深度学习

2017 年 9 月 4 日

目录

中文版致谢

首先，我们要感谢原作者在本书翻译时给予我们的大力帮助。特别是，原作者和我们分享了书中的原图和参考文献库，这极大节省了我们的时间和精力。

本书涉及的内容博大且思想深刻，如果没有众多同学和网友的帮助，我们不可能顺利完成翻译。

我们才疏学浅而受此重任，深知自身水平难以将本书翻译得很准确。因此我们完成草稿后，将书稿公开于 Github，及早接受网友的批评和建议。以下网友为本书的翻译草稿提供了很多及时的反馈和宝贵的修改意见：@tttwwy @tankeco @fairmiracle @GageGao @huangpingchun @MaHongP @acgtyrant @yanhuibin315 @Buttonwood @titicacafz @weijy026a @RuiZhang1993 @zymiboxpay @xingkongliang @oisc @tielei @yuduowu @Qingmu @HC-2016 @xiaomingabc @bengordai @Bojian @JoyFYan @minoriwww @khty2000 @gump88 @zdx3578 @PassStory @imwebson @wlbksy @roachsinai @Elvinczp @endymecy @9578577 @linzhp @cnscottzheng @germany-zhu @zhangyafeikimi @showgood163 @kangqf @NeutronT @badpoem @kkpoker @Seaball @wheaio @angrymidiao @ZhiweiYang @corenel @zhaoyu611 @SiriusXDJ @dfcv24 @EmisXXY @FlyingFire @vsooda @friskit-china @poerin @ninesunqian @JiaqiYao @Sofring @wenlei @wizyoung @imageslr @indam @XuLYC @zhouqing-ping @freedomRen @runPenguin @piantou

在此期间，我们四位译者再次进行了校对并且相互之间也校对了一遍。然而仅仅通过我们的校对，实在难以发现翻译中存在的问题。因此，我们邀请一些同学和网友帮助我们校对。经过他们的校对，本书的翻译质量得到了极大的提升。在此我们一一列出，以表示我们由衷的感谢！

- 第一章（引言）：刘畅、许丁杰、潘雨粟和 NeutronT 对本章进行了阅读，并对

很多语句提出了不少修改建议。林中鹏进行了校对，他提出了很多独到的修改建议。

- 第二章（线性代数）：许丁杰和骆徐圣阅读本章，并修改语句。李若愚进行了校对，提出了很多细心的建议。
- 第三章（概率与信息论）：许丁杰阅读本章，并修改语句。李培炎和何翊卓进行了校对，并修改了很多中文用词，使翻译更加准确。
- 第四章（数值计算）：张亚霏阅读本章，并对其他章节也有提出了一些修改建议。张源源进行了校对，并指出了原文可能存在的问题，非常仔细。
- 第五章（机器学习基础）：郭浩和黄平春阅读本章，并修改语句。李东和林中鹏进行了校对。本章篇幅较长，能够有现在的翻译质量离不开这四位的贡献。
- 第六章（深度前馈网络）：周卫林、林中鹏和张远航阅读本章，并提出修改意见。
- 第七章（深度学习中的正则化）：周柏村进行了非常细心的校对，指出了大量问题，令翻译更加准确。
- 第八章（深度模型中的优化）：房晓宇和吴翔阅读本章。黄平春进行了校对，他提出的很多建议让行文更加流畅易懂。
- 第九章（卷积网络）：赵雨和潘雨粟阅读本章，并润色语句。丁志铭进行了非常仔细的校对，并指出很多翻译问题。
- 第十章（序列建模：循环和递归网络）：刘畅阅读本章。赵雨提供了详细的校对建议，尹瑞清根据他的翻译版本，给我们的版本提出了很多建议。虽然仍存在一些分歧，但我们两个版本的整合，让翻译质量提升很多。
- 第十二章（应用）：潘雨粟进行了校对，在他的校对之前，本章阅读起来比较困难。他提供的修改建议，不仅提高了行文流畅度，还提升了译文的准确度。
- 第十三章（线性因子模型）：贺天行阅读本章，修改语句。杨志伟校对本章，润色大量语句。
- 第十四章（自编码器）：李雨慧和黄平春进行了校对。李雨慧提升了语言的流畅度，黄平春纠正了不少错误，提高了准确性。

