Terminology

Grouping Terms

$(\cdot)$	parentheses are used to group scalar terms
$[\cdot]$	brackets are used to group vectors, 1-forms, and matrices

Scalars and Vectors

a, b, c, d, r	real number, possibly indexed; typically, known values
h	real number or function, *context dependent*; e.g., $h(x)$ or h_{ij}
i, j, k, q	index; positive integer
l	*context dependent* positive integer as an index or real number as a level of a function
m	size of an objective vector, i.e., vector is in $\mathbb{R}^m$
n	size of a search vector, i.e., vector is in $\mathbb{R}^n$
p	*context dependent*; probability or positive integer or scalar-valued function, possibly indexed
w_0, x_0, y_0	real number; typically, a specific value
x_i, y_i, z_i	real number, *context dependent*; entry of a vector or i^{th} iterant of a sequence of scalars
$\vec{a}$, $\vec{b}$, $\vec{c}$, $\vec{d}$, $\vec{r}$	real vector, possibly indexed; typically, known values
$\vec{e}_i$	elementary vector; j^{th} entry is 0 if $j \neq i$ and is 1 if $j = i$
f, g	scalar-valued function, possibly indexed; e.g., $f : \mathbb{R} \to \mathbb{R}$ or $f : \mathbb{R}^n \to \mathbb{R}$
$\vec{f}$, $\vec{g}$, $\vec{p}$	vector-valued function, e.g., $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$
$H(\underline{P})$	entropy of probability distribution $\underline{P}$
$\mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\mu})$	Lagrange function of vector $\vec{x}$ and multiplier(s) $\vec{\lambda}$ and $\vec{\mu}$
$\mathcal{L}_D(\vec{\lambda}, \vec{\mu})$	Lagrange dual function of multiplier(s) $\vec{\lambda}$ and $\vec{\mu}$
$\vec{q}$, $\vec{u}$, $\vec{w}$, $\vec{x}$, $\vec{y}$, $\vec{z}$	real vector, possibly indexed
$\vec{v}$	real vector, possibly indexed, *context dependent*; general vector or eigenvector of a matrix
$\vec{w}_0$, $\vec{x}_0$, $\vec{y}_0$	real vector; typically, a specific value
$\vec{w}_j$, $\vec{x}_j$, $\vec{y}_j$, $\vec{z}_j$	real vector; typically, j^{th} iterant of a sequence or j^{th} instance of data
${}_i\vec{w}$, ${}_i\vec{x}$	real vector; typically, weight vector or data vector for the i^{th} layer of an artificial neural network
${}^i\vec{w}$	real vector; for an artificial neural network with p layers, ${}^i\vec{w} \stackrel{\text{def}}{=} [{}_i\vec{w} \, \ldots \, {}_p\vec{w}]^T$

1-Forms

$\underline{a}$, $\underline{u}$, $\underline{v}$, $\underline{x}$	real 1-form in $\mathbb{R}^n$ with i^{th} entry a_i, etc.; computed as a $1 \times n$ row matrix $\underline{a} = [a_1\ a_2\ \cdots\ a_n]$
$\underline{0}$, $\underline{1}$	real 1-forms with constant entries
$\underline{z}(M)$	"mean" 1-form of a matrix, where entry z_j is the mean of column j of M
$\underline{P}$, $\underline{Q}$	Probability distributions

General Matrices

A, C, M, X, Y, Z	matrix; entries are a_{ij}, etc.
B	symmetric positive definite matrix
D, K	symmetric matrix, positive definite or positive semidefinite
R	upper-triangular matrix
Q, U	orthogonal matrix
V	orthogonal matrix; may give eigenvectors of another matrix
W	*context dependent*; general matrix or symmetric positive semidefinite matrix

Special Functions

$H(u)$	Heaviside step function; $H(v) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if} \quad v < 0 \\ 1 & \text{if} \quad v \geq 0 \end{cases}$

Greek Symbols

α, α_i, $\vec{\alpha}$	real number, *context dependent*; typically a known value or, in SVM, Lagrange multiplier
β, γ, ρ	real numbers, especially as known values
η	positive real number; typically, $0 < \eta < 1$ as a learning rate
λ, λ_i	real number, *context dependent*; eigenvalue or Lagrange multiplier or regularizer
μ, μ_i	real number, *context dependent*; free parameter or Lagrange multiplier or mean value
ν	real number; generally, variable parameter
ϕ, θ, ψ	real number
$\vec{\rho}$	real vector, possibly indexed
Λ	diagonal matrix where λ_i is an eigenvalue of another matrix
Σ	diagonal matrix, or diagonal-like matrix such as the singular-value matrix

Differentials

$f', \dfrac{df}{dt}$	derivative of $f:\mathbb{R}\to\mathbb{R}$
$f'', \dfrac{d^2f}{dt^2}$	derivative of $f':\mathbb{R}\to\mathbb{R}$
$\dfrac{\partial f}{\partial w_i}(\vec{w})$	partial derivative of $\vec{f}:\mathbb{R}^n\to\mathbb{R}$ with respect to w_i
$\underline{\nabla}\, f(\vec{w})$	1-form derivative of $f:\mathbb{R}^n\to\mathbb{R}$; i^{th} entry is $\dfrac{\partial f}{\partial w_i}$
$\dfrac{\partial^2 f}{\partial w_i \partial w_j}(\vec{w})$	partial derivative of $\dfrac{\partial f}{\partial w_j}:\mathbb{R}^n\to\mathbb{R}$ with respect to w_i
$D_{\vec{v}} f(\vec{w})$	directional derivative of $f:\mathbb{R}^n\to\mathbb{R}$ at $\vec{w}$ in direction $\vec{v}$
$J_{\vec{f}}(\vec{w}_0)$	Jacobian matrix of $\vec{f}:\mathbb{R}^n\to\mathbb{R}^m$, evaluated at $\vec{w}_0$; entry (i,j) is $\dfrac{\partial f_i}{\partial w_j}(\vec{w}_0)$ so the i^{th} row of $J_{\vec{f}}(\vec{w}_0)$ is $\underline{\nabla}\, f_i(\vec{w}_0)$
$\nabla^2 f(\vec{w}_0)$	Hessian matrix of $f:\mathbb{R}^n\to\mathbb{R}$, evaluated at $\vec{w}_0$; entry (i,j) is $\dfrac{\partial^2 f}{\partial w_i \partial w_j}(\vec{w}_0)$

Sets

$\mathbb{N}$	set of natural numbers, including zero
$\mathbb{N}_+$	set of positive natural numbers
$\mathbb{R}$	set of real numbers
$\mathbb{R}^n$	set of real vectors, each vector having n entries
$\mathbb{R}^n_+$	set of real vectors where each entry of each vector is non-negative
$\mathbb{R}^n_{++}$	set of real vectors where each entry of each vector is positive
$\mathbb{S}_C(f,l)$	level curve of function f at level l
$\mathbb{S}_L(f,l)$	level set of function f at level l
$\mathbb{V}$	vector space over the field of real numbers
$\mathbb{A}$	events a_i
$\mathbb{A}(M,\vec{c})$	Affine space; set of vectors $\vec{u}$ that are solutions to $M\vec{w}=\vec{c}$
$\mathbb{F}$	Feasible solutions of a constrained optimization problem