`latex-math` Macros

compiled: 2021-10-17

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
\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
\fp	$\frac{\partial \cdot}{\partial \cdot}$	partial derivative
\pd	$\frac{\partial \cdot}{\partial \cdot}$	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	$\langle \cdot, \cdot \rangle$	<.,.>, scalarproduct
\mat	$\begin{pmatrix} \cdot \end{pmatrix}$	short pmatrix command
\Amat	$\mathbf{A}$	matrix A
\xv	$\mathbf{x}$	vector x (bold)
\xtil	$\tilde{\mathbf{x}}$	vector x-tilde (bold)

`\yv`	$\mathbf{y}$	vector y (bold)
`\Deltab`	$\boldsymbol{\Delta}$	error term for vectors
`\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`	$\dot{\sim}$	... is distributed as ...

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
\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
\D	$\mathcal{D}$	D, data
\obs	$\left(\mathbf{x}^{(\cdot)}, y^{(\cdot)}\right)$	observation (x^(i), y^(i))
\Dn	$\mathcal{D}_n$	D_n, data of size n
\allDatasetsn	$\mathbb{D}_n$	The set of all datasets of size n
\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
\ydat	$\mathbf{y}$	y (bold), vector of outcomes
\yvec	$\left(y^{(1)}, \ldots, y^{(n)}\right)^T$	(y1, ..., yn), vector of outcomes
\yi	$y^{(\cdot)}$	y^i, i-th observed value of y
\xyi	$\left(\mathbf{x}^{(\cdot)}, y^{(\cdot)}\right)$	(x^i, y^i), i-th observation
\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
\Dtrain	$\mathcal{D}_{\text{train}}$	D_train, training set
\Dtest	$\mathcal{D}_{\text{test}}$	D_test, test set
\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
\fix	$f_i(\mathbf{x})$	f_i(x), discriminant component function
\fjx	$f_j(\mathbf{x})$	f_j(x), discriminant component function
\fkx	$f_k(\mathbf{x})$	f_k(x), discriminant component function
\fgx	$f_g(\mathbf{x})$	f_g(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
\fhDlambda	$\hat{f}_{\mathcal{D}, \boldsymbol{\lambda}}$	model learned on D with hp lambda
\fhDnlambdastar	$\hat{f}_{\mathcal{D}_n, \boldsymbol{\lambda}^*}$	model learned on Dn with optimal hp lambda
\fhDlambdastar	$\hat{f}_{\mathcal{D}, \boldsymbol{\lambda}^*}$	model learned on D with optimal hp lambda
\hx	$h(\mathbf{x})$	h(x), discrete prediction function
\hxv	$h(\mathbf{x})$	h(x), discrete prediction function with x (vector) as input
\hh	$\hat{h}$	h hat
\hxh	$\hat{h}(\mathbf{x})$	hhat(x)
\hxt	$h(\mathbf{x}\vert\boldsymbol{\theta})$	h(x | theta)
\hxi	$h\left(\mathbf{x}^{(i)}\right)$	h(x^(i))
\hxit	$h\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right)$	h(x^(i) | theta)
\yh	$\hat{y}$	yhat for prediction of target
\yih	$\hat{y}^{(i)}$	yhat^(i) for prediction of ith targiet
\thetah	$\hat{\theta}$	theta hat
\thetab	$\boldsymbol{\theta}$	theta vector
\thetabh	$\hat{\boldsymbol{\theta}}$	theta vector hat
\thetat	$\boldsymbol{\theta}^{[t]}$	theta^[t] in optimization
\thetatn	$\boldsymbol{\theta}^{[t+1]}$	theta^[t+1] in optimization