- 第十五章（表示学习）：cnscottzheng 阅读本章，并修改语句。
- 第十七章（蒙特卡罗方法）：张远航提供了非常细致的校对，后续还校对了一遍，使译文质量大大提升。
- 第十八章（直面配分函数）：吴家楠进行了校对，提升了译文准确性和可读性。
- 第十九章（近似推断）：黄浩军、张远航和张源源进行了校对。这章虽篇幅不大，但内容有深度，译文在三位的帮助下提高了准确度。

所有校对的修改建议都保存在 Github 上，再次感谢以上同学和网友的付出。经过这五个多月的修改，草稿慢慢变成了初稿。尽管还有很多问题，但大部分内容是可读的，并且是准确的。当然目前的翻译仍存在一些没有及时发现的问题，因此翻译也将持续更新，不断修改。我们非常希望读者能到 Github 提建议，并且非常欢迎，无论多么小的修改建议，都是非常宝贵的。

此外，我们还要感谢魏太云学长，他帮助我们与出版社沟通交流，并给予了我们很多排版上的指导。

最后，感谢我们的导师张志华教授，没有老师的支持，我们难以完成翻译。

原书致谢

如果没有他人的贡献，这本书将不可能完成。我们感谢为本书提出建议和帮助组织内容结构的人：Guillaume Alain, Kyunghyun Cho, Çağlar Gülçehre, David Krueger, Hugo Larochelle, Razvan Pascanu and Thomas Rohée 。

我们感谢为本书内容提供反馈的人。其中一些人对许多章都给出了建议：Martín Abadi, Guillaume Alain, Ion Androutsopoulos, Fred Bertsch, Olexa Bilaniuk, Ufuk Can Biçici, Matko Bošnjak, John Boersma, Greg Brockman, Alexandre de Brébisson, Pierre Luc Carrier, Sarath Chandar, Pawel Chilinski, Mark Daoust, Oleg Dashevskii, Laurent Dinh, Stephan Dreseitl, Jim Fan, Miao Fan, Meire Fortunato, Frédéric Francis, Nando de Freitas, Çağlar Gülçehre, Jurgen Van Gael, Javier Alonso García, Jonathan Hunt, Gopi Jeyaram, Chingiz Kabytayev, Lukasz Kaiser, Varun Kanade, Asifullah Khan, Akiel Khan, John King, Diederik P. Kingma, Yann LeCun, Rudolf Mathey, Matías Mattamala, Abhinav Maurya, Kevin Murphy, Oleg Mürk, Roman Novak, Augustus Q. Odena, Simon Pavlik, Karl Pichotta, Eddie Pierce, Kari Pulli, Roussel Rahman, Tapani Raiko, Anurag Ranjan, Johannes Roith, Mihaela Rosca, Halis Sak, César Salgado, Grigory Sapunov, Yoshinori Sasaki, Mike Schuster, Julian Serban, Nir Shabat, Ken Shirriff, Andre Simpelo, Scott Stanley, David Sussillo, Ilya Sutskever, Carles Gelada Sáez, Graham Taylor, Valentin Tolmer, Massimiliano Tomassoli, An Tran, Shubhendu Trivedi, Alexey Umnov, Vincent Vanhoucke, Marco Visentini-Scarzanella, Martin Vita, David Warde-Farley, Dustin Webb, Kelvin Xu, Wei Xue, Ke Yang, Li Yao, Zygmunt Zając and Ozan Çağlayan.

我们也要感谢对单个章节提供有效反馈的人：

- 数学符号：Zhang Yuanhang.

