`latex-math` Macros

compiled: 2021-10-18

Latex macros like `\frac{#1}{#2}` with arguments are displayed as $\frac{\#1}{\#2}$.

Note that macro declarations may only span a single line to be displayed correctly in the below tables.

Contents

basic-math

Macro	Notation	Comment
`\N`	$\mathbb{N}$	N, naturals
`\Z`	$\mathbb{Z}$	Z, integers
`\Q`	$\mathbb{Q}$	Q, rationals
`\R`	$\mathbb{R}$	R, reals
`\C`	$\mathbb{C}$	C, complex
`\continuous`	$\mathcal{C}$	C, space of continuous functions
`\M`	$\mathcal{M}$	machine numbers
`\epsm`	ϵ_m	maximum error
`\setzo`	$\{0, 1\}$	set 0, 1
`\setmp`	$\{-1, +1\}$	set -1, 1
`\unitint`	$[0, 1]$	unit interval
`\xt`	$\tilde{x}$	x tilde
`\argmax`	$\arg\max$	argmax
`\argmin`	$\arg\min$	argmin
`\argminlim`	$\arg\min$	argmax with limits
`\argmaxlim`	$\arg\max$	argmin with limits
`\sign`	sign	sign, signum
`\I`	$\mathbb{I}$	I, indicator
`\order`	$\mathcal{O}$	O, order
`\pd`	$\frac{\partial \#1}{\partial \#2}$	partial derivative
`\sumin`	$\sum\limits_{i=1}^{n}$	summation from i=1 to n
`\sumim`	$\sum\limits_{i=1}^{m}$	summation from i=1 to m
`\sumjp`	$\sum\limits_{j=1}^{p}$	summation from j=1 to p
`\sumik`	$\sum\limits_{i=1}^{k}$	summation from i=1 to k
`\sumkg`	$\sum\limits_{k=1}^{g}$	summation from k=1 to g
`\sumjg`	$\sum\limits_{j=1}^{g}$	summation from j=1 to g
`\meanin`	$\frac{1}{n}\sum\limits_{i=1}^{n}$	mean from i=1 to n
`\meankg`	$\frac{1}{g}\sum\limits_{k=1}^{g}$	mean from k=1 to g
`\prodin`	$\prod\limits_{i=1}^{n}$	product from i=1 to n
`\prodkg`	$\prod\limits_{k=1}^{g}$	product from k=1 to g
`\prodjp`	$\prod\limits_{j=1}^{p}$	product from j=1 to p
`\one`	$\mathbf{1}$	1, unitvector
`\zero`	$\mathbf{0}$	0-vector
`\id`	$\boldsymbol{I}$	I, identity
`\diag`	diag	diag, diagonal
`\trace`	tr	tr, trace
`\spn`	span	span
`\scp`	$\langle \#1, \#2 \rangle$	<.,.>, scalarproduct
`\mat`	$\begin{pmatrix} \#1 \end{pmatrix}$	short pmatrix command
`\Amat`	$\mathbf{A}$	matrix A
`\Deltab`	$\boldsymbol{\Delta}$	error term for vectors
`\P`	$\mathbb{P}$	P, probability

`\E`	$\mathbb{E}$	E, expectation
`\var`	Var	Var, variance
`\cov`	Cov	Cov, covariance
`\corr`	Corr	Corr, correlation
`\normal`	$\mathcal{N}$	N of the normal distribution
`\iid`	$\overset{i.i.d}{\sim}$	dist with i.i.d superscript
`\distas`	$\overset{\#1}{\sim}$	... is distributed as ...

