标题 title

作者 *author*

2023 年 8 月 9 日

前言

目录

前言 i

第一部分　科学的逻辑 1

第一章　合情推理 2

§1.1 回顾：命题逻辑的演绎推理 2
§1.2 合情推理的数学模型 4
1.2.1 似然，合情推理的原则 4
1.2.2 似然与概率 6
§1.3 合情推理的归纳强论证 8
1.3.1 先验与基率谬误 8
1.3.2 归纳强论证 9
1.3.3 有效论证和归纳强论证的比较 12

第二章　Markov 链与决策 15

§2.1 Markov 链 15
§2.2 Markov 奖励过程（MRP） 19
§2.3 Markov 决策过程（MDP） 22
§2.4 隐 Markov 模型（HMM） 26
2.4.1 评估问题 27
2.4.2 解释问题 28

第二部分　信息与数据 30

第三章　信息论基础 31

§3.1 熵 31
3.1.1 概念的导出 31
3.1.2 概念与性质 34
3.1.3 熵与通信理论 39
§3.2 Kullback-Leibler 散度 42
3.2.1 定义 42
3.2.2 两个关于信息的不等式 44
3.2.3 在机器学习中的应用：语言生成模型 45
§3.3 附录：Shannon 定理的证明 46
§3.4 习题 47
§3.5 章末注记 49

第四章 Johnson-Lindenstrauss 引理 **51**
§4.1 机器学习中的数据 51
§4.2 矩法与集中不等式 52
§4.3 J-L 引理的陈述与证明 56
§4.4 J-L 引理的应用 60
§4.5 习题 61
§4.6 章末注记 61

第五章 差分隐私 **62**
§5.1 数据隐私问题 62
§5.2 差分隐私的定义与性质 64
§5.3 差分隐私的应用 68
5.3.1 随机反应算法 68
5.3.2 全局灵敏度与 Laplace 机制 69
5.3.3 DP 版本 Llyod 算法 71
§5.4 差分隐私与信息论 72
§5.5 习题 73
§5.6 章末注记 73

第三部分 决策与优化 **74**

第六章 凸分析 **75**

§6.1 决策与优化的基本原理 75
6.1.1 统计决策理论 75
6.1.2 优化问题 76
6.1.3 例子：网格搜索算法 79
§6.2 凸函数 81
§6.3 凸集 84
6.3.1 基本定义和性质 84
6.3.2 分离超平面定理 86

第七章 对偶理论 **88**
§7.1 条件极值与 Lagrange 乘子法 89
§7.2 Karush–Kuhn–Tucker 条件 92
§7.3 Lagrange 对偶 95
7.3.1 Lagrange 定理 95
7.3.2 弱对偶定理，强对偶定理 99
§7.4 应用：支持向量机（SVM） 103

第八章 不动点理论 **106**
§8.1 Banach 不动点定理 106
§8.2 Brouwer 不动点定理 109
§8.3 不动点的一般视角 112

第四部分 逻辑与博弈 **113**

第九章 动态博弈 **114**
§9.1 输赢博弈 114
§9.2 随机博弈（Markov 博弈） 119

第十章 静态博弈 **125**
§10.1 正则形式博弈 125
10.1.1 生成对抗网络 126
10.1.2 混合策略 128
§10.2 不完全信息博弈（Bayes 博弈） 129

