Example Notation for Deep Learning

伊恩
Yoshua Bengio
Aaron Courville

目录

Notation

下面列出了本文所用的数学符号对照表，具体可参照Goodfellow *et al.* (2016)的2-4章节。

数字和数组

a	标量(整数或实数)
$\boldsymbol{a}$	向量(矢量)
$\boldsymbol{A}$	矩阵
A	张量
$\boldsymbol{I}_n$	n 行 n 列的单位矩阵
$\boldsymbol{I}$	单位矩阵，维度参见上下文
$\boldsymbol{e}^{(i)}$	标准基向量 $[0,\dots,0,1,0,\dots,0]$ ，其中第 i 位为 1
$\text{diag}(\boldsymbol{a})$	正方形对角矩阵，其中对角线元素为 $\boldsymbol{a}$
a	随机变量，标量
$\mathbf{a}$	随机变量，向量
$\mathbf{A}$	随机变量，矩阵

集合和图

$\mathbb{A}$	集合
$\mathbb{R}$	实数集合
$\{0,1\}$	包含 0 和 1 的集合
$\{0,1,\ldots,n\}$	从 0 到 n 的所有整数的集合
$[a,b]$	a 到 b 的实数闭区间
$(a,b]$	a 到 b 的实数半开区间
$\mathbb{A}\backslash\mathbb{B}$	差集, 即集合包含了在 $\mathbb{A}$ 中但不在 $\mathbb{B}$ 中的元素
$\mathcal{G}$	图
$Pa_{\mathcal{G}}(\mathrm{x}_i)$	图$\mathcal{G}$ 中节点 x_i 的双亲

角标

a_i	向量 $\boldsymbol{a}$ 的第 i 个元素, 其中角标从 1 开始
a_{-i}	向量 $\boldsymbol{a}$ 除第i个元素以外的所有元素
$A_{i,j}$	矩阵 $\boldsymbol{A}$ 的第 i 行第 j 列的元素
$\boldsymbol{A}_{i,:}$	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-D 张量 A 的元素 (i,j,k)
$\mathsf{A}_{:,:,i}$	3-D 张量的 2-D 切片
a_i	随机向量 $\mathbf{a}$ 的第 i 个元素

线性代数

$\boldsymbol{A}^\top$	矩阵 $\boldsymbol{A}$ 的转置
$\boldsymbol{A}^+$	矩阵 $\boldsymbol{A}$ 的摩尔彭罗斯伪逆（广义逆）
$\boldsymbol{A}\odot\boldsymbol{B}$	矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的元素积（Hadamard乘积）
$\det(\boldsymbol{A})$	矩阵 $\boldsymbol{A}$ 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 的偏导数
$\nabla_{\boldsymbol{x}} y$	y 关于向量$\boldsymbol{x}$ 的梯度
$\nabla_{\boldsymbol{X}} y$	y 关于矩阵$\boldsymbol{X}$ 的导数
$\nabla_{\mathbf{X}} y$	y 关于张量$\mathbf{X}$ 的导数
$\frac{\partial f}{\partial \boldsymbol{x}}$	$f : \mathbb{R}^n \to \mathbb{R}^m$ 的雅克比矩阵 $\boldsymbol{J} \in \mathbb{R}^{m \times n}$
$\nabla^2_{\boldsymbol{x}} f(\boldsymbol{x})$ or $\boldsymbol{H}(f)(\boldsymbol{x})$	f 在输入向量 $\boldsymbol{x}$ 的海森矩阵
$\int f(\boldsymbol{x}) d\boldsymbol{x}$	在整个定义域上f关于 $\boldsymbol{x}$ 的定积分
$\int_{\mathbb{S}} f(\boldsymbol{x}) d\boldsymbol{x}$	在集合 $\mathbb{S}$ 上f关于 $\boldsymbol{x}$ 的定积分

