Machine Learning Notation

Shan-Hung Wu

1 Numbers & Arrays

a	A scalar (integer or real)
A	A scalar constant
$\boldsymbol{a}$	A vector
$\boldsymbol{A}$	A matrix
$\mathsf{\mathbf{A}}$	A tensor
$\boldsymbol{I}_n$	The $n \times n$ identity matrix
$\boldsymbol{D}$	A diagonal matrix
$\mathrm{diag}(\boldsymbol{a})$	A square, diagonal matrix with diagonal entries given by $\boldsymbol{a}$
a	A scalar random variable
$\mathbf{a}$	A vector-valued random variable
$\mathbf{A}$	A matrix-valued random variable

2 Sets & Graphs

$\mathbb{A}$	A set
$\mathbb{R}$	The set of real numbers
$\{0, 1\}$	The set containing 0 and 1
$\{0, 1, \cdots, n\}$	The set of all integers between 0 and n
$[a, b]$	The real interval including a and b
$(a, b]$	The real interval excluding a but including b
$\mathbb{A} \backslash \mathbb{B}$	Set subtraction, i.e., the set containing the elements of $\mathbb{A}$ that are not in $\mathbb{B}$
$\mathcal{G}$	A graph whose each vertex $\mathbf{x}^{(i)}$ denotes a random variable and edge denotes conditional dependency (directed) or correlation (undirected)
$\mathrm{Pa}(\mathbf{x}^{(i)})$	The parents of a vertex $\mathbf{x}^{(i)}$ in $\mathcal{G}$

3 Indexing

a_i	Element i of vector $\boldsymbol{a}$, with indexing starting at 1
a_{-i}	All elements of vector $\boldsymbol{a}$ except for element i
$A_{i,j}$	Element (i, j) of matrix $\boldsymbol{A}$
$\boldsymbol{A}_{i,:}$	Row i of matrix $\boldsymbol{A}$
$\boldsymbol{A}_{:,i}$	Column i of matrix $\boldsymbol{A}$
$\mathsf{A}_{i,j,k}$	Element (i, j, k) of a 3-D tensor $\mathsf{\mathbf{A}}$
$\mathsf{\mathbf{A}}_{:,:,i}$	2-D slice of a 3-D tensor
a_i	Element i of the random vector $\mathbf{a}$

4 Functions

$f : \mathbb{A} \to \mathbb{B}$	A function f with domain $\mathbb{A}$ and range $\mathbb{B}$
$f \circ g$	Composition of functions f and g
$f(\boldsymbol{x}; \boldsymbol{\theta})$	A function of $\boldsymbol{x}$ parametrized by $\boldsymbol{\theta}$ (with $\boldsymbol{\theta}$ omitted sometimes)
$\ln x$	Natural logarithm of x
$\sigma(x)$	Logistic sigmoid, i..e, $(1 + \exp(-x))^{-1}$
$\zeta(x)$	Softplus, $\ln(1 + \exp(x))$
$\\|\boldsymbol{x}\\|_p$	L^p norm of $\boldsymbol{x}$
$\\|\boldsymbol{x}\\|$	L^2 norm of $\boldsymbol{x}$
x^+	Positive part of x, i.e., $\max(0, x)$
$1(x; \mathrm{cond})$	The indicator function of x: 1 if the condition is true, 0 otherwise
$g[f; x]$	A functional that maps f to $f(x)$

Sometimes we use a function f whose argument is a scalar, but apply it to a vector, matrix, or tensor: $f(\boldsymbol{x})$, $f(\boldsymbol{X})$, or $f(\mathsf{\mathbf{X}})$. This means to apply f to the array element-wise. For example, if $\mathsf{\mathbf{C}} = \sigma(\mathsf{\mathbf{X}})$, then $\mathsf{C}_{\mathsf{i,j,k}} = \sigma(\mathsf{X}_{\mathsf{i,j,k}})$ for all i, j and k.