第五部分　认知逻辑 **134**

第十一章　模态逻辑基础 **135**

§11.1 模态逻辑的起源 135

11.1.1 三段论 135

11.1.2 非经典逻辑 136

§11.2 模态语言 137

§11.3 Kripke 语义与框架语义 140

§11.4 模态可定义性 145

第十二章　认知逻辑与共同知识 **147**

§12.1 “泥泞的孩童”谜题 147

§12.2 认知逻辑的基本模型与性质 149

12.2.1 “泥泞的孩童”再回顾 153

12.2.2 Aumann 结构 154

§12.3 对不一致达成一致 155

§12.4 Rubinstein 电子邮件博弈 158

附录 A　线性代数基础 **162**

§A.1 线性空间 162

§A.2 线性映射 165

附录 B　微积分基础 **163**

附录 C　概率统计基础 **164**

第一部分

科学的逻辑

第二部分

信息与数据

第三部分

决策与优化

第四部分

逻辑与博弈

第五部分

认知逻辑

附录 A 线性代数基础

§A.1 线性空间

从动机上说，线性空间试图将 $\mathbb{R}^n$ 或者 $\mathbb{C}^n$ 这样的集合连同他们上面的代数结构抽象出来。除此之外，函数和无穷数列的集合也是非常重要的对象，比如说 $\mathbb{R}$ 上的连续函数组成的集合 $C(\mathbb{R})$，或者具有“模长”的无穷实数列（$\ell^2(\mathbb{R})$-空间）：

$$\ell^2(\mathbb{R}) = \left\{ x = (x_1, x_2, \cdots) \in \mathbb{R}^\infty : \sum_{i=1}^{\infty} x_i^2 < \infty \right\}.$$

我们将这些对象的共性抽象出来，得到线性空间的概念。线性空间都是基于某个域定义的，我们先给出域的定义。

定义 A.1 (域) 一个域是一个集合 F，其上定义了两种二元运算：加法 $+$ 和乘法 $\cdot$，他们都是 $F \times F$ 到 F 的映射，满足下面的公理：

1. (结合律) 对于任意的 $a, b, c \in F$, 有 $(a+b)+c = a+(b+c)$ 和 $(a \cdot b) \cdot c = a \cdot (b \cdot c)$;
2. (交换律) 对于任意的 $a, b \in F$，有 $a+b = b+a$ 和 $a \cdot b = b \cdot a$;
3. (分配律) 对于任意的 $a, b, c \in F$，有 $a \cdot (b+c) = a \cdot b + a \cdot c$。
4. (单位元) 存在唯一的两个元素 $0, 1 \in F$，使得对于任意的 $a \in F$，有 $a+0 = a$ 和 $a \cdot 1 = a$;
5. (加法逆元) 对于任意的 $a \in F$，存在唯一 $b \in F$，使得 $a+b = 0$，记 b 作 $-a$;
6. (乘法逆元) 对于任意的 $a \in F$，如果 $a \neq 0$，则存在唯一 $b \in F$，使得 $a \cdot b = 1$，记 b 作 a^{-1};

通常将 $a \cdot b$ 写作 ab，并且乘法的优先级高于加法，即 $ab + c = (ab) + c$。

域的重要例子包括有理数域 $\mathbb{Q}$，实数域 $\mathbb{R}$ 和复数域 $\mathbb{C}$，他们都是无限域。我们将在后面的内容中使用这些域。

定义 A.2 (线性空间，向量空间) 设 V 是一个集合，F 是一个域。如果在 V 上定义了两种运算：加法 $+$ 和数乘 $\cdot$，使得 V 满足下面的公理：

1. （V 的结合律）对于任意的 $x,y,z \in V$，有 $(x+y)+z = x+(y+z)$；
2. （V 的交换律）对于任意的 $x,y \in V$，有 $x+y = y+x$；
3. （加法零元）存在唯一的元素 $0 \in V$，使得对于任意的 $x \in V$，有 $x+0 = x$；
4. （加法逆元）对于任意的 $x \in V$，存在唯一 $y \in V$，使得 $x+y=0$，记 y 作 $-x$；
5. 对于任意的 $x \in V$，有 $1 \cdot x = x$；
6. 对于任意的 $a,b \in F$ 和 $x \in V$，有 $(ab) \cdot x = a \cdot (b \cdot x)$；
7. 对于任意的 $a \in F$ 和 $x,y \in V$，有 $a \cdot (x+y) = a \cdot x + a \cdot y$；
8. 对于任意的 $a,b \in F$ 和 $x \in V$，有 $(a+b) \cdot x = a \cdot x + b \cdot x$。

