

# Notation

This section provides a concise reference describing the notation used throughout this book. If you are unfamiliar with any of these mathematical concepts, this notation reference may seem intimidating. However, do not despair, we describe most of these ideas in chapters 1-3.

## Numbers and Arrays

$a$	A scalar (integer or real) value with the name “a”
$\mathbf{a}$	A vector with the name “a”
$\mathbf{A}$	A matrix with the name “A”
$\mathbf{A}$	A tensor with the name “A”
$\mathbf{I}_n$	Identity matrix with $n$ rows and $n$ columns
$\mathbf{I}$	Identity matrix with dimensionality implied by context
$\mathbf{e}_i$	Standard basis vector $[0, \dots, 0, 1, 0, \dots, 0]$ with a 1 at position $i$ .
$\text{diag}(\mathbf{a})$	A square, diagonal matrix with entries given by $\mathbf{a}$
$a$	A scalar random variable with the name “a”
$\mathbf{a}$	A vector-valued random variable with the name “a”
$\mathbf{A}$	A matrix-valued random variable with the name “A”

## Sets and Graphs

$\mathbb{A}$	A set with the name “A”
$\mathbb{R}$	The set of real numbers
$\{0, 1\}$	The set containing 0 and 1
$\{0, 1, \dots, n\}$	The set of all integers between 0 and $n$
$[a, b]$	The real interval including $a$ and $b$
$(a, b]$	The real interval excluding $a$ but including $b$
$\mathbb{A} \setminus \mathbb{B}$	Set subtraction, i.e., the elements of $\mathbb{A}$ that are not in $\mathbb{B}$
$\mathcal{G}$	A graph with the name “G”
$Pa_{\mathcal{G}}(x_i)$	The parents of $x_i$ in $\mathcal{G}$ .

## Indexing

$a_i$	Element $i$ of vector $\mathbf{a}$ , with indexing starting at 1
$a_{-i}$	All elements of vector $a$ except for element $i$
$A_{i,j}$	Element $i, j$ of matrix $\mathbf{A}$
$\mathbf{A}_{i,:}$	Row $i$ of matrix $\mathbf{A}$
$\mathbf{A}_{:,i}$	Column $i$ of matrix $\mathbf{A}$
$\mathbf{A}_{i,j,k}$	Element $(i, j, k)$ of a 3-D tensor $\mathbf{A}$
$\mathbf{A}_{::,i}$	2-D slice of a 3-D tensor
$\mathbf{a}_i$	Element $i$ of the random vector $\mathbf{a}$

## Linear Algebra Operations

$\mathbf{A}^{\top}$	Transpose of matrix $\mathbf{A}$
$\mathbf{A}^{+}$	Moore-Penrose pseudoinverse of $\mathbf{A}$
$\mathbf{A} \odot \mathbf{B}$	Element-wise (Hadamard) product of $\mathbf{A}$ and $\mathbf{B}$

## Calculus

$\frac{dy}{dx}$	Derivative of $y$ with respect to $x$
$\frac{\partial y}{\partial x}$	Partial derivative of $y$ with respect to $x$
$\nabla_x y$	Gradient of $y$ with respect to $x$
$\nabla_{\mathbf{X}} y$	Matrix derivatives of $y$ with respect to $x$
$\frac{\partial f}{\partial \mathbf{x}}$	Jacobian matrix $\mathbf{J} \in \mathbb{R}^{n \times n}$ of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$
$\nabla_{\mathbf{x}}^2 f(\mathbf{x})$ or $\mathbf{H}(f)(\mathbf{x})$	The Hessian matrix of $f$ at input point $\mathbf{x}$
$\int f(\mathbf{x}) d\mathbf{x}$	Definite integral over the entire domain of $\mathbf{x}$
$\int_{\mathbb{S}} f(\mathbf{x}) d\mathbf{x}$	Definite integral with respect to $\mathbf{x}$ over the set $\mathbb{S}$

## Probability and Information Theory

$a \perp b$	The random variables $a$ and $b$ are independent.
$a \perp b \mid c$	They are conditionally independent given $c$ .
$\mathbb{E}_{x \sim P}[f(x)]$ or $\mathbb{E}f(x)$	Expectation of $f(x)$ with respect to $P(x)$
$\text{Var}(f(x))$	Variance of $f(x)$ under $P(x)$
$\text{Cov}(f(x), g(x))$	Covariance of $f(x)$ and $g(x)$ under $P(x, y)$
$H(x)$	Shannon entropy of the random variable $x$
$D_{\text{KL}}(P \parallel Q)$	Kullback-Leibler divergence of $P$ and $Q$

## Functions

$f \circ g$	Composition of the functions $f$ and $g$
$f(\mathbf{x}; \boldsymbol{\theta})$	A function of $\mathbf{x}$ parameterized by $\boldsymbol{\theta}$
$\log x$	Natural logarithm of $x$
$\sigma(x)$	Logistic sigmoid, $1/(1 + \exp(-x))$
$\zeta(x)$	Softplus, $\log(1 + \exp(x))$
$\ \mathbf{x}\ _p$	$L^p$ norm of $\mathbf{x}$
$x^+$	Positive part of $x$ , i.e., $\max(0, x)$

$\mathbf{1}_{condition}$  is 1 if the condition is true, 0 otherwise.

Sometimes we write  $f(\mathbf{x})$ ,  $f(\mathbf{X})$ , or  $f(\mathbf{X})$ , when  $f$  is a function of a scalar rather than a vector, matrix, or tensor. In this case, we mean to apply  $f$  to the array element-wise. For example, if  $\mathbf{C} = \sigma(\mathbf{X})$ , then  $C_{i,j,k} = \sigma(X_{i,j,k})$  for all valid values of  $i$ ,  $j$  and  $k$ .

## Datasets and distributions

$\mathbb{X}$	A set of training examples
$\mathbf{x}^{(i)}$	The $i$ -th example (input) from a dataset
$y^{(i)}$ or $\mathbf{y}^{(i)}$	The target associated with $\mathbf{x}^{(i)}$ for supervised learning
$\mathbf{X}$	The $m \times n$ matrix with input example $\mathbf{x}^{(i)}$ in row $\mathbf{X}_{i,:}$