5 Calculus

$f'(a)$ or $\frac{df}{dx}(a)$	Derivative of $f : \mathbb{R} \to \mathbb{R}$ at input point a
$\frac{\partial f}{\partial x_i}(\boldsymbol{a})$	Partial derivative of $f : \mathbb{R}^n \to \mathbb{R}$ with respect to x_i at input $\boldsymbol{a}$
$\nabla f(\boldsymbol{a}) \in \mathbb{R}^n$	Gradient of $f : \mathbb{R}^n \to \mathbb{R}$ at input $\boldsymbol{a}$
$\nabla f(\boldsymbol{A}) \in \mathbb{R}^{m \times n}$	Matrix derivatives of $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ at input $\boldsymbol{A}$
$\nabla f(\mathsf{\mathbf{A}})$	Tensor derivatives of f at input $\mathsf{\mathbf{A}}$
$\boldsymbol{J}(\boldsymbol{f})(\boldsymbol{a}) \in \mathbb{R}^{m \times n}$	The Jacobian matrix of $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ at input $\boldsymbol{a}$
$\nabla^2 f(\boldsymbol{a})$ or $\boldsymbol{H}(f)(\boldsymbol{a}) \in \mathbb{R}^{n \times n}$	The Hessian matrix of $f : \mathbb{R}^n \to \mathbb{R}$ at input point $\boldsymbol{a}$
$\int f(\boldsymbol{x}) d\boldsymbol{x}$	Definite integral over the entire domain of $\boldsymbol{x}$
$\int_{\mathbb{S}} f(\boldsymbol{x}) d\boldsymbol{x}$	Definite integral with respect to $\boldsymbol{x}$ over the set $\mathbb{S}$

6 Linear Algebra

$\boldsymbol{A}^\top$	Transpose of matrix $\boldsymbol{A}$
$\boldsymbol{A}^\dagger$	Moore-Penrose pseudo-inverse of $\boldsymbol{A}$
$\boldsymbol{A} \odot \boldsymbol{B}$	Element-wise (Hadamard) product of $\boldsymbol{A}$ and $\boldsymbol{B}$
$\det(\boldsymbol{A})$	Determinant of $\boldsymbol{A}$
$\mathrm{tr}(\boldsymbol{A})$	Trace of $\boldsymbol{A}$
$\boldsymbol{e}^{(i)}$	The i-th standard basis vector (a one-hot vector)

7 Probability & Info. Theory

$\mathrm{a} \perp \mathrm{b}$	Random variables a and b are independent
$\mathrm{a} \perp \mathrm{b} \mid \mathrm{c}$	They are conditionally independent given c
$\Pr(a \mid b)$ or $\Pr(\mathrm{a} \mid \mathrm{b})$	Shorthand for the probability $\Pr(\mathrm{a} = a \mid \mathrm{b} = b)$
$P_{\mathrm{a}}(a)$	A probability mass function of the discrete random variable a
$p_{\mathrm{a}}(a)$	A probability density function of the continuous random variable a
$\mathrm{P}(\mathrm{a} = a)$	Either $P_{\mathrm{a}}(a)$ or $p_{\mathrm{a}}(a)$
$\mathrm{P}(\theta)$	A probability distribution parametrized by θ
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	The Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\mathrm{x} \sim \mathrm{P}(\theta)$	Random variable x has distribution P
$\mathrm{E}_{\mathrm{x} \sim \mathrm{P}}[\mathrm{f}(\mathrm{x})]$	Expectation of f(x) with respect to P
$\mathrm{Var}[\mathrm{f}(\mathrm{x})]$	Variance of f(x)
$\mathrm{Cov}[\mathrm{f}(\mathrm{x}), \mathrm{g}(\mathrm{x})]$	Covariance of f(x) and g(x)
$\mathrm{H}(\mathrm{x})$	Shannon entropy of the random variable x
$\mathrm{D}_{\mathrm{KL}}(\mathrm{P} \\| \mathrm{Q})$	Kullback-Leibler (KL) divergence from distribution Q to P

8 Machine Learning

$\mathbb{X}$	The set of training examples
N	Size of $\mathbb{X}$
$(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)})$	The i-th example pair in $\mathbb{X}$ (supervised learning)
$\boldsymbol{x}^{(i)}$	The i-th example in $\mathbb{X}$ (unsupervised learning)
D	Dimension of a data point $\boldsymbol{x}^{(i)}$
K	Dimension of a label $\boldsymbol{y}^{(i)}$
$\boldsymbol{X} \in \mathbb{R}^{N \times D}$	Design matrix, where $\boldsymbol{X}_{i,:}$ denotes $\boldsymbol{x}^{(i)}$
$P(\mathbf{x}, \mathbf{y})$	A data generating distribution
$\mathbb{F}$	Hypothesis space of functions to be learnt, i.e., a model
$C[f]$	A cost functional of $f \in \mathbb{F}$
$C(\theta)$	A cost function of θ parametrizing $f \in \mathbb{F}$
$(\boldsymbol{x}', \boldsymbol{y}')$	A testing pair
$\hat{\boldsymbol{y}}$	Label predicted by a function f, i.e., $\hat{\boldsymbol{y}} = f(\boldsymbol{x}')$ (supervised learning)