\thetahDnlambda	$\hat{\boldsymbol{\theta}}\mathcal{D}_n, \boldsymbol{\lambda}$	theta learned on Dn with hp lambda
\thetahDlambda	$\hat{\boldsymbol{\theta}}\mathcal{D}, \boldsymbol{\lambda}$	theta learned on D with hp lambda
\pdf	p	p
\pdfx	$p(\mathbf{x})$	p(x)
\pixt	$\pi(\mathbf{x} \mid \boldsymbol{\theta})$	pi(x|theta), pdf of x given theta
\pixit	$\pi\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right)$	pi(x^i|theta), pdf of x given theta
\pixii	$\pi\left(\mathbf{x}^{(i)}\right)$	pi(x^i), pdf of i-th x
\pdfxy	$p(\mathbf{x}, y)$	p(x, y)
\pdfxyt	$p(\mathbf{x}, y \mid \boldsymbol{\theta})$	p(x, y | theta)
\pdfxyit	$p\left(\mathbf{x}^{(i)}, y^{(i)} \mid \boldsymbol{\theta}\right)$	p(x^(i), y^(i) | theta)
\pdfxyk	$p(\mathbf{x}\vert y=k)$	p(x | y = k)
\pdfxyj	$p(\mathbf{x}\vert y=j)$	p(x | y = j)
\lpdfxyk	$\log p(\mathbf{x}\vert y=k)$	log p(x | y = k)
\pdfxiyk	$p\left(\mathbf{x}^{(i)}\vert y=k\right)$	p(x^i | y = k)
\pik	π_k	pi_k, prior
\lpik	$\log \pi_k$	log pi_k, log of the prior
\pit	$\pi(\boldsymbol{\theta})$	Prior probability of parameter theta
\post	$\mathbb{P}(y=1 \mid \mathbf{x})$	P(y = 1 | x), post. prob for y=1
\pix	$\pi(\mathbf{x})$	pi(x), P(y = 1 | x)
\postk	$\mathbb{P}(y=k \mid \mathbf{x})$	P(y = k | y), post. prob for y=k
\pikx	$\pi_k(\mathbf{x})$	pi_k(x), P(y = k | x)
\pikxt	$\pi_k(\mathbf{x} \mid \boldsymbol{\theta})$	pi_k(x | theta), P(y = k | x, theta)
\pijx	$\pi_j(\mathbf{x})$	pi_j(x), P(y = j | x)
\pigx	$\pi_g(\mathbf{x})$	pi_g(x), P(y = g | x)
\pdfygxt	$p(y \mid \mathbf{x}, \boldsymbol{\theta})$	p(y | x, theta)
\pdfyigxit	$p\left(y^{(i)} \mid \mathbf{x}^{(i)}, \boldsymbol{\theta}\right)$	p(y^i |x^i, theta)
\lpdfygxt	$\log p(y \mid \mathbf{x}, \boldsymbol{\theta})$	log p(y | x, theta)
\lpdfyigxit	$\log p\left(y^{(i)} \mid \mathbf{x}^{(i)}, \boldsymbol{\theta}\right)$	log p(y^i |x^i, theta)
\pixh	$\hat{\pi}(\mathbf{x})$	pi(x) hat, P(y = 1 | x) hat
\pikxh	$\hat{\pi}_k(\mathbf{x})$	pi_k(x) hat, P(y = k | x) hat
\pixih	$\hat{\pi}(\mathbf{x}^{(i)})$	pi(x^(i)) with hat
\pikxih	$\hat{\pi}_k(\mathbf{x}^{(i)})$	pi_k(x^(i)) with hat
\eps	ϵ	residual, stochastic
\epsi	$\epsilon^{(i)}$	epsilon^i, residual, stochastic

\epsh	$\hat{\epsilon}$	residual, estimated
\yf	$yf(\mathbf{x})$	y f(x), margin
\yfi	$y^{(i)}f\left(\mathbf{x}^{(i)}\right)$	y^i f(x^i), margin
\Sigmah	$\hat{\Sigma}$	estimated covariance matrix
\Sigmahj	$\hat{\Sigma}_j$	estimated covariance matrix for the j-th class
\Lyf	$L\left(y, f\right)$	L(y, f), loss function
\Lxy	$L\left(y, f(\mathbf{x})\right)$	L(y, f(x)), loss function
\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 f 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 (without factor 1 / n
\riskeb	$\bar{\mathcal{R}}_{\text{emp}}$	R_emp, empirical risk with 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
\LS	$\mathfrak{L}$	???????????
\TS	$\mathfrak{T}$	?????????????
\errtrain	$\text{err}_{\text{train}}$	training error
\errtest	err_{test}	training error
\errexp	$\overline{\text{err}_{\text{test}}}$	training error