Notation and Symbols

$\mathbb{R}$	the set of reals
$\mathbb{N}$	the set of natural numbers, $\mathbb{N} = \{1, 2, \ldots\}$
$\mathcal{X}$	the input domain
N	(used if $\mathcal{X}$ is a vector space) dimension of $\mathcal{X}$
x_i	input patterns
y_i	target values $y_i \in \mathbb{R}$, or (in pattern recognition) classes $y_i \in \{\pm 1\}$
m	number of training examples
$[m]$	compact notation for $\{1, \ldots, m\}$
i, j	indices, by default running over $[m]$
X	a sample of input patterns, $X = (x_1, \ldots, x_m)$
Y	a sample of output targets, $Y = (y_1, \ldots, y_m)$
$\mathcal{H}$	feature space
$\mathbf{\Phi}$	feature map, $\mathbf{\Phi} : \mathcal{X} \to \mathcal{H}$
$\mathbf{x}_i$	a vector with entries $[\mathbf{x}_i]_j$; usually a mapped pattern in $\mathcal{H}$, $\mathbf{x}_i = \mathbf{\Phi}(x_i)$
$\mathbf{w}$	weight vector in feature space
b	constant offset (or threshold)
k	(positive definite) kernel
K	kernel matrix or Gram matrix, $K_{ij} = k(x_i, x_j)$
$\mathbf{E}[\xi]$	expectation of a random variable ξ (Section B.1.3)
$\mathrm{P}\{\cdot\}$	probability of a logical formula
$\mathrm{P}(C)$	probability of a set (event) C
$p(x)$	density evaluated at $x \in \mathcal{X}$
$\mathcal{N}(\varepsilon, \mathcal{F}, d)$	covering number of a set $\mathcal{F}$ in the metric d with precision ε
$\mathcal{N}(\mu, \sigma)$	normal distribution with mean μ and variance σ
ε	parameter of the ε-insensitive loss function
α_i	Lagrange multiplier or expansion coefficient
β_i	Lagrange multiplier
$\boldsymbol{\alpha}, \boldsymbol{\beta}$	vectors of Lagrange multipliers
ξ_i	slack variables
$\boldsymbol{\xi}$	vector of all slack variables
Q	Hessian of a quadratic program

$\langle \mathbf{x}, \mathbf{x}' \rangle$	dot product between $\mathbf{x}$ and $\mathbf{x}'$
$\|\cdot\|$	2-norm, $\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$
$\|\cdot\|_p$	p-norm , $\|x\|_p := \left(\sum_{i=1}^N \|x_i\|^p\right)^{1/p}$, $N \in \mathbb{N} \cup \{\infty\}$
$\|\cdot\|_\infty$	∞-norm , $\|x\|_\infty := \max_{i=1}^N \|x_i\|$ on $\mathbb{R}^N$, $\|x\|_\infty := \sup_{i=1}^\infty \|x_i\|$ on ℓ_∞
ln	logarithm to base e
$\log_2$	logarithm to base 2
f	a function $\mathcal{X} \to \mathbb{R}$ or $\mathcal{X} \to \{\pm 1\}$
$\mathcal{F}$	a family of functions
$\rho_f(x, y)$	margin of function f on the example (x, y), i.e., $y \cdot f(x)$
ρ_f	margin of f on the training set, i.e., $\min_{i=1}^m \rho_f(x_i, y_i)$
h	VC dimension
C	regularization parameter in front of the empirical risk term
λ	regularization parameter in front of the regularizer
$x \in [a, b]$	interval $a \leq x \leq b$
$x \in (a, b]$	interval $a < x \leq b$
$x \in (a, b)$	interval $a < x < b$
A^{-1}	inverse matrix (in some cases, pseudo-inverse)
$A^\top$	transposed matrix (or vector)
A^*	adjoint matrix (or: operator, vector), i.e., transposed and complex conjugate
$(x_j)_j$ or (x_j)	shorthand for a sequence $(x_j) = (x_1, x_2, \ldots)$
ℓ_p	sequence spaces, $1 \leq p \leq \infty$ (Section B.3.1)
$L_p(\mathcal{X})$	function spaces, $1 \leq p \leq \infty$ (Section B.3.1)
I_A	characteristic (or indicator) function on a set A i.e., $I_A(x) = 1$ if $x \in A$ and 0 otherwise
$\mathbf{1}$	unit matrix, or identity map ($\mathbf{1}(x) = x$ for all x)
$\|C\|$	cardinality of a set C (for finite sets, the number of elements)
Υ	regularization operator
δ_{ij}	Kronecker δ (Section B.2.1)
δ_x	Dirac δ, satisfying $\int \delta_x(y) f(y) dy = f(x)$
$O(g(n))$	a function $f(n)$ is said to be $O(g(n))$ if there exists a constant C such that $\|f(n)\| \leq C g(n)$ for all n
$o(g(n))$	a function is said to be $o(g(n))$ if there exists a constant c such that $\|f(n)\| \geq c g(n)$ for all n
rhs/lhs	shorthand for "right/left hand side"
■	the end of a proof
•	easy problem
••	intermediate problem
•••	difficult problem
○○○	open problem