ITエンジニアのための
機械学習理論入門

机器学习入门之道

[日]中井悦司◎著　姚待艳◎译

中国工信出版集团　人民邮电出版社
POSTS & TELECOM PRESS

IT エンジニアのための
機械学習理論入門

机器学习入门之道

[日]中井悦司◎著　姚待艳◎译

人民邮电出版社
北　京

图书在版编目（CIP）数据

机器学习入门之道／（日）中井悦司著；姚待艳译
. 北京：人民邮电出版社，2018.5(2018.9重印)
ISBN 978-7-115-47934-1

Ⅰ. ①机… Ⅱ. ①中… ②姚… Ⅲ. ①机器学习
Ⅳ. ① TP181

中国版本图书馆 CIP 数据核字（2018）第 032931 号

内 容 提 要

人工智能正在形成一股新的浪潮，它将从技术、经济、社会等各个层面改变我们的工作和生活方式。作为实现人工智能的重要技术，机器学习正在受到人工智能专家之外的更广泛人群的关注，想要了解机器学习相关知识和技术的人日益增多。

本书紧紧围绕“机器学习的商业应用”这个主题，从数学原理上解释了机器学习的一些基础算法，如最小二乘法、最优推断法、感知器、Logistic 回归、*K* 均值算法、EM 算法、贝叶斯推断等。全书的主旨在于帮助读者理解机器学习的本质，因此作者介绍具体的例题时，基本的着眼点是教会读者使用什么样的思维方式，以及如何进行计算，为读者探索更加复杂的深度学习领域或神经网络算法打下坚实的基础。

本书适合所有对机器学习感兴趣的读者阅读，尤其适合具备一定数学基础的 IT 工程师阅读，也可作为高等院校相关专业师生的参考读物。

◆ 著 ［日］中井悦司
译 姚待艳
责任编辑 陈 宏
责任印制 焦志炜
◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 http://www.ptpress.com.cn
固安县铭成印刷有限公司印刷
◆ 开本：700×1000 1/16
印张：14 2018 年 5 月第 1 版
字数：180 千字 2018 年 9 月河北第 2 次印刷
著作权合同登记号 图字：01-2016-5087

定价：59.00 元

读者服务热线：（010）81055656 印装质量热线：（010）81055316
反盗版热线：（010）81055315
广告经营许可证：京东工商广登字 20170147 号

自　序

差不多一年前，我就产生了这样一个怀疑——从事机器学习相关领域工作的IT工程师可能会以超过预想的速度不断增加。从“数据科学”到“深度学习”，甚至是“人工智能”，全都充斥着流行媒体，即使对不是从事数据分析的普通IT工程师来说，期待已久的机器学习应用时代也已经到来了。当今甚至还有机构宣称自己可以提供“无需专业知识也能使用”的机器学习服务。

但是，这里面有很大的陷阱。市面上各种关于机器学习的工具和程序库都是开源的，好像谁都可以掌握机器学习，输入数据、运行程序就可以得出一定的结果。可是，这样的结果到底有什么“意义”呢？既然是把机器学习的结果应用到商业领域，就必须理解其中的算法并正确掌握结果所传达的含义。

本书始终围绕“机器学习的商业应用”这个主题，从原理上解释了机器学习的基础算法。对于具体的例题，本书从“使用什么样的思维方式以及如何进行计算”的视角展开详细说明。因此，本书旨在帮助读者理解机器学习以及数据科学的本质。机器学习涵盖了各种各样的算法，本质上算法之间具有一个共通的思维方式，即“数据建模和参数调优”。本书把这种思维方式作为重点，尽量以通俗易懂的语言对各种数学公式进行说明。如果读者能够理解这种观点，那么面对本书没有涵盖的、更加复杂的深度学习或神经网络算法时，至少不会有畏难情绪。

“正在头疼于为委托方制定机器学习的商业应用方案”“突然决定参加销售额分析应用软件的开发项目”，我从IT工程师朋友那里听到了这样的声音，这引发了本文开头的那个怀疑。当今时代，如果能理解并熟练运用机器学习技术，IT工程师们就能获得开启新的职业道路的机会。

最重要的是，机器学习中蕴含的趣味性可以充分满足 IT 工程师对知识的好奇心和对技术的探究之心。希望本书的读者都能以本书为起点，成功地踏出迈向机器学习世界的第一步。

2015 年初秋

中井悦司

写给读者的话

本书面向的主要读者是希望理解机器学习算法背后的原理并将机器学习灵活运用到商业领域中的 IT 工程师。机器学习有各种用途，本书将从数据分析的视角出发，对各种算法进行解释说明。但请各位读者注意，本书并非介绍机器学习工具和程序库使用方法的书籍。

