

一、逻辑回归 00:05

1. 人口增长 03:13

1) 19世纪初的人口增长研究 03:20

- **研究背景**：19世纪初数学家开始关注人口增长问题，比利时数学家阿道夫·凯特勒（Adolphe Quetelet）是该领域先驱者。需注意与阿道夫·希特勒区分。

2) 凯特勒的人口增长研究 03:30



- **模型建立**：用 $w(t)$ 表示 t 时刻人口量，其一阶导数 $\frac{dw(t)}{dt}$ 表示人口增长率

- **核心发现**：人口增长率与人口量呈线性关系，即 $\frac{dw(t)}{dt} = \beta w(t)$

3) 凯特勒模型的求解 04:32

- **求解方法**：

- 分离变量： $\frac{dw(t)}{w(t)} = \beta dt$
- 两边积分得： $\ln w(t) = \beta t + C$
- 指数化得解： $w(t) = Ae^{\beta t}$ ，其中 $A = e^C = w(0)$

- **模型特性**：呈现指数爆炸增长， A 表示初始人口量

4) 凯特勒模型的局限性 06:20

- **问题发现**：地球资源有限，人口不能无限增长
- **关键缺陷**：未考虑环境承载力，导致预测结果与现实不符

5) 威尔赫斯特的改进模型 07:34



P. F. VERHULST.

-
- 改进思路：在增长率中引入阻力项，改为 $\frac{dw(t)}{dt} = \sim\beta w(t)(\Omega - w(t))$
- 参数含义：
 - Ω ：环境最大承载人口量
 - $\sim\beta$ ：调整后的增长系数 ($\beta = \sim\beta\Omega$)

6) 威尔赫斯特模型求解 09:02

- 模型特性：
 - 当 $w(t) = \Omega / 2$ 时增长率最大
 - 当 $w(t) \rightarrow \Omega$ 时增长率 $\rightarrow 0$
- 合理性验证：通过绘制 $\frac{dw}{dt}$ 与 $w(t)$ 的关系曲线，呈现先增后减的抛物线形态

7) 威尔赫斯特模型合理性验证 10:46

$= A \exp(\beta t)$
↓
 $w(0)$

$\beta = \tilde{\beta} \Omega$

$\frac{dw}{dt} = \beta w(\Omega - w)$
 $P(t) = \frac{w(t)}{\Omega}$

$\frac{dP(t)}{dt} = \frac{d}{dt} \frac{w(t)}{\Omega}$
 $= \frac{1}{\Omega} \frac{dw(t)}{dt}$

-
- 关键特征点：
 - 顶点位置： $w(t) = \Omega / 2$ 处取得最大增长率
 - 零点位置： $w(t) = 0$ 和 $w(t) = \Omega$ 时增长率为零

8) 威尔赫斯特模型的比例研究 11:42

- 比例定义：引入 $P(t) = w(t) / \Omega$ 表示人口占比
- 转化方法：将原方程两边除以 Ω ，得到 $\frac{dP(t)}{dt} = \beta P(t)(1 - P(t))$

9) 威尔赫斯特模型的求解 (PT) 12:30

Logistic

简博士 bilibili

人口增长

$$\frac{dW(t)}{dt} = \beta W(t)$$

$$\frac{dW(t)}{W(t)} = \beta dt$$

$$\log W(t) = \beta t + C$$

$$\Rightarrow W(t) = \exp(\beta t + C)$$

$$= A \exp(\beta t)$$

$\downarrow$

$W(0)$

Ω : 最大值



P.F. VERHULST.

$\beta = \tilde{\beta} \Omega$

Verhulst

$$\frac{dW(t)}{dt} = \beta W(t) - \phi(W(t))$$

$$\frac{dW(t)}{dt} = \tilde{\beta} W(t) [\Omega - W(t)]$$

$$P(t) = \frac{W(t)}{\Omega}$$

Janneil (简博士数据分析吧)

统计学习方法

August 10, 2021 2 / 16

- 方程特性：逻辑斯蒂微分方程，描述受限增长
 - 实际意义：当 $P(t) \rightarrow 1$ 时增长停止，符合资源有限的实际约束
 - 参数关系：保持 $\beta = \tilde{\beta} \Omega$ 的原始定义不变
- #### 10) 威尔赫斯特模型的最终求解 14:41

Logistic

简博士 bilibili

$$\frac{dP(t)}{dt} = \frac{d \frac{W(t)}{\Omega}}{dt}$$

$$= \frac{1}{\Omega} \frac{dW(t)}{dt}$$

$$= \frac{1}{\Omega} \tilde{\beta} W(t) [\Omega - W(t)]$$

$$= \tilde{\beta} P(t) [1 - P(t)]$$

$$= \beta P(t) [1 - P(t)]$$



P.F. VERHULST.