- 第一章（引言）：Yusuf Akgul, Sebastien Bratieres, Samira Ebrahimi, Charlie Gorichanaz, Brendan Loudermilk, Eric Morris, Cosmin Pârvulescu and Alfredo Solano.
- 第二章（线性代数）：Amjad Almahairi, Nikola Banić, Kevin Bennett, Philippe Castonguay, Oscar Chang, Eric Fosler-Lussier, Andrey Khalyavin, Sergey Oreshkov, István Petrás, Dennis Prangle, Thomas Rohée, Gitanjali Gulve Sehgal, Colby Toland, Alessandro Vitale and Bob Welland.
- 第三章（概率与信息论）：John Philip Anderson, Kai Arulkumaran, Vincent Dumoulin, Rui Fa, Stephan Gouws, Artem Oboturov, Antti Rasmus, Alexey Surkov and Volker Tresp.
- 第四章（数值计算）：Tran Lam AnIan Fischer and Hu Yuhuang.
- 第五章（机器学习基础）：Dzmitry Bahdanau, Justin Domingue, Nikhil Garg, Makoto Otsuka, Bob Pepin, Philip Popien, Emmanuel Rayner, Peter Shepard, Kee-Bong Song, Zheng Sun and Andy Wu.
- 第六章（深度前馈网络）：Uriel Berdugo, Fabrizio Bottarel, Elizabeth Burl, Ishan Durugkar, Jeff Hlywa, Jong Wook Kim, David Krueger and Aditya Kumar Praharaj.
- 第七章（深度学习中的正则化）：Morten Kolbæk, Kshitij Lauria, Inkyu Lee, Sunil Mohan, Hai Phong Phan and Joshua Salisbury.
- 第八章（深度模型中的优化）：Marcel Ackermann, Peter Armitage, Rowel Atienza, Andrew Brock, Tegan Maharaj, James Martens, Kashif Rasul, Klaus Strobl and Nicholas Turner.
- 第九章（卷积网络）：Martín Arjovsky, Eugene Brevdo, Konstantin Divilov, Eric Jensen, Mehdi Mirza, Alex Paino, Marjorie Sayer, Ryan Stout and Wentao Wu.
- 第十章（序列建模：循环和递归网络）：Gökçen Eraslan, Steven Hickson, Razvan Pascanu, Lorenzo von Ritter, Rui Rodrigues, Dmitriy Serdyuk, Dongyu Shi and Kaiyu Yang.
- 第十一章（实践方法论）：Daniel Beckstein.

- 第十二章（应用）：George Dahl, Vladimir Nekrasov and Ribana Roscher.
- 第十三章（线性因子模型）：Jayanth Koushik.
- 第十五章（表示学习）：Kunal Ghosh.
- 第十六章（深度学习中的结构化概率模型）：Minh Lê and Anton Varfolom.
- 第十八章（直面配分函数）：Sam Bowman.
- 第十九章（近似推断）：Yujia Bao.
- 第二十章（深度生成模型）：Nicolas Chapados, Daniel Galvez, Wenming Ma, Fady Medhat, Shakir Mohamed and Grégoire Montavon.
- 参考文献：Lukas Michelbacher and Leslie N. Smith.

我们还要感谢那些允许我们从他们的出版物中复制图片、数据的人。我们在图片标题的文字中注明了他们的贡献。

我们还要感谢 Lu Wang 为我们写了 pdf2htmlEX，我们用它来制作这本书的网页版本，Lu Wang 还帮助我们改进了生成的 HTML 的质量。

我们还要感谢 Ian 的妻子 Daniela Flori Goodfellow，在 Ian 的写作过程中的耐心支持和检查。

我们还要感谢 Google Brain 团队提供了学术环境，从而使得 Ian 能够花费大量时间写作此书并接受同行的反馈和指导。我们特别感谢 Ian 的前任经理 Greg Corrado 和他的现任经理 Samy Bengio 对这个项目的支持。最后我们还要感谢 Geoffrey Hinton 在写作困难时的鼓励。

