 KNN

距离度量

特征向量: 数据集中每个样本被表示为一个特征向量

对于包含 n 个特征的样本实例, 其特征向量为 $X = (x_1, x_2, \dots, x_n)$, 其中 x_i 是对应特征的值。

特征空间: 一个 n 维实数空间 R^n

一个样本 → 一个特征向量

一个特征向量 → 特征空间中的一个点

特征空间中, 样本实例之间的**距离**反映它们之间的**相似程度**

101

K-近邻算法 KNN

距离度量

特征向量: 数据集中每个样本被表示为一个特征向量

对于包含 n 个特征的样本实例, 其特征向量为 $X = (x_1, x_2, \dots, x_n)$, 其中 x_i 是对应特征的值。

特征空间: 一个 n 维实数空间 R^n

一个样本 → 一个特征向量

一个特征向量 → 特征空间中的一个点

特征空间中, 样本实例之间的**距离**反映它们之间的**相似程度**

如何度量这个距离?

102

K-近邻算法 KNN

距离度量

欧氏距离(Euclidean distance)

$$x = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \dots \\ x^{(n)} \end{bmatrix} \quad y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \dots \\ y^{(n)} \end{bmatrix} \quad d(x, y) = \sqrt{\sum_{l=1}^n (x^{(l)} - y^{(l)})^2}$$

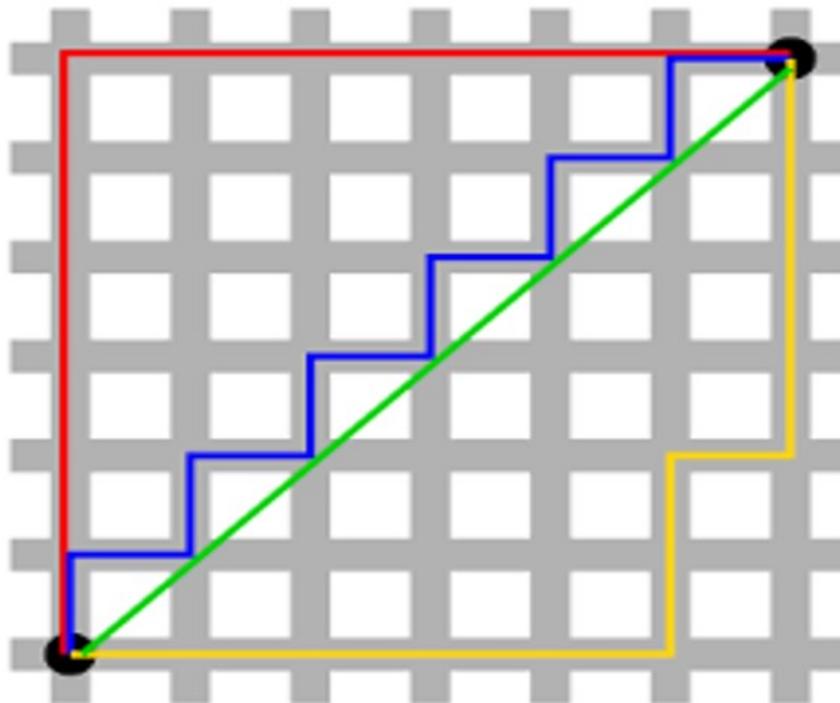
欧氏距离是一种常用的距离定义，指在 n 维空间中两个点之间的真实距离，或者向量的自然长度
(即该点到原点的距离)。

103

K-近邻算法 KNN

距离度量

曼哈顿距离(Manhattan distance)



$$d(x, y) = \sum_{l=1}^n (|x^{(l)} - y^{(l)}|)$$

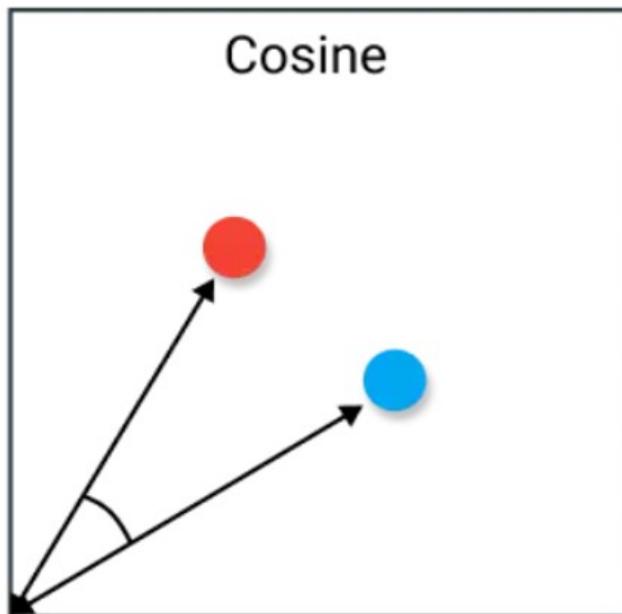
两点在南北方向上的距离加上在东西方向上的距离。

104

K-近邻算法 KNN

距离度量

余弦距离 (Cosine Distance)