Back to contents

basic-ml

Macro	Notation	Comment
`\Xspace`	$\mathcal{X}$	X, input space
`\Yspace`	$\mathcal{Y}$	Y, output space
`\nset`	$\{1, \ldots, n\}$	set from 1 to n
`\pset`	$\{1, \ldots, p\}$	set from 1 to p
`\gset`	$\{1, \ldots, g\}$	set from 1 to g
`\Pxy`	$\mathbb{P}_{xy}$	P_xy
`\Exy`	$\mathbb{E}_{xy}$	E_xy: Expectation over random variables xy
`\xv`	$\mathbf{x}$	vector x (bold)
`\xtil`	$\tilde{\mathbf{x}}$	vector x-tilde (bold)
`\yv`	$\mathbf{y}$	vector y (bold)
`\xy`	$(\mathbf{x}, y)$	observation (x, y)
`\xvec`	$(x_1, \ldots, x_p)^T$	(x1, ..., xp)
`\Xmat`	$\mathbf{X}$	Design matrix
`\allDatasets`	$\mathbb{D}$	The set of all datasets
`\allDatasetsn`	$\mathbb{D}_n$	The set of all datasets of size n
`\D`	$\mathcal{D}$	D, data
`\Dn`	$\mathcal{D}_n$	D_n, data of size n
`\Dtrain`	$\mathcal{D}_{\text{train}}$	D_train, training set
`\Dtest`	$\mathcal{D}_{\text{test}}$	D_test, test set
`\xyi`	$\left(\mathbf{x}^{(\#1)}, y^{(\#1)}\right)$	(x^i, y^i), i-th observation
`\Dset`	$\left(\left(\mathbf{x}^{(1)}, y^{(1)}\right), \ldots, \left(\mathbf{x}^{(n)}, y^{(n)}\right)\right)$	{(x1,y1)), ..., (xn,yn)}, data
`\defAllDatasetsn`	$(\mathcal{X} \times \mathcal{Y})^n$	Def. of the set of all datasets of size n
`\defAllDatasets`	$\bigcup_{n \in \mathbb{N}} (\mathcal{X} \times \mathcal{Y})^n$	Def. of the set of all datasets
`\xdat`	$\left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\right\}$	{x1, ..., xn}, input data
`\yvec`	$\left(y^{(1)}, \ldots, y^{(n)}\right)^T$	(y1, ..., yn), vector of outcomes
`\xi`	$\mathbf{x}^{(\#1)}$	x^i, i-th observed value of x
`\yi`	$y^{(\#1)}$	y^i, i-th observed value of y
`\xivec`	$\left(x_1^{(i)}, \ldots, x_p^{(i)}\right)^T$	(x1^i, ..., xp^i), i-th observation vector
`\xj`	$\mathbf{x}_j$	x_j, j-th feature
`\xjvec`	$\left(x_j^{(1)}, \ldots, x_j^{(n)}\right)^T$	(x^1_j, ..., x^n_j), j-th feature vector
`\phiv`	ϕ	Basis transformation function phi
`\phixi`	$\phi^{(i)}$	Basis transformation of xi: phi^i := phi(xi)
`\lamv`	$\boldsymbol{\lambda}$	lambda vector, hyperconfiguration vector
`\Lam`	$\boldsymbol{\Lambda}$	Lambda, space of all hpos
`\preimageInducer`	$\left(\bigcup_{n \in \mathbb{N}} (\mathcal{X} \times \mathcal{Y})^n\right) \times \boldsymbol{\Lambda}$	Set of all datasets times the hyperparameter space
`\preimageInducerShort`	$\mathbb{D} \times \boldsymbol{\Lambda}$	Set of all datasets times the hyperparameter space
`\inducer`	$\mathcal{I}$	Inducer, inducing algorithm, learning algorithm
`\ftrue`	f_{true}	True underlying function (if a statistical model is assumed)
`\ftruex`	$f_{\text{true}}(\mathbf{x})$	True underlying function (if a statistical model is assumed)
`\fx`	$f(\mathbf{x})$	f(x), continuous prediction function
`\Hspace`	$\mathcal{H}$	hypothesis space where f is from
`\fkx`	$f_{\#1}(\mathbf{x})$	f_j(x), discriminant component function
`\fh`	$\hat{f}$	f hat, estimated prediction function
`\fxh`	$\hat{f}(\mathbf{x})$	fhat(x)
`\fxt`	$f(\mathbf{x} \mid \boldsymbol{\theta})$	f(x | theta)
`\fxi`	$f\left(\mathbf{x}^{(i)}\right)$	f(x^(i))
`\fxih`	$\hat{f}\left(\mathbf{x}^{(i)}\right)$	f(x^(i))
`\fxit`	$f\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right)$	f(x^(i) | theta)
`\fhD`	$\hat{f}_{\mathcal{D}}$	fhat_D, estimate of f based on D
`\fhDtrain`	$\hat{f}_{\mathcal{D}_{\text{train}}}$	fhat_Dtrain, estimate of f based on D
`\fhDnlambda`	$\hat{f}_{\mathcal{D}_n, \boldsymbol{\lambda}}$	model learned on Dn with hp lambda