9 Typesetting

Section*	Section that can be skipped for the first time reading
Section**	Section for reference only (will not be taught)
[Proof]	Prove it yourself
[Homework]	You have homework

3 Mathematical Foundations

Symbol	Meaning
$\lfloor x \rfloor$	Floor of x, i.e., round down to nearest integer
$\lceil x \rceil$	Ceiling of x, i.e., round up to nearest integer
$\vec{x} \otimes \vec{y}$	Convolution of $\vec{x}$ and $\vec{y}$
$\vec{x} \odot \vec{y}$	Hadamard (elementwise) product of $\vec{x}$ and $\vec{y}$
$a \wedge b$	logical AND
$a \vee b$	logical OR
$\neg a$	logical NOT
$\mathbb{I}(x)$	Indicator function, $\mathbb{I}(x) = 1$ if x is true, else $\mathbb{I}(x) = 0$
∞	Infinity
$\rightarrow$	Tends towards, e.g., $n \rightarrow \infty$
$\leftarrow$	in an algorithm: assign to variable t the new value $t+1$, e.g., $t \leftarrow t+1$
$\propto$	Proportional to, so $y = ax$ can be written as $y \propto x$
$\|x\|$	Absolute value
$\|\mathcal{S}\|$	Size (cardinality) of a set
$n!$	Factorial function
∇	Vector of first derivatives
∇^2	Hessian matrix of second derivatives
$\triangleq$	Defined as
$O(\cdot)$	Big-O: roughly means order of magnitude
$\mathbb{R}$	The real numbers
$1:n$	Range (Matlab convention): $1:n = 1, 2, ..., n$
$\approx$	Approximately equal to
$\arg\max\limits_x f(x)$	Argmax: the value x that maximizes f
$\arg\min\limits_x f(x)$	Argmin: the value x that minimizes f
$B(a,b)$	Beta function, $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$
$B(\vec{\alpha})$	Multivariate beta function, $\dfrac{\prod\limits_k \Gamma(\alpha_k)}{\Gamma(\sum\limits_k \alpha_k)}$
$n!$	n factorial $= n*(n-1)*(n-2)*...*1$
$\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$	n choose k , equal to $n!/(k!(nk)!)$
$\delta(x)$	Dirac delta function,$\delta(x) = \infty$ if $x = 0$, else $\delta(x) = 0$
$\Gamma(x)$	Gamma function, $\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}\mathrm{d}u$
$\Psi(x)$	Digamma function,$Psi(x) = \dfrac{d}{dx}\log\Gamma(x)$
$\mathcal{X}$	A set from which values are drawn (e.g.,$\mathcal{X} = \mathbb{R}^D$)
$\equiv$	equivalent to (or defined to be)
$\lim\limits_{a\rightarrow\infty} f(x)$	the value of $f(x)$ in the limit as x approaches a
$m \mod n$	m modulo n, the remainder when m is divided by n (e.g. $7mod5 = 2$)
$\ln$	logarithm base e, or natural logarithm of x
$\log$	logarithm base 10 of x
$\log_2$	logarithm base 2 of x
$\exp(x)$ or e^x	exponential of x, i.e., e raised the power of x
$\partial f(x)/\partial x$	partial derivative of f with respect to x
$\int_a^b f(x)dx$	the integral of $f(x)$ between a and b. If no limits are written, the full space is assumed.
$F(X;\theta)$	function of x, with implied dependence upon θ
$< x >$	expected value of random variable x
$\bar{x}$	mean or averrage value of x
$\mathcal{E}[f(x)]$	the expected value of function $f(x)$ where x is a random variable
$\mathcal{E}_y[f(x,y)]$	the expected value of function over several variables, $f(x)$, taken over a subset y of them
$\sum_{i=1}^n a_i$	the sum from $i = 1$ to $n : a_1 + a_2 + ... + a_n$
$\prod_{i=1}^n a_i$	the product from $i = 1$ to $n : a_1 * a_2 * ... * a_n$