本书采用的大部分例题都可以说是机器学习领域的经典例子，引用自下面的书籍：

C. M. Bishop. 模式识别和机器学习（上下册）. 东京：丸善出版社，2012.

虽然打开这本书的各位已经决定通过这本书学习机器学习，但很有可能一些读者并不能深入理解其中的原理。本书的读者当然并不局限于 IT 工程师，这本书也可以作为读者突破“权威经典”的入门书籍。

本书的阅读方法

本书将对机器学习的各种算法进行系统的介绍，请从第 1 章开始按顺序阅读。在第 1 章中，为了明确“机器学习的商业应用”这个主题，我会把机器学习放在体系更庞大的数据科学中进行介绍。在之后的第 2 章至第 8 章中，我将具体的算法应用到了第 1 章介绍过的具有代表性的例题中。我会对同一个问题应用不同的算法，这能让读者理解各种算法的特征并掌握通用的思考方法。

此外，本书还提供了用 Python 编写的各个算法的可运行示例代码。运行示例代码并观察具体的输出结果，就可以捕捉到数学公式无法呈现出来的算法的本质。

当然，要想理解机器学习的算法，就必须具备一定程度的数学知识。

本书尽量以“该数学公式用于什么计算”这种通俗易懂的方式进行解释，如果读者具备大学初级程度的数学知识，就能顺畅地理解书中内容。对机器学习相关的数学知识感兴趣的读者，可阅读微积分、线性代数和概率统计方面的书籍。

最后，针对“已经有将近 10 年没有碰过数学了”的读者，本书使用的主要数学符号和基本公式都归纳总结在了接下来的几页。需要的话，读者可以参考这部分内容。

致　谢

在此向在本书的撰写和出版过程中给我提供过帮助的各位表示衷心的感谢。

本书的策划萌芽于技术评论出版社池本公平先生的提案。感谢各位对我的机器学习相关知识体系进行整理，总结并出版了面向 IT 工程师的书籍，这是一个非常难得的机会。

我还要感谢本书的审校织学先生，他不仅快速高效地提出了修改意见，还提供了 Mac OS X/Windows 版本的安装步骤。

书中很多内容都受到了日本国立信息学研究所旗下项目“TopSE”的志愿者举办的学习讨论会的启发。在此感谢参与学习讨论会的各方人士。

编写本书时，很多时候我都是先把今年刚上小学的女儿送到车站，再到早间营业的星巴克进行写作。感谢照料我起居的妻子，她始终信奉“早睡早起身体好”，提倡健康的生活方式。我要对她说：“不胜感激！”

主要数学符号和基本公式

■ 求和符号

符号 Σ 表示求和。下式为x_1到x_N的求和运算：

$$\sum_{n=1}^{N} x_n = x_1 + x_2 + \cdots + x_N \tag{1}$$

■ 乘积符号

符号 Π 表示乘积。下式为x_1到x_N的乘法运算：

$$\prod_{n=1}^{N} x_n = x_1 \times x_2 \times \cdots \times x_N \tag{2}$$

■ 指数函数

符号 exp 表示以自然常数 $e \approx 2.718$ 为底的指数函数。下式表示 e 的 x 次方函数：

$$\exp x = e^x \tag{3}$$

指数函数的积可变换为自变量的和。

$$\prod_{n=1}^{N} e^{x_n} = e^{x_1} \times \cdots \times e^{x_N} = e^{x_1 + \cdots + x_N} = \exp\left\{\sum_{n=1}^{N} x_n\right\} \tag{4}$$

指数函数 e^x 的微分形式函数不变。

$$\frac{d}{dx} e^x = e^x \tag{5}$$

■ 对数函数

符号 ln 表示以自然常数 $e \approx 2.718$ 为底的对数函数。

$$\ln x = \log_e x \tag{6}$$

代入 $x = e$，则其值变为 1。

$$\ln e = 1 \tag{7}$$

对数函数满足下面的对数法则：

$$\ln \frac{ab}{c} = \ln a + \ln b - \ln c \tag{8}$$

$$\ln a^b = b \ln a \tag{9}$$

因此，将公式（4）的指数函数代入对数函数中，可简化为下面的公式：

$$\ln\left(\exp \sum_{n=1}^{N} x_n\right) = \sum_{n=1}^{N} x_n \times \ln e = \sum_{n=1}^{N} x_n \tag{10}$$