<table>
<tr><td>\fhDlambda</td><td>$\hat{f}_{\mathcal{D}, \boldsymbol{\lambda}}$</td><td>model learned on D with hp lambda</td></tr>
<tr><td>\fhDnlambdastar</td><td>$\hat{f}_{\mathcal{D}_n, \boldsymbol{\lambda}^\ast}$</td><td>model learned on Dn with optimal hp lambda</td></tr>
<tr><td>\fhDlambdastar</td><td>$\hat{f}_{\mathcal{D}, \boldsymbol{\lambda}^\ast}$</td><td>model learned on D with optimal hp lambda</td></tr>
<tr><td>\hx</td><td>$h(\mathbf{x})$</td><td>h(x), discrete prediction function</td></tr>
<tr><td>\hh</td><td>$\hat{h}$</td><td>h hat</td></tr>
<tr><td>\hxh</td><td>$\hat{h}(\mathbf{x})$</td><td>hhat(x)</td></tr>
<tr><td>\hxt</td><td>$h(\mathbf{x}|\boldsymbol{\theta})$</td><td>h(x | theta)</td></tr>
<tr><td>\hxi</td><td>$h\left(\mathbf{x}^{(i)}\right)$</td><td>h(x^(i))</td></tr>
<tr><td>\hxit</td><td>$h\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right)$</td><td>h(x^(i) | theta)</td></tr>
<tr><td>\yh</td><td>$\hat{y}$</td><td>yhat for prediction of target</td></tr>
<tr><td>\yih</td><td>$\hat{y}^{(i)}$</td><td>yhat^(i) for prediction of ith targiet</td></tr>
<tr><td>\thetah</td><td>$\hat{\theta}$</td><td>theta hat</td></tr>
<tr><td>\thetab</td><td>$\boldsymbol{\theta}$</td><td>theta vector</td></tr>
<tr><td>\thetabh</td><td>$\hat{\boldsymbol{\theta}}$</td><td>theta vector hat</td></tr>
<tr><td>\thetat</td><td>$\boldsymbol{\theta}^{[\#1]}$</td><td>theta^[t] in optimization</td></tr>
<tr><td>\thetatn</td><td>$\boldsymbol{\theta}^{[\#1 +1]}$</td><td>theta^[t+1] in optimization</td></tr>
<tr><td>\thx</td><td>$\boldsymbol{\theta}^T \mathbf{x}$</td><td>linear combination with theta</td></tr>
<tr><td>\thetahDnlambda</td><td>$\hat{\boldsymbol{\theta}}_{\mathcal{D}_n, \boldsymbol{\lambda}}$</td><td>theta learned on Dn with hp lambda</td></tr>
<tr><td>\thetahDlambda</td><td>$\hat{\boldsymbol{\theta}}_{\mathcal{D}, \boldsymbol{\lambda}}$</td><td>theta learned on D with hp lambda</td></tr>
<tr><td>\pdf</td><td>p</td><td>p</td></tr>
<tr><td>\pdfx</td><td>$p(\mathbf{x})$</td><td>p(x)</td></tr>
<tr><td>\pixt</td><td>$\pi(\mathbf{x} \mid \boldsymbol{\theta})$</td><td>pi(x|theta), pdf of x given theta</td></tr>
<tr><td>\pixit</td><td>$\pi\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right)$</td><td>pi(x^i|theta), pdf of x given theta</td></tr>
<tr><td>\pixii</td><td>$\pi\left(\mathbf{x}^{(i)}\right)$</td><td>pi(x^i), pdf of i-th x</td></tr>
<tr><td>\pdfxy</td><td>$p(\mathbf{x}, y)$</td><td>p(x, y)</td></tr>
<tr><td>\pdfxyt</td><td>$p(\mathbf{x}, y \mid \boldsymbol{\theta})$</td><td>p(x, y | theta)</td></tr>
<tr><td>\pdfxyit</td><td>$p\left(\mathbf{x}^{(i)}, y^{(i)} \mid \boldsymbol{\theta}\right)$</td><td>p(x^(i), y^(i) | theta)</td></tr>
<tr><td>\pdfxyk</td><td>$p(\mathbf{x}|y = \#1)$</td><td>p(x | y = k)</td></tr>
<tr><td>\lpdfxyk</td><td>$\log p(\mathbf{x}|y = \#1)$</td><td>log p(x | y = k)</td></tr>
<tr><td>\pdfxiyk</td><td>$p\left(\mathbf{x}^{(i)}|y = \#1\right)$</td><td>p(x^i | y = k)</td></tr>
<tr><td>\pik</td><td>$\pi_{\#1}$</td><td>pi_k, prior</td></tr>
<tr><td>\lpik</td><td>$\log \pi_{\#1}$</td><td>log pi_k, log of the prior</td></tr>
<tr><td>\pit</td><td>$\pi(\boldsymbol{\theta})$</td><td>Prior probability of parameter theta</td></tr>
<tr><td>\post</td><td>$\mathbb{P}(y = 1 \mid \mathbf{x})$</td><td>P(y = 1 | x), post. prob for y=1</td></tr>
<tr><td>\postk</td><td>$\mathbb{P}(y = \#1 \mid \mathbf{x})$</td><td>P(y = k | y), post. prob for y=k</td></tr>
<tr><td>\pix</td><td>$\pi(\mathbf{x})$</td><td>pi(x), P(y = 1 | x)</td></tr>
<tr><td>\pikx</td><td>$\pi_{\#1}(\mathbf{x})$</td><td>pi_k(x), P(y = k | x)</td></tr>
<tr><td>\pikxt</td><td>$\pi_{\#1}(\mathbf{x} \mid \boldsymbol{\theta})$</td><td>pi_k(x | theta), P(y = k | x, theta)</td></tr>
<tr><td>\pixh</td><td>$\hat{\pi}(\mathbf{x})$</td><td>pi(x) hat, P(y = 1 | x) hat</td></tr>
<tr><td>\pikxh</td><td>$\hat{\pi}_{\#1}(\mathbf{x})$</td><td>pi_k(x) hat, P(y = k | x) hat</td></tr>
<tr><td>\pixih</td><td>$\hat{\pi}(\mathbf{x}^{(i)})$</td><td>pi(x^(i)) with hat</td></tr>
<tr><td>\pikxih</td><td>$\hat{\pi}_{\#1}(\mathbf{x}^{(i)})$</td><td>pi_k(x^(i)) with hat</td></tr>
<tr><td>\pdfygxt</td><td>$p(y \mid \mathbf{x}, \boldsymbol{\theta})$</td><td>p(y | x, theta)</td></tr>
<tr><td>\pdfyigxit</td><td>$p\left(y^{(i)} \mid \mathbf{x}^{(i)}, \boldsymbol{\theta}\right)$</td><td>p(y^i |x^i, theta)</td></tr>
<tr><td>\lpdfygxt</td><td>$\log p(y \mid \mathbf{x}, \boldsymbol{\theta})$</td><td>log p(y | x, theta)</td></tr>
<tr><td>\lpdfyigxit</td><td>$\log p\left(y^{(i)} \mid \mathbf{x}^{(i)}, \boldsymbol{\theta}\right)$</td><td>log p(y^i |x^i, theta)</td></tr>
<tr><td>\eps</td><td>ϵ</td><td>residual, stochastic</td></tr>
<tr><td>\epsi</td><td>$\epsilon^{(i)}$</td><td>epsilon^i, residual, stochastic</td></tr>
<tr><td>\epsh</td><td>$\hat{\epsilon}$</td><td>residual, estimated</td></tr>
<tr><td>\yf</td><td>$y f(\mathbf{x})$</td><td>y f(x), margin</td></tr>
<tr><td>\yfi</td><td>$y^{(i)} f\left(\mathbf{x}^{(i)}\right)$</td><td>y^i f(x^i), margin</td></tr>
<tr><td>\Sigmah</td><td>$\hat{\Sigma}$</td><td>estimated covariance matrix</td></tr>
<tr><td>\Sigmahj</td><td>$\hat{\Sigma}_j$</td><td>estimated covariance matrix for the j-th class</td></tr>
<tr><td>\Lyf</td><td>$L\left(y, f\right)$</td><td>L(y, f), loss function</td></tr>
<tr><td>\Lxy</td><td>$L\left(y, f(\mathbf{x})\right)$</td><td>L(y, f(x)), loss function</td></tr>
</table>