$$D(x, y) = \cos(\theta) = \frac{x \cdot y}{\|x\| \|y\|}$$

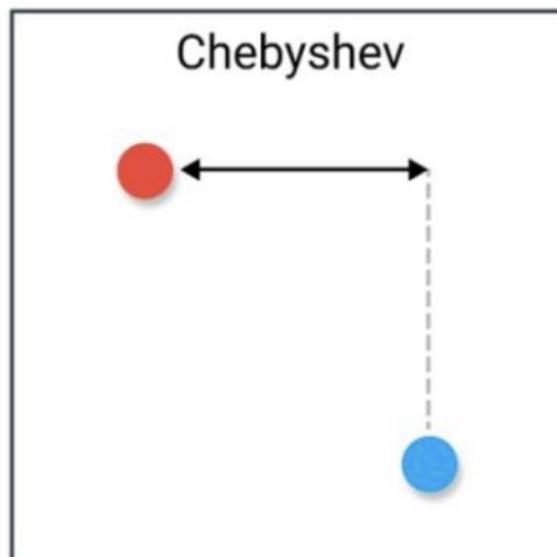
余弦距离是一种基于**余弦相似度**的距离度量方法，用于衡量两个向量在方向上的差异，忽略向量长度的影响。

105

K-近邻算法 KNN

距离度量

切比雪夫距离(Chebyshev distance)



$$d(x, y) = \max_l |x^{(l)} - y^{(l)}|$$

两个点之间的距离定义为其各坐标数值差绝对值的最大值。

106

K-近邻算法 KNN

距离度量

汉明距离(Hamming distance)

Hamming

A	1	0	1	1	0	0
	↑		↑			
B	1	1	1	0	0	0

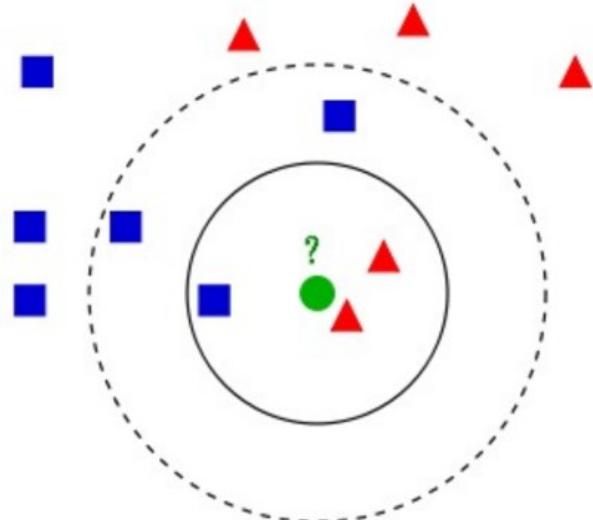
汉明距离表示两个字（相同长度）对应位不同的数量。

对两个字符串进行异或运算 (XOR, 相同为0, 不同为1) , 并统计结果为1的个数, 即该个数为汉明距离

107

K-近邻算法 KNN

K值的选择

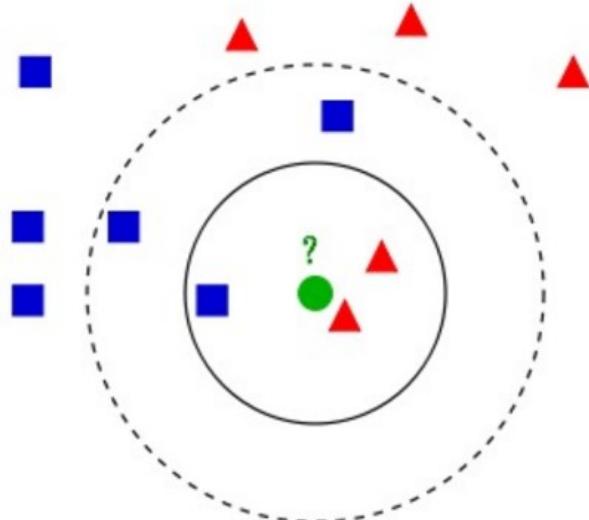


- 如果 $K = 3$ ，绿色圆点的最邻近的3个点是2个红色小三角形和1个蓝色正方形，则其属于红色三角形
- 如果 $K = 5$ ，绿色圆点的最邻近的5个邻居是2个红色三角形和3个蓝色正方形，则其属于蓝色正方形

108

K-近邻算法 KNN

K值的选择



- 如果 $K = 3$ ，绿色圆点的最邻近的3个点是2个红色小三角形和1个蓝色正方形，则其属于红色三角形
- 如果 $K = 5$ ，绿色圆点的最邻近的5个邻居是2个红色三角形和3个蓝色正方形，则其属于蓝色正方形