Janneil (简博士数据分析吧)

统计学习方法

August 10, 2021 2 / 16

- 微分方程形式： $\frac{dP(t)}{dt} = \beta P(t)[1 - P(t)]$
- 求解方法：采用分离变量法，将关于 $P(t)$ 的部分移到等式左边，时间 t 相关项移到右边
- 推导过程：
 - 变形为 $\frac{1}{P(t)[1 - P(t)]} dP(t) = \beta dt$
 - 两边积分得到 $\ln P(t) - \ln(1 - P(t)) = \beta t + \alpha$

- 整理得 $\ln \frac{P(t)}{1-P(t)} = \beta t + \alpha$

11) 威尔赫斯特模型求解示例 15:00

Logistic
简博士 bilibili

人口增长

$$\frac{dW(t)}{dt} = \beta W(t)$$

$$\frac{dW(t)}{W(t)} = \beta dt$$

$$\log W(t) = \beta t + C$$

$$\Rightarrow W(t) = \exp(\beta t + C)$$

$$= A \exp(\beta t)$$

$\downarrow$

$W(0)$



P. F. VERHULST.

Ω : 最大值



Verhulst

$$\frac{dW(t)}{dt} = \beta W(t) - \phi(W(t))$$

$$\frac{dW(t)}{dt} = \tilde{\beta} W(t) [\Omega - W(t)]$$

$$P(t) = \frac{W(t)}{\Omega}$$

Janneil (简博士数据分析吧)
August 10, 2021
2 / 16

- 解的形式: $P(t) = \frac{e^{\alpha + \beta t}}{1 + e^{\alpha + \beta t}}$
- 参数解释:
 - β : 增长速率参数
 - α : 积分常数, 由初始条件确定
- 求解技巧: 通过指数运算将对数方程转化为显式表达式

12) 威尔赫斯特模型的含义 16:27

- 人口比例解释:
 - $P(t)$ 表示 t 时刻已存在人口比例
 - $1 - P(t)$ 表示尚未出现人口比例
 - 两者之和恒为 1
- 概率联系: 模型可视为二项分布, 已存在记为 1 (概率 $P(t)$), 未出现记为 0 (概率 $1 - P(t)$)

13) logistic 模型 17:28

Logistic
简博士 bilibili

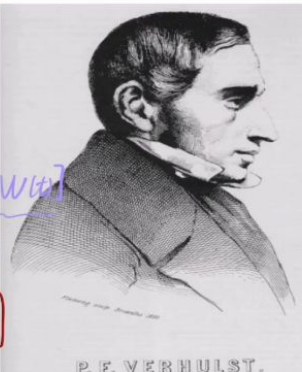
$$\frac{dP(t)}{dt} = \frac{d \frac{W(t)}{\Omega}}{dt}$$

$$= \frac{1}{\Omega} \frac{dW(t)}{dt}$$

$$= \frac{1}{\Omega} \tilde{\beta} W(t) [\Omega - W(t)]$$

$$= \tilde{\beta} P(t) [1 - P(t)]$$

$$\frac{1}{P(t) [1 - P(t)]} dP(t) = \tilde{\beta} dt$$



P. F. VERHULST.

$$\Rightarrow \log P(t) - \log(1 - P(t))$$

$$= \beta t + \alpha$$

$$\Rightarrow \log \frac{P(t)}{1 - P(t)} = \beta t + \alpha$$

$$\Rightarrow P(t) = \frac{\exp(\alpha + \beta t)}{1 + \exp(\alpha + \beta t)}$$

$P(t)$	1
$1 - P(t)$	0

Logistic

Janneil (简博士数据分析吧)
August 10, 2021
2 / 16

- 标准形式: $P(t) = \frac{1}{1 + e^{-(\alpha + \beta t)}}$
- 发明者: P.F. Verhulst (韦尔赫斯特)
- 函数特性:
 - 值域在[0,1]之间
 - 呈现S型增长曲线
 - 具有上下渐近线

14) logistic模型命名由来 17:42

- 命名可能原因:
 - 与对数增长曲线相对应 (法语原文"logistique")
 - 与数学术语后缀"-tic"保持一致 (如arithmetic, geometric)
 - 反映二分类问题的逻辑特性 (是/否, 1/0)
- 历史趣闻: 实际命名原因可能仅是作者个人偏好