$f(x) * g(x)$	convolution of $f(x)$ with $g(x)$
$\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, ...$	"Calligraphic" font generally denotes sets or lists, e.g., data set $\mathcal{D} = x_1, ..., x_n$
$x \in \mathcal{D}$	x is an element of set $\mathcal{D}$
$x \notin \mathcal{D}$	x is not an element of set $\mathcal{D}$
$\mathcal{D} \cup \mathcal{D}$	x union of two sets, i.e., the set containing all elements of $\mathcal{D}$ and $\mathcal{D}$
$\lvert\mathcal{D}\rvert$	cardinality of set $\mathcal{D}$, i.e., the number of (possibly non-distinct) elements in it
$\max_x[\mathcal{D}]$	the maximum x value in set $\mathcal{D}$
$dom(x)$	Domain of variable x
$x = x$	The variable x is in the state x
$dim(x)$	For a discrete variable x, this denotes the number of states x can take
$x_{a:b}$	$x_a, x_{a+1}, ..., x_b$
$\not<, \not>$	not less than; not greater than
$\neq$	not equal to
$\ll, \gg$	much less than; much greater than
d/dx	the derivative with respect to x
$\mathcal{M} \subset \mathcal{N}$	$\mathcal{M}$ is a subset of $\mathcal{N}$
$\mathcal{M} \supset \mathcal{N}$	$\mathcal{M}$ contains $\mathcal{N}$
$\mathcal{M} \bigcap \mathcal{N}$	intersection of $\mathcal{M}$ and $\mathcal{N}$
$\Longrightarrow$	implies
$\longleftrightarrow$	equivalent to
$\exists$	there exists
$\forall$	for every

4 Linear algebra notations

We use boldface lower-case to denote vectors, such as $\vec{x}$, and boldface upper-case to denote matrices, such as $\vec{X}$. We denote entries in a matrix by non-bold upper case letters, such as X_{ij}.

Vectors are assumed to be column vectors, unless noted otherwise. We use $(x_1, \cdots, x_D)$ to denote a column vector created by stacking D scalars. If we write $\vec{X} = (\vec{x}_1, \cdots, \vec{x}_n)$, where the left hand side is a matrix, we mean to stack the $\vec{x}_i$ along the columns, creating a matrix.

Symbol	**Meaning**
$\vec{X} \succ 0$	$\vec{X}$ is a positive definite matrix
$tr(\vec{X})$	Trace of a matrix
$det(\vec{X})$	Determinant of matrix $\vec{X}$
$\lvert\vec{X}\rvert$	Determinant of matrix $\vec{X}$
$\vec{X}^{-1}$	Inverse of a matrix
$\vec{X}^{\dagger}$	Pseudo-inverse of a matrix
$\vec{X}^T$	Transpose of a matrix
$\vec{x}^T$	Transpose of a vector
$\text{diag}(x)$	Diagonal matrix made from vector $\vec{x}$
$\text{diag}(X)$	Diagonal vector extracted from matrix $\vec{X}$
$\vec{I}$ or $\vec{I}_d$	Identity matrix of size $d \times d$ (ones on diagonal, zeros of)
$\vec{1}$ or $\vec{1}_d$	Vector of ones (of length d)
$\vec{0}$ or $\vec{0}_d$	Vector of zeros (of length d)
$\lvert\lvert\vec{x}\rvert\rvert = \lvert\lvert\vec{x}\rvert\rvert_2$	Euclidean or ℓ_2 norm $\sqrt{\sum_{j=1}^{d} x_j^2}$
$\lvert\lvert\vec{x}\rvert\rvert_1$	ℓ_1 norm $\sum_{j=1}^{d} \lvert x_j\rvert$
$\vec{X}_{:,j}$	j'th column of matrix
$\vec{X}_{i,:}$	transpose of i'th row of matrix (a column vector)
$\vec{X}_{i,j}$	Element (i,j) of matrix $\vec{X}$