\Lxyi	$L\left(y^{(i)}, f\left(\mathbf{x}^{(i)}\right)\right)$	loss of observation
\Lxyt	$L\left(y, f(\mathbf{x} \mid \boldsymbol{\theta})\right)$	loss with f parameterized
\Lxyit	$L\left(y^{(i)}, f\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right)\right)$	loss of observation with f parameterized
\Lxym	$L\left(y^{(i)}, f\left(\tilde{\boldsymbol{x}}^{(i)} \mid \boldsymbol{\theta}\right)\right)$	loss of observation with f parameterized
\Lpixy	$L\left(y, \pi(\mathbf{x})\right)$	loss in classification
\Lpixyi	$L\left(y^{(i)}, \pi\left(\mathbf{x}^{(i)}\right)\right)$	loss of observation in classification
\Lpixyt	$L\left(y, \pi(\mathbf{x} \mid \boldsymbol{\theta})\right)$	loss with pi parameterized
\Lpixyit	$L\left(y^{(i)}, \pi\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right)\right)$	loss of observation with pi parameterized
\Lhxy	$L\left(y, h(\mathbf{x})\right)$	L(y, h(x)), loss function on discrete classes
\Lr	$L\left(r\right)$	L(r), loss defined on residual (reg) / margin (classif)
\risk	$\mathcal{R}$	R, risk
\riskf	$\mathcal{R}(f)$	R(f), risk
\riskt	$\mathcal{R}(\boldsymbol{\theta})$	R(theta), risk
\riske	$\mathcal{R}_{\text{emp}}$	R_emp, empirical risk w/o factor 1 / n
\riskeb	$\bar{\mathcal{R}}_{\text{emp}}$	R_emp, empirical risk w/ factor 1 / n
\riskef	$\mathcal{R}_{\text{emp}}(f)$	R_emp(f)
\risket	$\mathcal{R}_{\text{emp}}(\boldsymbol{\theta})$	R_emp(theta)
\riskr	$\mathcal{R}_{\text{reg}}$	R_reg, regularized risk
\riskrt	$\mathcal{R}_{\text{reg}}(\boldsymbol{\theta})$	R_reg(theta)
\riskrf	$\mathcal{R}_{\text{reg}}(f)$	R_reg(f)
\riskrth	$\hat{\mathcal{R}}_{\text{reg}}(\boldsymbol{\theta})$	hat R_reg(theta)
\risketh	$\hat{\mathcal{R}}_{\text{emp}}(\boldsymbol{\theta})$	hat R_emp(theta)
\LL	$\mathcal{L}$	L, likelihood
\LLt	$\mathcal{L}(\boldsymbol{\theta})$	L(theta), likelihood
\logl	ℓ	l, log-likelihood
\loglt	$\ell(\boldsymbol{\theta})$	l(theta), log-likelihood
\errtrain	$\text{err}_{\text{train}}$	training error
\errtest	err_{test}	test error
\errexp	$\overline{\text{err}_{\text{test}}}$	avg training error