概率论与信息论

$\mathrm{a} \perp \mathrm{b}$	随机变量 a 与 b 相互独立
$\mathrm{a} \perp \mathrm{b} \mid \mathrm{c}$	随机变量 a 与 b对于给定的 c 条件独立
$P(\mathrm{a})$	离散变量的概率分布
$p(\mathrm{a})$	连续变量的概率分布或类型不确定的变量的概率分布
$\mathrm{a} \sim P$	随机变量 a 服从 P 分布
$\mathbb{E}_{\mathrm{x} \sim P}[f(x)]$ or $\mathbb{E} f(x)$	$f(x)$ 在概率分布 $P(\mathrm{x})$ 的期望
$\mathrm{Var}(f(x))$	$f(x)$ 在概率分布 $P(\mathrm{x})$ 下的方差
$\mathrm{Cov}(f(x), g(x))$	$f(x)$ 和 $g(x)$ 在概率分布 $P(\mathrm{x})$ 下的协方差
$H(\mathrm{x})$	随机变量 x 的信息熵
$D_{\mathrm{KL}}(P \| Q)$	随机变量P 与 Q 的相对熵(KL散度)
$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	均值为 $\boldsymbol{\mu}$ 方差为 $\boldsymbol{\Sigma}$ 的 $\boldsymbol{x}$ 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
$f \circ g$	函数 f 和 g 的复合函数
$f(\boldsymbol{x}; \boldsymbol{\theta})$	参数为 $\boldsymbol{\theta}$ 的关于$\boldsymbol{x}$ 的函数. （有时我们写作 $f(\boldsymbol{x})$ 而忽略参数 $\boldsymbol{\theta}$ 来简化符号）
$\log x$	x 的自然对数
$\sigma(x)$	Logistic sigmoid 函数, $\dfrac{1}{1+\exp(-x)}$
$\zeta(x)$	Softplus 函数, $\log(1+\exp(x))$
$\|\|\boldsymbol{x}\|\|_p$	$\boldsymbol{x}$ 的 L^p 范数
$\|\|\boldsymbol{x}\|\|$	$\boldsymbol{x}$ 的 L^2 范数
x^+	x 的正值部分, 即 $\max(0, x)$
$\mathbf{1}_{\text{condition}}$	如果条件为真则值为1，条件为假则值为0

有时我们把参数为标量的函数 f 应用到矢量、矩阵或张量中：$f(\boldsymbol{x})$, $f(\boldsymbol{X})$, 或 $f(\mathsf{X})$。这代表着将 f 应用到数组元素层面，例如，如果 $\mathsf{C} = \sigma(\mathsf{X})$, 那么任意 i, j 和 k ，都有 $C_{i,j,k} = \sigma(X_{i,j,k})$

数据集和分布

p_{data}	数据生成的概率分布
$\hat{p}_{\text{data}}$	训练集生成(定义)的经验分布
$\mathbb{X}$	训练集
$\boldsymbol{x}^{(i)}$	数据集中的第 i 个（输入）实例
$y^{(i)}$ or $\boldsymbol{y}^{(i)}$	有监督学习下 $\boldsymbol{x}^{(i)}$ 对应的标记
$\boldsymbol{X}$	$m \times n$ 的矩阵，其中输入实例 $\boldsymbol{x}^{(i)}$ 在 $\boldsymbol{X}_{i,:}$ 行

Chapter 1

Commentary

This document is an example of how to use the accompanying files as well as some commentary on them. The files are `math_commands.tex` and `notation.tex`. The file `math_commands.tex` includes several useful LaTeX macros and `notation.tex` defines a notation page that could be used at the front of any publication.

We developed these files while writing Goodfellow *et al.* (2016). We release these files for anyone to use freely, in order to help establish some standard notation in the deep learning community.

1.1 二项逻辑回归模型

二项逻辑回归模型是如下的条件概率分布

$$\begin{aligned} P(Y=1|\boldsymbol{x}) =& \frac{\exp(\boldsymbol{\theta}^T\boldsymbol{x}+b)}{1+\exp(\boldsymbol{\theta}^T\boldsymbol{x}+b)} \\ P(Y=0|\boldsymbol{x}) =& \frac{1}{1+\exp(\boldsymbol{\theta}^T\boldsymbol{x}+b)} \end{aligned} \tag{1.1}$$

公式 (1.1)的含义将会在第 1.2节中说明，也可以参见 李航 (2012)。

1.2 Examples

We include this section as an example of some LaTeX commands and the macros we created for the book.

Citations that support a sentence without actually being used in the sentence should appear at the end of the sentence using `citep`:

> Inventors have long dreamed of creating machines that think. This desire dates back to at least the time of ancient Greece. The mythical figures Pygmalion, Daedalus, and Hephaestus may all be interpreted as legendary inventors, and Galatea, Talos, and Pandora may all be regarded as artificial life (Ovid and Martin, 2004; Sparkes, 1996; Tandy, 1997).