15) 对数增长曲线与logistic曲线 17:53

Logistic
简博士 bilibili

人口增长

$$\frac{dW(t)}{dt} = \beta W(t)$$

$$\frac{dW(t)}{W(t)} = \beta dt$$

$$\log W(t) = \beta t + C$$

$$\Rightarrow W(t) = \exp(\beta t + C)$$

$$= A \exp(\beta t)$$

$\downarrow$

$W(0)$



P. F. VERHULST.

Ω : 最大值

Verhulst

$$\frac{dW(t)}{dt} = \beta W(t) - \phi(W(t))$$

$$\frac{dW(t)}{dt} = \tilde{\beta} W(t) [\Omega - W(t)]$$

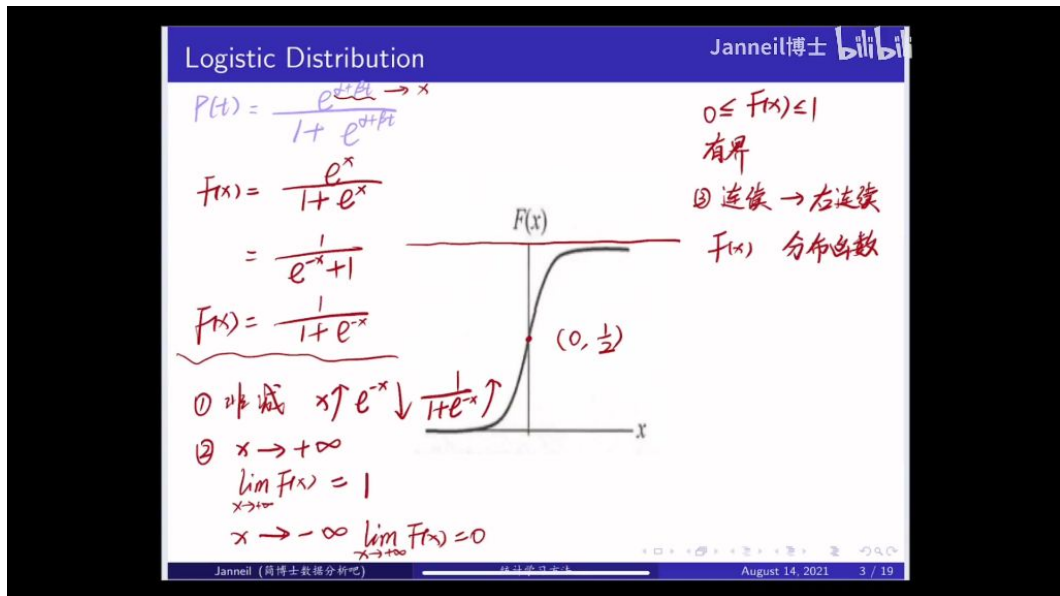
$$P(t) = \frac{W(t)}{\Omega}$$

$\beta = \tilde{\beta} \Omega$

Janneil (简博士数据分析吧)
August 10, 2021
2 / 16

- 对数增长曲线:
 - 无限制的指数增长: $W(t) = Ae^{\beta t}$

- 适用于资源无限情况
 - **logistic曲线:**
 - 受环境容量限制的S型增长
 - 增长速率随人口接近上限而减缓
 - **核心区别:** logistic模型引入了 $1 - P(t)$ 的限制因子
- 16) logistic分布



- 分布函数: $F(x) = \frac{1}{1 + e^{-x}}$
- 密度函数: $f(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$
- 特性:
 - 有界性: $0 \leq F(x) \leq 1$
 - 连续性: 右连续函数
 - 极限性质: $\lim_{x \rightarrow +\infty} F(x) = 1, \lim_{x \rightarrow -\infty} F(x) = 0$

17) 课程总结与下节预告 21:31

- **本节重点:**
 - 掌握了logistic函数的推导过程
 - 理解了人口增长模型的数学表达
 - 认识了logistic分布的基本性质
- **下节内容:** logistic回归模型的构建与应用

2. 逻辑回归模型的定义 37:02

1) 回归模型的定义 37:05

Logistic Distribution
Janneil博士 bilibili

Definition

设X是连续型随机变量，X服从逻辑斯谛分布是指X具有以下的分布函数和分布密度：

$$F(x) = P(X \leq x) = \frac{1}{1 + e^{-\frac{(x-\mu)}{\gamma}}} \quad (1)$$

$$f(x) = F'(x) = \frac{e^{-\frac{(x-\mu)}{\gamma}}}{\gamma(1 + e^{-\frac{(x-\mu)}{\gamma}})^2} \quad (2)$$