该公式表示对数函数 $\ln x$ 为指数函数 e^x 的反函数。

对数函数的微分形式如下：

$$\frac{d}{dx} \ln x = \frac{1}{x} \tag{11}$$

■ 偏微分

对于多变量函数，特定变量的微分称为偏微分（符号为 ∂）。

$\frac{\partial f(x, y)}{\partial x}$ ：y 不变，对 x 微分

$\frac{\partial f(x,y)}{\partial y}$ ：x 不变，对 y 微分

复合函数的微分公式对偏微分也成立。

$$\frac{\partial f(g(x,y))}{\partial x}=f'(g(x,y))\times\frac{\partial g(x,y)}{\partial x} \tag{12}$$

$f'(x)$表示一次微分系数。

$$f'(x)=\frac{\mathrm{d}f(x)}{\mathrm{d}x} \tag{13}$$

■ 向量的内积和外积

公式中的粗体变量表示向量以及行列式。纵向排列的向量称为基础“列向量”。

$$\mathbf{x}=\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} \tag{14}$$

根据书写方法的不同，使用行向量进行描述时，使用转置符号 T 表示列向量。

$$\mathbf{x}=(x_1,x_2,x_3)^{\mathrm{T}} \tag{15}$$

反之，对列向量进行转换可得到行向量。

$$\mathbf{x}^{\mathrm{T}}=(x_1,x_2,x_3) \tag{16}$$

内积表示“行向量 × 列向量”。

$$\mathbf{w}^{\mathrm{T}}\mathbf{x}=(w_1,w_2,w_3)\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\sum_{i=1}^{3}w_i x_i \tag{17}$$

外积表示“列向量 × 行向量”。

$$\mathbf{w}\mathbf{x}^{\mathrm{T}}=\begin{pmatrix}w_1\\w_2\\w_3\end{pmatrix}(x_1,\ x_2,\ x_3)=\begin{pmatrix}w_1x_1 & w_1x_2 & w_1x_3\\w_2x_1 & w_2x_2 & w_2x_3\\w_3x_1 & w_3x_2 & w_3x_3\end{pmatrix} \tag{18}$$

使用式（12），对代入向量内积的函数，可以求出特定元素的偏微分。

$$\frac{f(\mathbf{w}^{\mathrm{T}}\mathbf{x})}{\partial w_i}=f'(\mathbf{w}^{\mathrm{T}}\mathbf{x})\frac{\partial(\mathbf{w}^{\mathrm{T}}\mathbf{x})}{\partial w_i}=f'(\mathbf{w}^{\mathrm{T}}\mathbf{x})x_i \tag{19}$$

向量的大小用下面的符号表示：

$$\|\mathbf{x}\|=\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{x}}=\sqrt{x_1^2+x_2^2+x_3^2} \tag{20}$$

■ 随机变量的期望与方差

以一定概率取得各种不同的值的变量 X 称为随机变量，用 $P(x)$ 表示 $X=x$ 时的概率。随机变量的期望 E 和方差 V 由如下公式定义：

$$E[X]=\sum_x xP(x) \tag{21}$$

$$V[X]=E[\{X-E(X)\}^2] \tag{22}$$

式（21）中的和 $\sum_x$ 为所有满足条件的 x 的和。

期望和方差之间有如下公式成立：

$$E[aX+b]=aE[X]+b \tag{23}$$

$$V[aX]=a^2V[X] \tag{24}$$

$$V[X]=E[X^2]-(E[X])^2 \tag{25}$$

$\overline{x}=E[X]$，由式（23）可知下式成立：

$$E[X-\overline{x}]=E[X]-\overline{x}=0 \tag{26}$$

假设有两个“独立的”随机变量 X 和 Y，$X=x$ 且 $Y=y$ 时的概率（同时发生的概率）$P(x,y)$ 可表示为 $X=x$ 时的概率与 $Y=y$ 时的概率的乘积。

$$P(x,y)=P_X(x)\,P_Y(y) \tag{27}$$

例如，掷两个骰子同时掷出 1 点的概率，为各个骰子分别掷出 1 点时的概率的乘积的 1/6。这表明不同的骰子掷出某个点的概率是相互独立的。

随机变量 X 和 Y 独立时，$\overline{x}=E[X]$，$\overline{y}=E[Y]$，则下式成立[①]：

$$\begin{aligned}E[(X-\overline{x})(Y-\overline{y})]&=\sum_{x,y}(x-\overline{x})(y-\overline{y})P(x,\ y)\\&=\sum_{x}(x-\overline{x})P_X(x)\sum_{y}(y-\overline{y})P_Y(y)\\&=E[X-\overline{x}]E[Y-\overline{y}]=0\end{aligned} \tag{28}$$

① 这种关系会在 3.3.1 节中用到。