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 \([m]\)
- \( \mathcal{X} \) a sample of input patterns, \( \mathcal{X} = (x_1, \ldots, x_m) \)
- \( \mathcal{Y} \) a sample of output targets, \( \mathcal{Y} = (y_1, \ldots, y_m) \)
- \( \mathcal{H} \) feature space
- \( \Phi \) feature map, \( \Phi : \mathcal{X} \rightarrow \mathcal{H} \)
- \( x_i \) a vector with entries \( [x_i]_j \); usually a mapped pattern in \( \mathcal{H} \), \( x_i = \Phi(x_i) \)
- \( w \) weight vector in feature space
- \( b \) constant offset (or threshold)
- \( k \) (positive definite) kernel
- \( K \) kernel matrix or Gram matrix, \( K_{ij} = k(x_i, x_j) \)
- \( \mathbb{E}[\xi] \) expectation of a random variable \( \xi \) (Section B.1.3)
- \( P(\cdot) \) probability of a logical formula
- \( P(C) \) 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 \( \varepsilon \)
- \( \mathcal{N}(\mu, \sigma) \) normal distribution with mean \( \mu \) and variance \( \sigma \)
- \( \varepsilon \) parameter of the \( \varepsilon \)-insensitive loss function
- \( \alpha_i \) Lagrange multiplier or expansion coefficient
- \( \beta_i \) Lagrange multiplier
- \( \alpha, \beta \) vectors of Lagrange multipliers
- \( \xi_i \) slack variables
- \( \xi \) 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, \( \| \mathbf{x} \|_p := \left( \sum_{i=1}^{N} |x_i|^p \right)^{1/p} \), \( N \in \mathbb{N} \cup \{ \infty \} \)

\[ \| \cdot \|_\infty \] \( \infty \)-norm, \( \| \mathbf{x} \|_\infty := \max_{1 \leq i \leq N} |x_i| \) on \( \mathbb{R}^N \), \( \| \mathbf{x} \|_\infty := \sup_{1 \leq i \leq N} |x_i| \) on \( \ell_\infty \)

\[ \ln \] logarithm to base \( e \)

\[ \log_2 \] logarithm to base 2

\[ f \] a function \( \mathcal{X} \rightarrow \mathbb{R} \) or \( \overline{\mathcal{X}} \rightarrow \{\pm 1\} \)

\( \mathcal{F} \) a family of functions

\( \rho_f(x, y) \) margin of function \( f \) on the example \( (x, y) \), i.e., \( y \cdot f(x) \)

\( \rho_f \) margin of \( f \) on the training set, i.e., \( \min_{i=1}^{n} \rho_f(x_i, y_i) \)

\( h \) VC dimension

\( C \) regularization parameter in front of the empirical risk term

\( \lambda \) 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

\( (x_j) \), or \( (x_j) \) shorthand for a sequence \( (x_j) = (x_1, x_2, \ldots) \)

\( \ell_p \) 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

\( 1 \) unit matrix, or identity map \( (1(x) = x \) for all \( x \) \)

\( |C| \) cardinality of a set \( C \) (for finite sets, the number of elements)

\( \gamma \) regularization operator

\( \delta_{ij} \) Kronecker \( \delta \) (Section B.2.1)

\( \delta_x \) Dirac \( \delta \), 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 \( |f(n)| \leq Cg(n) \) for all \( n \)

\( o(g(n)) \) a function is said to be \( o(g(n)) \) if there exists a constant \( c \)

such that \( |f(n)| \leq cg(n) \) for all \( n \)

\( \text{rhs/lhs} \) shorthand for “right/left hand side”

\( \bullet \) the end of a proof

\( \bullet \) easy problem

\( \bullet \bullet \) intermediate problem

\( \bullet \bullet \bullet \) difficult problem

\( \bullet \bullet \bullet \bullet \bullet \bullet \) open problem