式中 μ 是位置参数， γ 是形状参数。

- 分布函数是一条S形曲线，该曲线也被称为sigmoid曲线，关于点 $(\mu, \frac{1}{2})$ 中心对称。
- 概率密度函数一条钟型曲线，中间高两端低，关于 $x = \mu$ 对称，在此处取得最大值 $1/4\gamma$ 。

Janneil (同济大学数据分析师)
August 14, 2021
5 / 19

- 本质：虽然名为回归模型，但实际上是一个二值概率分布模型，分别计算 $y = 1$ 和 $y = 0$ 两个类别的概率

- 数学形式：基于Logistic分布函数 $F(x) = \frac{1}{1 + e^{-\frac{(x-\mu)}{\gamma}}}$ ，其概率密度函数 $f(x) = \frac{e^{-\frac{(x-\mu)}{\gamma}}}{\gamma(1 + e^{-\frac{(x-\mu)}{\gamma}})^2}$

2) 回归模型中的变量与参数 37:30

- 输入变量：x为N维输入变量
- 输出变量：y为二分类输出，取值0或1
- 模型参数：
 - ω ：N维权值参数（权重向量）
 - b：偏置项（标量参数）

3) 模型参数的个数与原因 38:14

- 参数总数：共 $n + 1$ 个参数（ ω 有n维，加上b）
- 类比说明：类比线性回归 $y = ax + b$ ，1维输入对应2个参数；扩展到N维输入自然对应 $n + 1$ 个参数

4) 模型类别的判断方法 40:01

- 基本方法：计算 $P(y = 1|x_0)$ 和 $P(y = 0|x_0)$ ，比较两者大小
 - 若 $P(y = 1|x_0) \geq P(y = 0|x_0)$ ，则归为类别1
 - 否则归为类别0
- 概率最大化：体现了最大概率决策思想

5) 模型类别的简化判断方法 41:51

- 阈值法：直接比较 $P(y = 1|x_0)$ 与0.5的关系
 - ≥ 0.5 归为类别1
 - < 0.5 归为类别0
- 数学简化：可通过调整分布函数形式使表达式更简洁

3. 广义模型和线性模型的联系 42:33

1) 线性回归模型公式及变形 42:36

Regression

一元线性
 $y = x\beta + \alpha + \varepsilon$

多元
 $y = \vec{x}\vec{\beta} + \alpha + \varepsilon$

$\vec{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$

$\vec{x} = (x_1, x_2, \dots, x_p)$

$E(y) = \vec{x}\vec{\beta} + \alpha$

线性

广义线性模型
logistic f

$f(y) = \vec{x}\vec{\beta} + \alpha$

Regress

简博士 bilibili

统计学习方法

August 14, 2021 6 / 19

- 标准形式: $y = x\beta + \alpha + \varepsilon$
- 向量形式: $y = \vec{x} \cdot \vec{\beta} + \alpha$
- 参数合并: 偏置项 α 可并入权重向量 $\vec{\beta}$

2) 逻辑回归模型公式及变形 43:02

- 参数扩充:
 - 输入 x 增加常数1维度 (从 n 维 $\rightarrow n+1$ 维)
 - 权重 ω 相应增加1维 (包含原偏置项 b)
- 简化形式: $\omega \cdot x$ (内积形式)

3) 逻辑回归模型的好处: 解决分类问题 45:38

- 核心优势: 将离散分类问题转化为连续概率估计
- 连接函数: 通过sigmoid函数将线性组合 $\omega \cdot x$ 映射到 $(0,1)$ 概率区间
- 非线性转化: 突破输入输出间的线性关系限制

4) 逻辑回归模型的好处: 输入变量自由 50:40

- 变量兼容性:
 - 替代单位阶跃函数, 解决不连续问题
 - 输入变量既可以是连续的也可以是离散的
- 平滑性: Logistic曲线连续可微, 便于求导优化

5) 逻辑回归模型的估计方法: 最大自然估计法 50:52

- 估计原理: 概率最大化思想
- 方法选择:
 - 最小二乘法不适用
 - 采用最大似然估计 (参考朴素贝叶斯章节)
- 参数估计: 通过最大化似然函数求解 ω 和 b

4. 逻辑回归模型的参数估计方法 51:58

1) 逻辑回归模型 52:27

Definition (二项逻辑斯谛回归模型)