Back to contents

ml-ensembles

Macro	Notation	Comment
`\bl`	$b^{[\#1]}$	baselearner, default m
`\blh`	$\hat{b}^{[\#1]}$	estimated base learner, default m
`\blx`	$b^{[\#1]}(\mathbf{x})$	baselearner, default m
`\fM`	$f^{[M]}(\mathbf{x})$	ensembled predictor
`\fMh`	$\hat{f}^{[M]}(\mathbf{x})$	estimated ensembled predictor
`\ambifM`	$\Delta\left(f^{[M]}(\mathbf{x})\right)$	ambiguity/instability of ensemble
`\betam`	$\beta^{[\#1]}$	weight of basemodel m
`\betamh`	$\hat{\beta}^{[\#1]}$	weight of basemodel m with hat
`\betaM`	$\beta^{[M]}$	last baselearner
`\fm`	$f^{[\#1]}$	prediction in iteration m
`\fmh`	$\hat{f}^{[\#1]}$	prediction in iteration m
`\fmd`	$f^{[\#1-1]}$	prediction m-1
`\fmdh`	$\hat{f}^{[\#1-1]}$	prediction m-1
`\errm`	$\mathrm{err}^{[\#1]}$	weighted in-sample misclassification rate
`\wm`	$w^{[\#1]}$	weight vector of basemodel m
`\wmi`	$w^{[\#1](i)}$	weight of obs i of basemodel m
`\thetam`	$\boldsymbol{\theta}^{[\#1]}$	parameters of basemodel m
`\thetamh`	$\hat{\boldsymbol{\theta}}^{[\#1]}$	parameters of basemodel m with hat
`\blxt`	$b(\mathbf{x}, \boldsymbol{\theta}^{[\#1]})$	baselearner, default m
`\ens`	$\sum_{m=1}^{M} \beta^{[m]} b(\mathbf{x}, \boldsymbol{\theta}^{[m]})$	ensemble
`\rmm`	$\tilde{r}^{[\#1]}$	pseudo residuals
`\rmi`	$\tilde{r}^{[\#1](i)}$	pseudo residuals
`\Rtm`	$R_t^{[\#1]}$	terminal-region
`\Tm`	$T^{[\#1]}$	
`\ctm`	$c_t^{[\#1]}$	mean, terminal-regions
`\ctmh`	$\hat{c}_t^{[\#1]}$	mean, terminal-regions with hat
`\ctmt`	$\tilde{c}_t^{[\#1]}$	mean, terminal-regions
`\Lp`	L'	
`\Ldp`	L''	
`\Lpleft`	L'_{left}	

