SYMBOLS AND NOTATION

Mathematical Relations

$a \triangleq b$	a is equal to b by definition
$a \overset{D}{=} b$	a is equal to b in distribution
$a \propto b$	a is proportional to b, i. e., $a = \text{const} \cdot b$
$a \approx b$	a is approximately equal to b, i. e., $\|a - b\| < \epsilon$ for small real number $\epsilon > 0$

Numbers, Arrays & Sets

a	A scalar (integer or real)
$\mathbf{a}$	A vector
$\mathbf{A}$	A matrix
$\mathbf{0}_n, \mathbf{0}$	A vector of zeros of length n or implied by context
$\mathbf{I}_n, \mathbf{I}$	Identity matrix with n rows and columns or dimensionality implied by context
$\operatorname{diag} \mathbf{a}$	A square, diagonal matrix with diagonal entries given by $\mathbf{a}$
$\mathbb{N}, \mathbb{Z}$	The set of natural numbers and integers, respectively
$\mathbb{R}, \mathbb{C}$	The set of real and complex numbers, respectively
$\mathbb{R}^d$	The d-dimensional vector space of real numbers

Linear Algebra

$\mathbf{A}^\top, \mathbf{a}^\top$	Transpose of a matrix or vector
$\mathbf{A}^{-1}$	Inverse of square matrix
$\det \mathbf{A}$	Determinant of square matrix
$\operatorname{tr} \mathbf{A}$	Trace of square matrix
$\mathbf{A} \succeq 0$	Matrix $\mathbf{A}$ is positive semidefinite
$\mathbf{A}^{\frac{1}{2}}$	Square root of a matrix, specifically the Cholesky decomposition: a lower-triangular matrix $\mathbf{L}$ that satisfies $\mathbf{L}\mathbf{L}^\top = \mathbf{A}$

Functions & Functional Analysis

$f : \mathcal{X} \to \mathcal{Y}$	A function with domain $\mathcal{X}$ and range $\mathcal{Y}$
$f : \mathbf{x} \mapsto g(\mathbf{x})$	A function that maps x to $g(x)$; i. e., $f(\mathbf{x}) \triangleq g(\mathbf{x})$
$f \circ g$	Composition of functions f and g; $f \circ g : \mathbf{x} \mapsto f(g(\mathbf{x}))$

$\mathcal{O}(\cdot)$	Asymptotic upper bound ("big O"); $f(n) = \mathcal{O}(g(n))$ for $f, g : \mathbb{N} \to \mathbb{N}$ if $f(n)/g(n)$ is bounded as $n \to \infty$
$\mathbb{R}^{\mathcal{X}}$	The space of functions $f : \mathcal{X} \to \mathbb{R}$
$\mathcal{H}_k, \mathcal{H}$	Reproducing kernel Hilbert space associated with kernel k or implied by context
$\langle \cdot, \cdot \rangle_{\mathcal{H}}, \langle \cdot, \cdot \rangle$	Inner product associated with Hilbert space $\mathcal{H}$ or implied by context
$\lVert \cdot \rVert, \lVert \cdot \rVert_p$	L^2 norm of a vector; L^p norm if subscript p is specified

Calculus

$\frac{\mathrm{d}y}{\mathrm{d}x}$	Total derivative of y with respect to x
$\frac{\partial y}{\partial x}$	Partial derivative of y with respect to x
$\frac{\partial f}{\partial \mathbf{x}}$	Jacobian matrix $\mathbf{J} \in \mathbb{R}^{m \times n}$ of $f : \mathbb{R}^n \to \mathbb{R}^m$
$\int f(\mathbf{x})\,\mathrm{d}\mathbf{x}$	Definite integral over the entire domain of $\mathbf{x}$
$\int_{\mathcal{X}} f(\mathbf{x})\,\mathrm{d}\mathbf{x}$	Definite integral with respect to $\mathbf{x}$ over the set $\mathcal{X}$

Probability and Information Theory

$p(\mathbf{x}), q(\mathbf{x})$	A probability density, latter used to emphasise approximation
$\mathbf{x} \sim p(\mathbf{x})$	Random variable $\mathbf{x}$ is distributed according to $p(\mathbf{x})$
$\mathbb{E}_{p(\mathbf{x})}[f(\mathbf{x})], \mathbb{E}[f(\mathbf{x})]$	Expectation of $f(\mathbf{x})$ under $p(\mathbf{x})$ or implied by context
$\mathrm{Cov}\,(\cdot, \cdot)$	Covariance between random variables
$\mathbb{H}[\cdot]$	Shannon entropy of a random variable
$\mathcal{D}_f\,[p \parallel q]$	f-divergence between distributions with densities p and q
$\mathcal{D}_{\mathrm{KL}}\,[p \parallel q]\,, \mathrm{KL}\,[p \parallel q]$	Kullback-Leibler divergence between distributions with densities p and q
$\mathcal{U}[a, b]$	Uniform distribution with lower and upper bounds a and b
$\mathrm{Bern}(\rho)$	Bernoulli distribution with parameter ρ
$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}), \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Multivariate Gaussian distribution (on $\mathbf{x}$) with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$
$\mathcal{GP}(f; m, k)$	Gaussian process; $f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ denotes $f(\mathbf{x})$ is a distributed as a Gaussian process with mean function m and covariance function (kernel) k
δ_{ij}	Kronecker delta; $\delta_{ij} = 1$ iff $i = j$ and 0 otherwise
$\delta(x - x_0)$	Dirac delta on x with point mass at x_0

Optimisation

$f^* = \min_{\mathbf{x}} f(\mathbf{x})$	A minimum of function $f(\mathbf{x})$

$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x})$	A minimiser of function $f(\mathbf{x})$

Special Functions

$\sigma(x)$	Sigmoid function, typically the logistic sigmoid $x \mapsto (1 + \exp(-x))^{-1}$
$\textsc{relu}(x)$	Rectified linear unit activation; positive part of x, i. e., $x \mapsto \max(0, x)$
$\text{Softplus}(x)$	Softplus activation; $x \mapsto \log(1 + \exp(x))$