设 $x \in R^n$ 是输入, $Y \in \{0, 1\}$ 是输出, $\omega \in R^n$ 和 $b \in R$ 是参数, ω 称为权值参数, b 称为偏置, $\omega \cdot x$ 表示 ω 与 x 的内积。二项逻辑斯谛回归模型为以下的条件概率:

$$P(Y = 1|x) = \frac{e^{\omega \cdot x + b}}{1 + e^{\omega \cdot x + b}} \quad (3)$$

$$P(Y = 0|x) = \frac{1}{1 + e^{\omega \cdot x + b}} \quad (4)$$

模型定义: 设 $x \in R^n$ 是输入, $Y \in \{0, 1\}$ 是输出, $\omega \in R^n$ 和 $b \in R$ 是参数, ω 称为权值参数, b 称为偏置。模型条件概率为:

$$P(Y = 1|x) = \frac{e^{\omega \cdot x + b}}{1 + e^{\omega \cdot x + b}}$$

$$P(Y = 0|x) = \frac{1}{1 + e^{\omega \cdot x + b}}$$

参数扩充: 为简化表示, 将权重向量和输入向量扩充为 $\omega = (\omega^{(1)}, \omega^{(2)}, \dots, \omega^{(n)}, b)^T$ 和 $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)}, 1)^T$, 此时模型可简化为:

$$P(Y = 1|x) = \frac{e^{\omega \cdot x}}{1 + e^{\omega \cdot x}}$$

$$P(Y = 0|x) = \frac{1}{1 + e^{\omega \cdot x}}$$

决策规则: 输入实例 x , 计算两类别的概率, 将实例 x 分到概率大的那一类去。

2) 参数估计采用最大似然估计法 01:00:25

Outline

① 逻辑回归模型

② 参数估计方法

似然函数构造:

- 单个样本 (x_i, y_i) 的概率可表示为 $(1 - p_i)^{1 - y_i} p_i^{y_i}$ ，其中 $p_i = \frac{e^{\omega \cdot x_i}}{1 + e^{\omega \cdot x_i}}$

对于N个独立样本的训练集，似然函数为：

- $L(\omega) = \prod_{i=1}^N P(Y = y_i | x_i) = \prod_{i=1}^N (1 - p_i)^{1 - y_i} p_i^{y_i}$

● 对数似然函数：

为简化计算，取对数得到：

- $\log L(\omega) = \sum_{i=1}^N [y_i \log p_i + (1 - y_i) \log(1 - p_i)]$

进一步化简可得：

- $\sum_{i=1}^N [y_i (\omega \cdot x_i) - \log(1 + \exp(\omega \cdot x_i))]$

● 参数求解方法：

- 解析法：对对数似然函数求导并令导数为零，但通常难以得到显式解
- 枚举法：尝试所有可能的 ω 值，选择使似然函数最大的解（计算量大）
- 迭代法：
 - 牛顿法：基于泰勒二阶展开，收敛速度快
 - 梯度上升法：将对数似然函数最大化问题转化为梯度上升问题

- 优化技巧：由于概率 p_i 与 ω 的关系是非线性的，通常需要采用数值优化方法求解参数估计值。

二、知识小结

知识点	核心内容	考试重点/易混淆点	难度系数
逻辑回归定义	二值概率分布模型，连接线性回归与分类问题	表面是回归实际解决分类问题	★★
人口增长模型	威尔赫斯特修正凯特勒指数模型，引入环境容量限制	微分方程推导过程及参数含义	★★★★
Logistic函数由来	从人口比例微分方程推导出S形曲线	函数命名与对数增长曲线的对应关系	★★★
模型参数构成	ω 权重向量+b偏置项→扩充为n+1维向量	参数维度与输入特征的对应关系	★★
分类决策规则	比较 $P(y=1)$ 与 $P(y=0)$ 概率/直接与0.5阈值比较	概率最大化思想的数学表达	★★

连接函数优势	用sigmoid替代单位阶跃函数保证连续性	函数可微性对参数估计的影响	★★★
似然函数构建	二项分布的联合概率→对数似然函数转换	连乘转连加的技巧及数学依据	★★★★
参数估计方法	极大似然估计→牛顿法/梯度上升法求解	显示解不可得时的迭代策略选择	★★★★
历史典故	凯特勒指数模型→威尔赫斯特修正过程	模型假设与实际人口规律的契合度	★★
命名渊源	法语词源考证/与几何级数命名对应关系	中文"逻辑"翻译的多重可能解释	★