When the authors of a document or the document itself are a noun in the sentence, use the `citet` command:

> Mitchell (1997) provides a succinct definition of machine learning: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

When introducing a new term, using the `newterm` macro to highlight it. If there is a corresponding acronym, put the acronym in parentheses afterward. If your document includes an index, also use the `index` command.

> Today, **artificial intelligence** (AI) is a thriving field with many practical applications and active research topics.

Sometimes you may want to make many entries in the index that all point to a canonical index entry:

> One of the simplest and most common kinds of parameter norm penalty is the squared L^2 parameter norm penalty commonly known as **weight decay**. In other academic communities, L^2 regularization is also known as **ridge regression** or **Tikhonov regularization**.

To refer to a figure, use either `figref` or `Figref` depending on whether you want to capitalize the resulting word in the sentence.

> See 图 1.1 for an example of a how to include graphics in your document. 图 1.1 shows how to include graphics in your document.

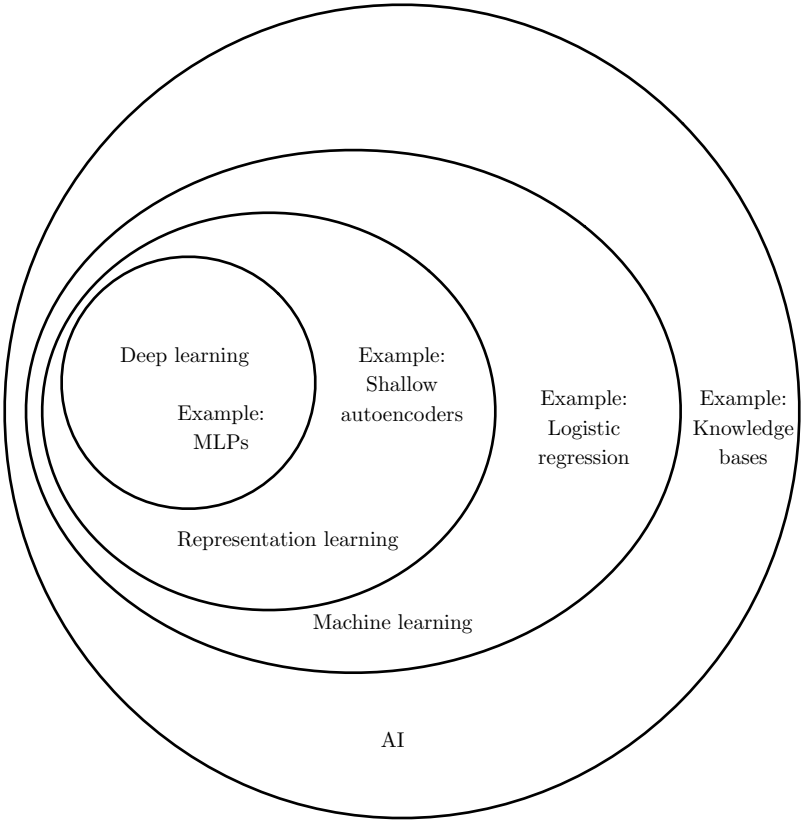

图 1.1: An example of a figure. The figure is a PDF displayed without being rescaled within LaTeX. The PDF was created at the right size to fit on the page, with the fonts at the size they should be displayed. The fonts in the figure are from the Computer Modern family so they match the fonts used by LaTeX.

Similarly, you can refer to different sections of the book using `partref`, `Partref`, `secref`, `Secref`, etc.

You are currently reading 第 1.2节.

Acknowledgments

We thank Catherine Olsson and Úlfar Erlingsson for proofreading and review of this manuscript.

参考文献

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. ii, 1

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, New York. 2

Ovid and Martin, C. (2004). *Metamorphoses*. W.W. Norton. 2

Sparkes, B. (1996). *The Red and the Black: Studies in Greek Pottery*. Routledge. 2

Tandy, D. W. (1997). *Works and Days: A Translation and Commentary for the Social Sciences*. University of California Press. 2

李航 (2012). 统计学习方法. 清华大学出版社. 1

索引