Notation for the Classification Lectures

Leandro Minku

September 23, 2023

This document contains the key notations used in this course. Please note that the font style for the calligraphic letters used in the slides has some variation compared to the font used in these notes, as the same calligraphic font is not available to be used in the two of them.

Mathematical notations:

- Scalar: lower case, e.g., a, b.
- Column vector: lower case, bold, e.g., $\mathbf{x}$. When using handwriting, lower case with arrow, e.g., $\vec{x}$.
- Vector element: lower case with subscript, e.g., x_1, x_2.
- Matrix: upper case, bold, e.g., $\mathbf{X}$.
- Matrix element: upper case with subscripts, e.g., $X_{1,2}$.
- If enumerating these (e.g., having multiple vectors), superscript will be used to differentiate this from indices, e.g., $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$.
- Use mathcal font for sets, e.g., $\mathcal{T}$ for training set.

Machine learning notations:

$\mathcal{T}$	training set
$\mathbf{x} = (x_1, x_2, \cdots, x_d)^T$	input variables
$\phi(\mathbf{x})$	basis expansion to create (higher dimensional embedding)
y	output variable
$(\mathbf{x}, y)$	a given example
$(\mathbf{x}^{(i)}, y^{(i)})$	example i, with its input variables $\mathbf{x}^{(i)}$ and output variable $y^{(i)}$
d	number of dimensions of the input space
N	Number of training examples
$\mathcal{X}$	domain of the input variables, input space
$\mathcal{Y}$	domain of the output variable, output space
$\mathcal{L}$	likelihood
c_1	class 1
c_0	class 0
p_1	$p_1 = p(c_1\|\mathbf{x}) = p(c_1\|\mathbf{x}, \mathbf{w})$
p_i	$p_i = p(c_i\|\mathbf{x}) = p(c_i\|\mathbf{x}, \mathbf{w})$
E	a given error function, e.g., the cross entropy loss function
$\tilde{E}$	a given regularised error function, e.g., the cross entropy loss function with L2 normalisation
L	function to be optimised in support vector machines
$\tilde{L}$	function to be optimised in the dual representation of the support vector machines problem
$\mathbf{w}$	parameters of a given machine learning model, e.g., the coefficients (a.k.a. weights) of logistic regression
b	bias parameter of a linear classifier (we are using it for support vector machines, but not for logistic regression)
$\mathbf{a}$	vector of Lagrange multipliers
$H_E(\mathbf{w})$	Hessian of the function E with respect to $\mathbf{w}$
$\nabla E(\mathbf{w})$	gradient of the function E with respect to $\mathbf{w}$
C	Upper value of the box constraint for the dual representation of the support vector machines optimisation problem
ξ	Slack variable for soft margin support vector machine

Equivalent terms:

- Input variable, independent variable, input attribute, input feature.
- Output variable, dependent variable, output attribute, output feature.
- Example, observation, data point, instance.