Back to contents

ml-eval

Macro	Notation	Comment
\ntest	n_{test}	size of the test set
\ntrain	n_{train}	size of the train set
\ntesti	$n_{\text{test},\#1}$	size of the i-th test set
\ntraini	$n_{\text{train},\#1}$	size of the i-th train set
\Jtrain	J_{train}	index vector associated to the train data
\Jtest	J_{test}	index vector associated to the test data
\Jtraini	$J_{\text{train},\#1}$	index vector associated to the i-th train dataset
\Jtesti	$J_{\text{test},\#1}$	index vector associated to the i-th test dataset
\Dtraini	$\mathcal{D}_{\text{train},\#1}$	D_train,i, i-th training set
\Dtesti	$\mathcal{D}_{\text{test},\#1}$	D_test,i, i-th test set
\JSpace	$\{1,\ldots,n\}^{\#1}$	space of train indices of size m_train
\JtrainSpace	$\{1,\ldots,n\}^{n_{\text{train}}}$	space of train indices of size m_train
\JtestSpace	$\{1,\ldots,n\}^{n_{\text{test}}}$	space of train indices of size m_test
\yJ	$\mathbf{y}_{\#1}$	output vector associated to index J
\yJDef	$\left(y^{(J^{(1)})},\ldots,y^{(J^{(m)})}\right)$	def of the output vector associated to index J
\JJ	$\mathcal{J}$	cali-J, set of all splits
\JJset	$((J_{\text{train},1},J_{\text{test},1}),\ldots,(J_{\text{train},B},J_{\text{test},B}))$	(Jtrain_1,Jtest_1) ...(Jtrain_B,Jtest_B)
\GE	GE	GE
\GEh	$\widehat{\mathrm{GE}}$	GE-hat
\GEfull	$\mathrm{GE}(\mathcal{I},\boldsymbol{\lambda},\#1,\rho)$	GE(I, lam, ?, rho)
\GEhholdout	$\widehat{\mathrm{GE}}_{J_{\text{train}},J_{\text{test}}}(\mathcal{I},\boldsymbol{\lambda},\lvert J_{\text{train}}\rvert,\rho)$	GE-hat_{Jtrain,Jtest} (I, lam, |J|, rho)
\GEhholdouti	$\widehat{\mathrm{GE}}_{J_{\text{train},\#1},J_{\text{test},\#1}}(\mathcal{I},\boldsymbol{\lambda},\lvert J_{\text{train},\#1}\rvert,\rho)$	GE-hat_{Jtrain_i,Jtest_i} (I, lam, |Jtrain_i|, rho)
\GEhlam	$\widehat{\mathrm{GE}}(\boldsymbol{\lambda})$	GE-hat(lam)
\GEhlamsubIJrho	$\widehat{\mathrm{GE}}_{\mathcal{I},\mathcal{J},\rho}(\boldsymbol{\lambda})$	GE-hat_I,J,rho(lam)
\GEhresa	$\widehat{\mathrm{GE}}(\mathcal{I},\mathcal{J},\rho,\boldsymbol{\lambda})$	GE-hat_I,J,rho(lam)
\GErhoDef	$\lim_{n_{\text{test}}\to\infty}\mathbb{E}\left[\rho\left(\mathbf{y}_{J_{\text{test}}},\boldsymbol{F}_{J_{\text{test}},\mathcal{I}(\mathcal{D}_{\text{train}},\boldsymbol{\lambda})}\right)\right]$	GE formal def
\agr	agr	aggregate function
\GEf	$\mathrm{GE}\left(\hat{f}\right)$	Generalization error of a fitted model
\GEind	$GE_n\left(\mathcal{I}_{L,O}\right)$	Generalization error of a fitted model
\GEnf	$GE_n\left(\hat{f}_{\#1}\right)$	Generalization error GE
\GEhat	$\widehat{\mathrm{GE}}$	Estimated train error
\GED	$\mathrm{GE}_{\mathcal{D}}$	Generalization error GE
\EGEn	EGE_n	expected GE
\EDn	$\mathbb{E}_{\lvert D\rvert=n}$	expectation wrt data of size n
\rhoL	ρ_L	perf. measure derived from pointwise loss function L
\F	$\boldsymbol{F}$	matrix of prediction scores
\Fi	$\boldsymbol{F}^{(\#1)}$	i'th row vector of the prediction scores matrix
\FJ	$\boldsymbol{F}_{\#1}$	predscore mat index vector J
\FJf	$\boldsymbol{F}_{J,f}$	predscore mat index vector J and model f
\FJtestfh	$\boldsymbol{F}_{J_{\text{test}},\hat{f}}$	predscore mat index vector Jtest and model f hat
\FJtestftrain	$\boldsymbol{F}_{J_{\text{test}},\mathcal{I}(\mathcal{D}_{\text{train}},\boldsymbol{\lambda})}$	predscore mat index vector Jtest and model f
\FJtestftraini	$\boldsymbol{F}_{J_{\text{test},\#1},\mathcal{I}(\mathcal{D}_{\text{train},\#1},\boldsymbol{\lambda})}$	predscore mat i-th index vector Jtest and model f
\FJfDef	$\left(f(\mathbf{x}^{(J^{(1)})}),\ldots,f(\mathbf{x}^{(J^{(m)})})\right)$	def of predscore mat index vector J and model f
\preimageRho	$\bigcup_{m\in\mathbb{N}}\left(\mathcal{Y}^m\times\mathbb{R}^{m\times g}\right)$	Set of all datasets times the hyperparameter space
\np	n_+	no. of positive instances
\nn	n_-	no. of negative instances
\rn	π_-	proportion negative instances
\rp	π_+	proportion negative instances
\tp	$\#\mathrm{TP}$	true pos
\fap	$\#\mathrm{FP}$	false pos (fp taken for partial derivs)
\tn	$\#\mathrm{TN}$	true neg