网站

www.deeplearningbook.org

这本书伴随有上述网站。网站提供了各种补充材料，包括练习、讲义幻灯片、错误更正以及其他应该对读者和讲师有用的资源。

数学符号

本节简要介绍本书所使用的数学符号。我们在第二章至第四章中描述大多数数学概念，如果你不熟悉任何相应的数学概念，可以参考对应的章节。

数和数组

a	标量 (整数或实数)
$\boldsymbol{a}$	向量
$\boldsymbol{A}$	矩阵
A	张量
$\boldsymbol{I}_n$	n 行 n 列的单位矩阵
$\boldsymbol{I}$	维度蕴含于上下文的单位矩阵
$\boldsymbol{e}^{(i)}$	标准基向量 $[0,\dots,0,1,0,\dots,0]$，其中索引 i 处值为 1
$\text{diag}(\boldsymbol{a})$	对角方阵，其中对角元素由 $\boldsymbol{a}$ 给定
a	标量随机变量
$\mathbf{a}$	向量随机变量
$\mathbf{A}$	矩阵随机变量

集合和图

$\mathbb{A}$	集合
$\mathbb{R}$	实数集
$\{0, 1\}$	包含 0 和 1 的集合
$\{0, 1, \dots, n\}$	包含 0 和 n 之间所有整数的集合
$[a, b]$	包含 a 和 b 的实数区间
$(a, b]$	不包含 a 但包含 b 的实数区间
$\mathbb{A} \backslash \mathbb{B}$	差集，即其元素包含于 $\mathbb{A}$ 但不包含于 $\mathbb{B}$
$\mathcal{G}$	图
$Pa_{\mathcal{G}}(\mathrm{x}_i)$	图 $\mathcal{G}$ 中 x_i 的父节点

索引

a_i	向量 $\boldsymbol{a}$ 的第 i 个元素，其中索引从 1 开始
a_{-i}	除了第 i 个元素，$\boldsymbol{a}$ 的所有元素
$A_{i,j}$	矩阵 $\boldsymbol{A}$ 的 i, j 元素
$\boldsymbol{A}_{i,:}$	矩阵 $\boldsymbol{A}$ 的第 i 行
$\boldsymbol{A}_{:,i}$	矩阵 $\boldsymbol{A}$ 的第 i 列
$\mathsf{A}_{i,j,k}$	3 维张量 A 的 (i, j, k) 元素
$\mathsf{A}_{:,:,i}$	3 维张量的 2 维切片
a_i	随机向量 $\mathbf{a}$ 的第 i 个元素

线性代数中的操作

$\boldsymbol{A}^\top$	矩阵 $\boldsymbol{A}$ 的转置
$\boldsymbol{A}^+$	$\boldsymbol{A}$ 的Moore-Penrose 伪逆
$\boldsymbol{A} \odot \boldsymbol{B}$	$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的逐元素乘积（Hadamard 乘积）
$\det(\boldsymbol{A})$	$\boldsymbol{A}$ 的行列式

微积分

$\frac{dy}{dx}$	y 关于 x 的导数
$\frac{\partial y}{\partial x}$	y 关于 x 的偏导
$\nabla_{\boldsymbol{x}} y$	y 关于 $\boldsymbol{x}$ 的梯度
$\nabla_{\boldsymbol{X}} y$	y 关于 $\boldsymbol{X}$ 的矩阵导数
$\nabla_{\mathbf{X}} y$	y 关于 $\mathbf{X}$ 求导后的张量
$\frac{\partial f}{\partial \boldsymbol{x}}$	$f : \mathbb{R}^n \to \mathbb{R}^m$ 的Jacobian矩阵 $\boldsymbol{J} \in \mathbb{R}^{m \times n}$
$\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x})$ or $\boldsymbol{H}(f)(\boldsymbol{x})$	f 在点 $\boldsymbol{x}$ 处的Hessian矩阵
$\int f(\boldsymbol{x}) d\boldsymbol{x}$	$\boldsymbol{x}$ 整个域上的定积分
$\int_{\mathbb{S}} f(\boldsymbol{x}) d\boldsymbol{x}$	集合 $\mathbb{S}$ 上关于 $\boldsymbol{x}$ 的定积分