$\vec{x} \otimes \vec{y}$	Tensor product of $\vec{x}$ and $\vec{y}$
R^d	d-dimensional Euclidean space
x, A, ...	boldface is used for (column) vectors and matrices
$f(x)$	vector-valued function (note the boldface) of a scalars
$f(\chi)$	vector-valued function (note the boldface) of a vector
I	identity matrix, square matrix having 1s on the diagonal and 0 everywhere else
$\sum$	covariance matrix
λ	eigenvalue
$\mathbf{e}$	eigenvector
$\mathbf{u}_i$	unit vector in the ith direction in Euclidean space
$dim\ x$	The dimension of vector/matrix x

5 Probability notations

We denote random and fixed scalars by lower case, random and fixed vectors by bold lower case, and random and fixed matrices by bold upper case. Occasionally we use non-bold upper case to denote scalar random variables. Also, we use $p()$ for both discrete and continuous random variables

Symbol	Meaning
X, Y	Random variable
$P()$	Probability of a random event
$F()$	Cumulative distribution function(CDF), also called distribution function
$p(x)$	Probability mass function(PMF)
$f(x)$	probability density function(PDF)
$F(x, y)$	Joint CDF
$p(x, y)$	Joint PMF
$f(x, y)$	Joint PDF
$p(X\|Y)$	Conditional PMF, also called conditional probability
$f_{X\|Y}(x\|y)$	Conditional PDF
$X \perp Y$	X is independent of Y
$X \not\perp Y$	X is not independent of Y
$X \perp Y\|Z$	X is conditionally independent of Y given Z
$X \not\perp Y\|Z$	X is not conditionally independent of Y given Z
$X \sim p$	X is distributed according to distribution p
$\vec{\alpha}$	Parameters of a Beta or Dirichlet distribution
$\mathrm{cov}[X]$	Covariance of X
$\mathbb{E}[X]$	Expected value of X
$\mathbb{E}_q[X]$	Expected value of X wrt distribution q
$\mathbb{H}(X)$ or $\mathbb{H}(p)$	Entropy of distribution $p(X)$
$\mathbb{I}(X; Y)$	Mutual information between X and Y
$\mathbb{KL}(p\|\|q)$	KL divergence from distribution p to q
$\ell(\vec{\theta})$	Log-likelihood function
$L(\theta, a)$	Loss function for taking action a when true state of nature is θ
λ	Precision (inverse variance) $\lambda = 1/\sigma^2$
Λ	Precision matrix $\Lambda = \Sigma^{-1}$
$\mathrm{mode}[\vec{X}]$	Most probable value of $\vec{X}$
μ	Mean of a scalar distribution
$\vec{\mu}$	Mean of a multivariate distribution
Φ	cdf of standard normal
ϕ	pdf of standard normal
$\vec{\pi}$	multinomial parameter vector, Stationary distribution of Markov chain
ρ	Correlation coefficient
$\mathrm{sigm}(x)$	Sigmoid (logistic) function, $\frac{1}{1+e^{-x}}$
σ^2	Variance
Σ	Covariance matrix

var[x]	Variance of x
ν	Degrees of freedom parameter
Z	Normalization constant of a probability distribution
~	has the distribution, e.g., $p(x)\tilde{}N(\mu, \sigma^2)$
$N(\mu, \sigma^2)$	multidimensional normal or Gaussian distribution with mean μ and variance σ^2
$O(h(x))$	big oh order of $h(x)$
$\Theta(h(x))$	big theta order of $h(x)$
$\Omega(h(x))$	big omega order of $h(x)$
$\sup\limits_{x} f(x)$	the supremum value of $f(x)$-the global maximum of $f(x)$ over all values of x
$p(x = tr)$	Probability of variable x being in the state true
$p(x = fa)$	Probability of variable x being in the state false
$p(x \cap y)$	Probability of x and y
$p(x \cup y)$	Probability of x or y
$p(x\|\|y)$	Probability of x contioned on y
$\langle f(x) \rangle_{g(x)}$	The average of the function $f(x)$ with respect to the distribtioin $p(x)$
$\sigma(x)$	The logistic sigmoid $\frac{1}{(1+\exp(-x))}$
$erf(x)$	The (Gaussian) error function

6 Specific Machine learning notations

We use upper case letters to denote constants, such as C, K, M, N, T, etc. We use lower case letters as dummy indexes of the appropriate range, such as $c = 1 : C$ to index classes, $i = 1 : M$ to index data cases, $j = 1 : N$ to index input features, $k = 1 : K$ to index states or clusters, $t = 1 : T$ to index time, etc.