`\fan`	#FN	false neg

Back to contents

ml-feature-sel

Macro	Notation	Comment
`\xjNull`	x_{j_0}	
`\xjEins`	x_{j_1}	
`\xl`	$\mathbf{x}_l$	
`\pushcode`		

Back to contents

ml-gp

Macro	Notation	Comment
`\gp`	$\mathcal{GP}\left(m(\mathbf{x}), k\left(\mathbf{x}, \mathbf{x}'\right)\right)$	Gaussian Process Definition
`\mvec`	$\mathbf{m}$	Gaussian process mean vector
`\Kmat`	$\mathbf{K}$	estimated base learner
`\kstarx`	$\mathbf{k}_*(x)$	cov of new obs with x
`\ls`	ℓ	length-scale

Back to contents

ml-hpo

Macro	Notation	Comment
\Ilam	$\mathcal{I}_{\boldsymbol{\lambda}}$	I_lambda
\lami	$\boldsymbol{\lambda}^{(\#1)}$	lambda i
\clam	$c(\boldsymbol{\lambda})$	c(lambda)
\clamh	$c(\hat{\boldsymbol{\lambda}})$	c(lambda-hat)
\lams	$\boldsymbol{\lambda}^*$	Theoretical min of c
\lamh	$\hat{\boldsymbol{\lambda}}$	returned lambda of HPO
\LamS	$\tilde{\boldsymbol{\Lambda}}$	search space
\lamp	$\boldsymbol{\lambda}^+$	proposed lambda
\clamp	$c(\boldsymbol{\lambda}^+)$	c of proposed lambda
\archive	$\mathcal{A}$	archive at time step t
\archivet	$\mathcal{A}^{[\#1]}$	archive at time step t
\tuner	$\mathcal{T}$	tuner
\tunerfull	$\mathcal{T}_{\mathcal{I},\tilde{\boldsymbol{\Lambda}},\rho,\mathcal{J}}$	tuner with inducer, search space, performance measure and resampling strategy
\chlam	$\hat{c}(\boldsymbol{\lambda})$	post mean of SM
\shlam	$\hat{\sigma}(\boldsymbol{\lambda})$	post sd of SM
\vhlam	$\hat{\sigma^2}(\boldsymbol{\lambda})$	post var of SM
\ulam	$u(\boldsymbol{\lambda})$	acquisition function
\lambdaopt	λ^*	Minimum of the black box function Psi
\metadata	$\left\{\left(\boldsymbol{\lambda}^{(i)}, \Psi^{[i]}\right)\right\}$	Metadata for the Gaussian process
\lamvec	$\left(\lambda^{[1]}, \dots, \lambda^{[m_{\text{init}}]}\right)$	Vector of different inputs
\minit	m_{init}	Size of the initial design
\lambu	λ_{budget}	single lambda_budget komponent HP
\lamfid	λ_{fid}	single lambda_budget komponent HP
\lamfidl	$\lambda_{\text{fid}}^{\text{low}}$	single lambda_budget komponent HP
\lamfidu	$\lambda_{\text{fid}}^{\text{upp}}$	single lambda_budget komponent HP
\etahb	η_{HB}	HB multiplier eta
\costs	$\mathcal{C}$	costs
\Celite	$\boldsymbol{\theta}^*$	elite configurations
\instances	$\mathcal{I}$	sequence of instances
\budget	$\mathcal{B}$	computational budget

Back to contents

ml-interpretable

Macro	Notation	Comment
`\fj`	f_j	marginal function f_j, depending on feature j
`\fnj`	f_{-j}	marginal function f_{-j}, depending on all features but j
`\fS`	f_S	marginal function f_S depending on feature set S
`\fC`	f_C	marginal function f_C depending on feature set C
`\fhj`	$\hat{f}_j$	marginal function fh_j, depending on feature j
`\fhnj`	$\hat{f}_{-j}$	marginal function fh_{-j}, depending on all features but j
`\fhS`	$\hat{f}_S$	marginal function fh_S depending on feature set S
`\fhC`	$\hat{f}_C$	marginal function fh_C depending on feature set C
`\XSmat`	$\mathbf{X}_S$	Design matrix subset
`\XCmat`	$\mathbf{X}_C$	Design matrix subset
`\Xnj`	$\mathbf{X}_{-j}$	Design matrix subset -j = {1, .., j-1, j+1, ..., p}
`\Scupj`	$S \cup \{j\}$	coalition S but without player j
`\Scupk`	$S \cup \{k\}$	coalition S but without player k
`\SsubP`	$S \subseteq P$	coalition S subset of P
`\SsubPnoj`	$S \subseteq P \setminus \{j\}$	coalition S subset of P without player j
`\SsubPnojk`	$S \subseteq P \setminus \{j, k\}$	coalition S subset of P without player k
`\phiij`	$\hat{\phi}_j^{(i)}$	Shapley value for feature j and observation i
`\Gspace`	$\mathcal{G}$	Hypothesis space for surrogate model
`\neigh`	$\phi_{\mathbf{x}}$	Proximity measure
`\zv`	$\mathbf{z}$	Sampled datapoints for surrogate
`\Zspace`	$\mathcal{Z}$	Space of sampled datapoints
`\Gower`	d_G	Gower distance