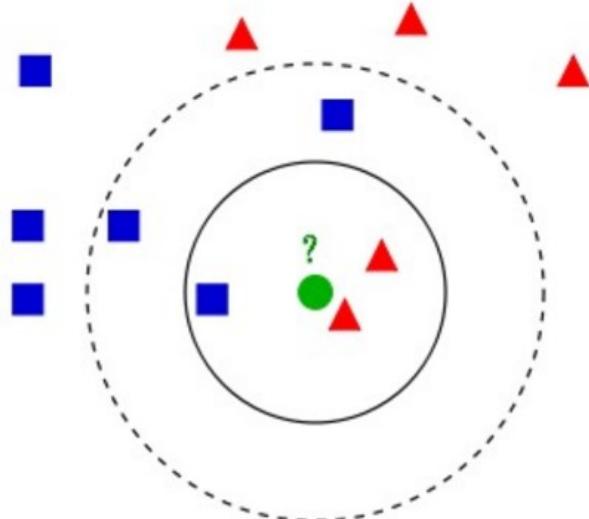
K值极端选择的影响：

- $K = 1$?

109

K-近邻算法 KNN

K值的选择



- 如果 $K = 3$ ，绿色圆点的最邻近的3个点是2个红色小三角形和1个蓝色正方形，则其属于红色三角形
- 如果 $K = 5$ ，绿色圆点的最邻近的5个邻居是2个红色三角形和3个蓝色正方形，则其属于蓝色正方形

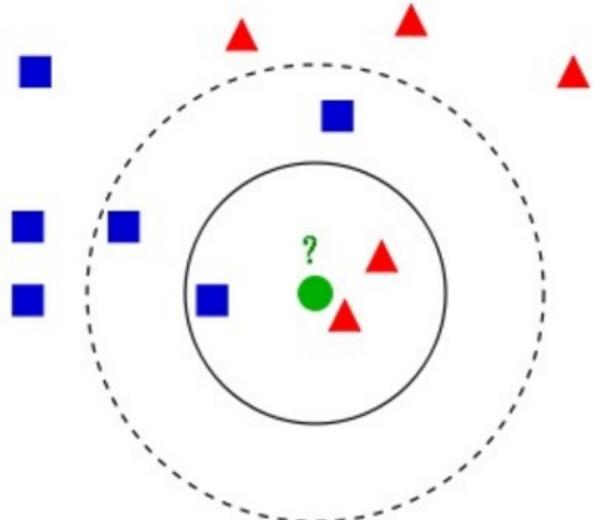
K值极端选择的影响：

- $K = 1$ ，仅考虑距离最近的样本，**容易过拟合，误差较大**

110

K-近邻算法 KNN

K值的选择



- 如果 $K = 3$ ，绿色圆点的最邻近的3个点是2个红色小三角形和1个蓝色正方形，则其属于红色三角形
- 如果 $K = 5$ ，绿色圆点的最邻近的5个邻居是2个红色三角形和3个蓝色正方形，则其属于蓝色正方形

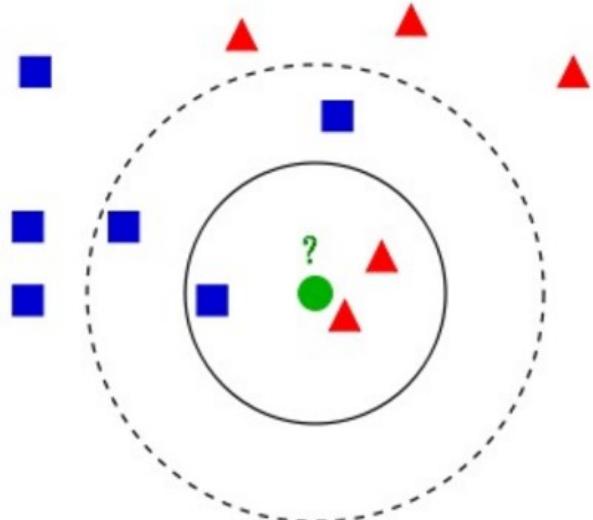
K值极端选择的影响：

- $K = 1$ ，仅考虑距离最近的样本，容易过拟合，误差较大
- $K = N$?

111

K-近邻算法 KNN

K值的选择



- 如果 $K = 3$ ，绿色圆点的最邻近的3个点是2个红色小三角形和1个蓝色正方形，则其属于红色三角形
- 如果 $K = 5$ ，绿色圆点的最邻近的5个邻居是2个红色三角形和3个蓝色正方形，则其属于蓝色正方形

K值极端选择的影响：

- $K = 1$ ，仅考虑距离最近的样本，容易过拟合，误差较大
- $K = N$ ，考虑全体训练样本，无意义

112

K-近邻算法 KNN

决策规则

- 多数投票

规则说明： k 个邻居中，哪个类别出现最多，就预测为哪个类别

- 加权投票

规则说明： k 个邻居中，离得更近的样本权重更大，最终加权下哪个类别最多，就预测为哪个类别

113

人工智能导论

Introduction to Artificial Intelligence

谢谢！



114