We use x to represent an observed data vector. In a supervised problem, we use y or $\vec{y}$ to represent the desired output label. We use $\vec{z}$ to represent a hidden variable. Sometimes we also use q to represent a hidden discrete variable.

We use uppercase bold roman letters to denote matrices $\mathbf{M}$

Symbol	**Meaning**
C	Number of classes
D	Dimensionality of data vector (number of features - of a feature vector gained)
N	Number of data cases
N_c	Number of examples of class c,$N_c = \sum_{i=1}^{N} \mathbb{I}(y_i = c)$
R	Number of outputs (response variables)
$\mathcal{D}$	Training data $\mathcal{D} = \{(\vec{x}_i, y_i) \| i = 1 : N\}$
$\mathcal{D}_{test}$	Test data
$\mathcal{X}$	Input space
$\mathcal{Y}$	Output space
K	Number of states or dimensions of a variable (often latent)
$k(x, y)$	Kernel function
$\vec{K}$	Kernel matrix
$\mathcal{H}$	Hypothesis space
L	Loss function
$J(\vec{\theta})$	Cost function
$f(\vec{x})$	Decision function
$P(y\|\vec{x})$	TODO
λ	Strength of ℓ_2 or $\ell_1 regularizer$
$\phi(x)$	Basis function expansion of feature vector $\vec{x}$
Φ	Basis function expansion of design matrix $\vec{X}$
$q()$	Approximate or proposal distribution
$Q(\vec{\theta}, \vec{\theta}_{old})$	Auxiliary function in EM
T	Length of a sequence
$T(\mathcal{D})$	Test statistic for data

$\vec{T}$	Transition matrix of Markov chain
$\vec{\theta}$	Parameter vector
$\vec{\theta}^{(s)}$	s'th sample of parameter vector
$\hat{\vec{\theta}}$	Estimate (usually MLE or MAP) of $\vec{\theta}$
$\hat{\vec{\theta}}_{MLE}$	Maximum likelihood estimate of $\vec{\theta}$
$\hat{\vec{\theta}}_{MAP}$	MAP estimate of $\vec{\theta}$
$\bar{\vec{\theta}}$	Estimate (usually posterior mean) of $\vec{\theta}$
$\vec{w}$	Vector of regression weights (called $\vec{\beta}$ in statistics)
b	intercept (called ε in statistics)
$\vec{W}$	Matrix of regression weights
x_{ij}	Component (i.e., feature) j of data case i ,for $i = 1 : N, j = 1 : D$
$\vec{x}_i$	Training case, $i = 1 : N$
$\vec{X}$	Design matrix of size $N \times D$
$\bar{\vec{x}}$	Empirical mean $\bar{\vec{x}} = \dfrac{1}{N} \sum_{i=1}^{N} \vec{x}_i$
$\tilde{\vec{x}}$	Future test case
$\vec{x}_*$	Feature test case
$\vec{y}$	Vector of all training labels $\vec{y} = (y_1, ..., y_N)$
z_{ij}	Latent component j for case i
S	Number of samples

7 Graphical model notations

In graphical models, we index nodes by $s, t, u \in V$, and states by $i, j, k \in \mathcal{X}$.

Symbol	Meaning
$\tilde{st}$	Node s is connected to node t
bel	Belief function
$\mathcal{C}$	Cliques of a graph
ch_j	Child of node j in a DAG (directed acyclic graph)
$desc_j$	Descendants of node j in a DAG
G	A graph
$\mathcal{E}$	Edges of a graph
mb_t	Markov blanket of node t
nbd_t	Neighborhood of node t
pa_t	Parents of node t in a DAG
$pred_t$	Predecessors of node t in a Direct Acyclic Graph (DAG) with respect to some ordering
$\psi_c(x_c)$	Potential function for clique c
$\mathcal{S}$	Separators of a graph
θ_{sjk}	prob. node s is in state k given its parents are in states j
$\mathcal{V}$	Nodes of a graph
$pa(x)$	The parents of x
$ch(x)$	The children of x
$ne(x)$	The neighbours of x
$\tilde{ij}$	The set of unique neighbouring edges on a graph

8 Abbreviations (incomplete)

cdf ... cumulative distribution function DAG ... directed acyclic graph HMM ... Hidden Markov Model iff ... if and only if pmf ... probability mass function