Back to contents

ml-nn

Macro	Notation	Comment
\neurons	$z_1, \dots, z_M$	vector of neurons
\hidz	$\mathbf{z}$	vector of hidden activations
\biasb	$\mathbf{b}$	bias vector
\biasc	c	bias in output
\wtw	$\mathbf{w}$	weight vector (general)
\Wmat	$\mathbf{W}$	weight vector (general)
\wtu	$\mathbf{u}$	weight vector of output neuron
\Oreg	$R_{reg}(\theta\|X, y)$	regularized objective function
\Ounreg	$R_{emp}(\theta\|X, y)$	unconstrained objective function
\Pen	$\Omega(\theta)$	penalty
\Oregweight	$R_{reg}(w\|X, y)$	regularized objective function with weight
\Oweight	$R_{emp}(w\|X, y)$	unconstrained objective function with weight
\Oweighti	$R_{emp}(w_i\|X, y)$	unconstrained objective function with weight w_i
\Oweightopt	$J(w^*\|X, y)$	unconstrained objective function withoptimal weight
\Oopt	$\hat{J}(\theta\|X, y)$	optimal objective function
\Odropout	$J(\theta, \mu\|X, y)$	dropout objective function
\Loss	$L(y, f(\mathbf{x}, \boldsymbol{\theta}))$	
\Lmomentumnest	$L(y^{(i)}, f(x^{(i)}, \boldsymbol{\theta} + \varphi\boldsymbol{\nu}))$	momentum risk
\Lmomentumtilde	$L(y^{(i)}, f(x^{(i)}, \tilde{\boldsymbol{\theta}}))$	Nesterov momentum risk
\Lmomentum	$L(y^{(i)}, f(x^{(i)}, \boldsymbol{\theta}))$	
\Hess	$\mathbf{H}$	
\nub	$\boldsymbol{\nu}$	
\uauto	$L(x, g(f(x)))$	undercomplete autoencoder objective function
\dauto	$L(x, g(f(\tilde{x})))$	denoising autoencoder objective function
\deltab	$\boldsymbol{\delta}$	
\Lossdeltai	$L(y^{(i)}, f(\mathbf{x}^{(i)} + \boldsymbol{\delta}\|\boldsymbol{\theta}))$	
\Lossdelta	$L(y, f(\mathbf{x} + \boldsymbol{\delta}\|\boldsymbol{\theta}))$	

Back to contents

ml-survival

Macro	Notation	Comment
`\Ti`	$T^{(\#1)}$	??
`\Ci`	$C^{(\#1)}$	??
`\oi`	$o^{(\#1)}$	??
`\ti`	$t^{(\#1)}$	??
`\deltai`	$\delta^{(\#1)}$	
`\Lxdi`	$L\left(\boldsymbol{\delta}, f(\mathbf{x})\right)$	

Back to contents

ml-svm

Macro	Notation	Comment
`\sv`	SV	supportvectors
`\HS`	Φ	H, hilbertspace
`\sl`	ζ	
`\slvec`	$\left(\zeta^{(1)}, \zeta^{(n)}\right)$	slack variables (SVM)
`\sli`	$\zeta^{(i)}$	slack variable (SVM)
`\alphah`	$\hat{\alpha}$	alpha-hat
`\alphav`	$\boldsymbol{\alpha}$	vector alpha (bold)
`\alphavh`	$\hat{\boldsymbol{\alpha}}$	vector alpha-hat
`\phix`	$\phi(\mathbf{x})$	\phi(x)
`\phixt`	$\phi(\tilde{\mathbf{x}})$	\phi(x-tilde)

Back to contents

ml-trees

Macro	Notation	Comment
`\Np`	$\mathcal{N}$	(Parent) node N
`\Npk`	$\mathcal{N}_k$	node N_k
`\Nl`	$\mathcal{N}_1$	Left node N_1
`\Nr`	$\mathcal{N}_2$	Right node N_2
`\pikN`	$\pi_k^{(\mathcal{N})}$	class probability node N
`\pikNh`	$\hat{\pi}_{\#1}^{(\mathcal{N})}$	estimated class probability node N
`\pikNlh`	$\hat{\pi}_{\#1}^{(\mathcal{N}_1)}$	
`\pikNrh`	$\hat{\pi}_{\#1}^{(\mathcal{N}_2)}$	

Back to contents

probmodel

Macro	Notation	Comment
`\muk`	$\boldsymbol{\mu_k}$	

Back to contents