概率和信息论

$\mathrm{a} \perp \mathrm{b}$	a 和 b 相互独立的随机变量
$\mathrm{a} \perp \mathrm{b} \mid \mathrm{c}$	给定 c 后条件独立
$P(\mathrm{a})$	离散变量上的概率分布
$p(\mathrm{a})$	连续变量（或变量类型未指定时）上的概率分布
$\mathrm{a} \sim P$	具有分布 P 的随机变量 a
$\mathbb{E}_{\mathrm{x} \sim P}[f(x)]$ or $\mathbb{E} f(x)$	$f(x)$ 关于 $P(\mathrm{x})$ 的期望
$\mathrm{Var}(f(x))$	$f(x)$ 在分布 $P(\mathrm{x})$ 下的方差
$\mathrm{Cov}(f(x), g(x))$	$f(x)$ 和 $g(x)$ 在分布 $P(\mathrm{x})$ 下的协方差
$H(\mathrm{x})$	随机变量 x 的香农熵
$D_{\mathrm{KL}}(P \Vert Q)$	P 和 Q 的KL 散度
$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	均值为 $\boldsymbol{\mu}$ 协方差为 $\boldsymbol{\Sigma}$，$\boldsymbol{x}$ 上的高斯分布

函数

$f : \mathbb{A} \to \mathbb{B}$	定义域为 $\mathbb{A}$ 值域为 $\mathbb{B}$ 的函数 f
$f \circ g$	f 和 g 的组合
$f(\boldsymbol{x}; \boldsymbol{\theta})$	由 $\boldsymbol{\theta}$ 参数化，关于 $\boldsymbol{x}$ 的函数（有时为简化表示，我们忽略 $\boldsymbol{\theta}$ 记为 $f(\boldsymbol{x})$ ）
$\log x$	x 的自然对数
$\sigma(x)$	Logistic sigmoid, $\dfrac{1}{1+\exp(-x)}$
$\zeta(x)$	Softplus, $\log(1+\exp(x))$
$\|\|\boldsymbol{x}\|\|_p$	$\boldsymbol{x}$ 的 L^p 范数
$\|\|\boldsymbol{x}\|\|$	$\boldsymbol{x}$ 的 L^2 范数
x^+	x 的正数部分, 即 $\max(0, x)$
$\mathbf{1}_{\text{condition}}$	如果条件为真则为 1，否则为 0

有时候我们使用函数 f，它的参数是一个标量，但应用到一个向量、矩阵或张量：$f(\boldsymbol{x})$, $f(\boldsymbol{X})$, or $f(\mathbf{X})$ 。这表示逐元素地将 f 应用于数组。例如，$\mathbf{C} = \sigma(\mathbf{X})$，则对于所有合法的 i、j 和 k，$C_{i,j,k} = \sigma(X_{i,j,k})$。

数据集和分布

p_{data}	数据生成分布
$\hat{p}_{\text{train}}$	由训练集定义的经验分布
$\mathbb{X}$	训练样本的集合
$\boldsymbol{x}^{(i)}$	数据集的第 i 个样本（输入）
$y^{(i)}$ or $\boldsymbol{y}^{(i)}$	监督学习中与 $\boldsymbol{x}^{(i)}$ 关联的目标
$\boldsymbol{X}$	$m \times n$ 的矩阵，其中行 $\boldsymbol{X}_{i,:}$ 为输入样本 $\boldsymbol{x}^{(i)}$