则称 V 是一个**F-线性空间**，简称**线性空间**，也称**向量空间**。V 中的元素被称为**向量**. 通常将数乘 $a \cdot x$ 写作 ax，并且乘法的优先级高于加法，即 $a \cdot x + y = (a \cdot x) + y$。

"线性"一词的含义是指的 $ax+by$ 这种形式的数学对象，线性代数就是研究这种对象的学科。线性空间的典型例子包括：

- $\mathbb{R}^n$ 和 $\mathbb{C}^n$.
- $M_{m\times n}(F)$，即所有 $m \times n$ 矩阵组成的集合.
- $C(\mathbb{R})$，即 $\mathbb{R}$ 上的连续函数组成的集合.
- $C^k(\mathbb{R})$，即 $\mathbb{R}$ 上的 k 次连续可微函数组成的集合.
- $\ell^2(\mathbb{R})$，即所有二次可和的实数序列组成的集合.

如同所有其他的代数结构，线性空间也有各式各样构造新的线性空间的方法。为了看出来线性空间本质的特性，我们有如下引理：

引理 A.1 设 V 是 F-线性空间，W 是 V 的一个子集。则 W 是一个线性空间当且仅当对任意 $a,b \in F$ 和 $x,y \in W$，有 $ax+by \in W$。

证明 按照定义即可验证。 □

我们给 $ax+by$ 这样的对象一个正式的定义。

定义 A.3 (线性组合) 设 V 是 F-线性空间，$x_1,\cdots,x_n\in V$，$a_1,\cdots,a_n\in F$，则称 $a_1x_1+\cdots+a_nx_n$ 是 $x_1,\cdots,x_n$ 的一个**线性组合**。

接下来，基于某些特定的线性空间，我们构造各种新的线性空间。

定义 A.4 (线性子空间) 设 V 是 F-线性空间，W 是 V 的一个子集。如果 W 是一个线性空间，则称 W 是 V 的一个**线性子空间**。

例如，$\mathbb{Q}$ 是 $\mathbb{R}$ 的一个线性子空间，但 $\mathbb{Z}$ 不是 $\mathbb{R}$ 的一个线性子空间。再比如，当 $k<l$，$C^k(\mathbb{R})$ 是 $C^l(\mathbb{R})$ 的一个线性子空间。

定义 A.5 (乘积空间) 设 $V_1,\cdots,V_n$ 是 F-线性空间，则 $V_1\times\cdots\times V_n$ 是一个 F-线性空间，其中加法和数乘分别定义为

$$
\begin{aligned}
&(x_1,\cdots,x_n)+(y_1,\cdots,y_n)=(x_1+y_1,\cdots,x_n+y_n),\\
&a(x_1,\cdots,x_n)=(ax_1,\cdots,ax_n).
\end{aligned}
$$

例如，$\mathbb{R}^n$ 就是 n 个 $\mathbb{R}$ 的乘积空间，$M_{m\times n}(F)$ 就是 $m\times n$ 个 F 的乘积空间。

接下来，我们按照表示论的观点，引入基的概念。线性空间是一个非常抽象的数学概念，因此我们需要一些具体的元素去表示这整个空间。

定义 A.6 (生成集) 设 V 是 F-线性空间，$S\subseteq V$，如果 V 中的每一个元素都是 S 的线性组合，则称 S 是 V 的一个**生成集**。

更一般地，任意一个 $S\subseteq V$，我们可以定义 S **生成的线性子空间**为所有 S 的线性组合的集合，记为 $\mathrm{Span}(S)$。

我们希望用尽可能少的元素来表示整个线性空间，为此，我们需要把“可表示”这样的概念严格化。

定义 A.7 (线性相关) 设 V 是 F-线性空间，$S\subseteq V$，如果存在 $x_1,\cdots,x_n\in S$，$a_1,\cdots,a_n\in F$，使得 $a_1x_1+\cdots+a_nx_n=0$，且至少有一个 $a_i\neq 0$，则称 S 是**线性相关**的，否则称 S 是**线性无关**的。

S 线性相关意味着 S 中的一些元素可以被另一些元素的线性组合表示出来，因而 S 中有一些冗余。线性无关意味着 S 中的元素都是必要的，没有冗余。由此，我们可以给出基的定义。

定义 A.8 (基) 设 V 是 F-线性空间，$S \subseteq V$，如果 S 是线性无关的，并且 $\mathrm{Span}(S) = V$，则称 S 是 V 的一个**基**。

线性空间的一个核心定理是基的存在性定理。

定理 A.1 (基的存在性定理) 设 V 是 F-线性空间，则 V 中存在一个基。

要注意，这一定理并不是平凡的。首先，基是线性无关的集合，所以 V 本身通常就不是基。此外，这一定理要求有一个线性无关的集合 $S \subseteq V$，任意向量 $x \in V$ 都可以用 S 中有限个元素的线性组合来表示，这样的 S 并不容易找到。该定理的证明是构造性的，这一构造依赖于选择公理（或者 Zorn 引理），我们在此略去。

基的典型例子包括：

- $\mathbb{R}^n$ 的标准基是 $\{e_1, \cdots, e_n\}$，其中 e_i 是第 i 个分量为 1，其余分量为 0 的向量；
- $M_{m\times n}(F)$ 的标准基是 $\{E_{ij} : 1 \le i \le m, 1 \le j \le n\}$，其中 E_{ij} 是第 i 行第 j 列为 1，其余元素为 0 的矩阵；
- $\ell^2(\mathbb{R})$ 的标准基是 $\{e_1, e_2, \cdots\}$，其中 e_i 是第 i 个分量为 1，其余分量为 0 的实数列。

线性空间的基可以衡量线性空间的复杂程度，基元素越少，线性空间越简单。我们可以定义维数来衡量线性空间的复杂程度。

定义 A.9 (维数) 设 V 是 F-线性空间，如果 V 的一个基有限，则称 V 是**有限维**的，否则称 V 是**无限维**的。有限维线性空间的基的元素个数称为 V 的**维数**，记为 $\dim V$。

这一定义隐含的事实是，如果 V 有有限基，那么所有基都是有限的，并且任意两个基的元素个数相同。我们这里略去证明。

例如，$\mathbb{R}^n$ 的维数是 n，$M_{m\times n}(F)$ 的维数是 mn，$C^k(\mathbb{R})$ 和 $\ell^2(\mathbb{R})$ 的维数是无穷。

§A.2 线性映射

参考文献

[Bre57] Leo Breiman. The Individual Ergodic Theorem of Information Theory. *The Annals of Mathematical Statistics*, 28(3):809–811, 1957.

[CT12] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.

[Huf52] David A. Huffman. A Method for the Construction of Minimum-Redundancy Codes. *Proceedings of the IRE*, 40(9):1098–1101, September 1952.

[Inf] Information | Etymology, origin and meaning of information by etymonline. https://www.etymonline.com/word/information.

[Jay02] Edwin T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2002.

[KL51] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[LLG+19] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension, October 2019.

[McM53] Brockway McMillan. The Basic Theorems of Information Theory. *The Annals of Mathematical Statistics*, 24(2):196–219, June 1953.

[RHW86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*, pages 318–362. MIT Press, Cambridge, MA, USA, January 1986.

[Rob49] Robert M. Fano. *The Transmission of Information*. March 1949.

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948.

[Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer, New York, NY, 1996.

[Tin62] Hu Kuo Ting. On the Amount of Information. *Theory of Probability & Its Applications*, 7(4):439–447, January 1962.

[Uff22] Jos Uffink. Boltzmann's Work in Statistical Physics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2022 edition, 2022.

[李 10] 李贤平. 概率论基础. 高等教育出版社, 2010.

索引

乘积空间, 164

向量, 163
向量空间, 163
域, 162
基, 165

生成集, 164
线性子空间, 164
线性相关, 164
线性空间, 163
线性组合, 164
维数, 165

表示论, 164