Probabilistic Graphical Models

Principles and Techniques

Daphne Koller

Nir Friedman

The MIT Press
Cambridge, Massachusetts
London, England

This book was set by the authors in LaTeX2$_\epsilon$.
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Koller, Daphne.
Probabilistic Graphical Models: Principles and Techniques / Daphne Koller and Nir Friedman.
p. cm. – (Adaptive computation and machine learning)
Includes bibliographical references and index.
ISBN 978-0-262-01319-2 (hardcover : alk. paper)
1. Graphical modeling (Statistics) 2. Bayesian statistical decision theory—Graphic methods. I. Koller, Daphne. II. Friedman, Nir.
QA279.5.K65 2010
519.5'420285–dc22
2009008615

10 9 8 7 6 5 4 3 2 1

Notation Index

$|A|$ — Cardinality of the set A, 20
$\phi_1 \times \phi_2$ — Factor product, 107
$\gamma_1 \bigoplus \gamma_2$ — Joint factor combination, 1102
$p(\boldsymbol{Z}) \bigoplus g(\boldsymbol{Z})$ — Marginal of $g(\boldsymbol{Z})$ based on $p(\boldsymbol{Z})$, 631
$\sum_Y \phi$ — Factor marginalization, 297
$X \rightleftharpoons Y$ — Bi-directional edge, 34
$X \rightarrow Y$ — Directed edge, 34
$X — Y$ — Undirected edge, 34
$X \leftrightarrow Y$ — Non-ancestor edge (PAGs), 1048
$X \circ\!\!\rightarrow Y$ — Ancestor edge (PAGs), 1048
$\langle x, y \rangle$ — Inner product of vectors x and y, 262
$\|P - Q\|_1$ — L_1 distance, 1141
$\|P - Q\|_2$ — L_2 distance, 1141
$\|P - Q\|_\infty$ — L_∞ distance, 1141
$(\boldsymbol{X} \perp \boldsymbol{Y})$ — Independence of random variables, 24
$(\boldsymbol{X} \perp \boldsymbol{Y} \mid \boldsymbol{Z})$ — Conditional independence of random variables, 24
$(\boldsymbol{X} \perp_c \boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{c})$ — Context-specific independence, 162

$\boldsymbol{I}\{\cdot\}$ — Indicator function, 32

$\mathcal{A}(\boldsymbol{x} \rightarrow \boldsymbol{x}')$ — Acceptance probability, 517
$\aleph$ — Template attributes, 214
$\alpha(A)$ — The argument signature of attribute A, 213
$\mathit{Ancestors}_X$ — Ancestors of X (in graph), 36
argmax, 26
A — A template attribute, 213

$\mathit{Beta}(\alpha_1, \alpha_0)$ — Beta distribution, 735
β_i — Belief potential, 352
$\mathcal{B}_{\mathcal{I}[\sigma]}$ — Induced Bayesian network, 1091
$\mathcal{B}$ — Bayesian network, 62
$\mathcal{B}_0$ — Initial Bayesian network (DBN), 204
$\mathcal{B}_\rightarrow$ — Transition Bayesian network (DBN), 204
$\mathcal{B}_{\boldsymbol{Z}=\boldsymbol{z}}$ — Mutilated Bayesian network, 499

$\mathcal{C}(K, \boldsymbol{h}, g)$ — Canonical form, 609
$\mathcal{C}(\boldsymbol{X}; K, \boldsymbol{h}, g)$ — Canonical form, 609
$\mathcal{C}[v]$ — Choices, 1083
Ch_X — Children of X (in graph), 34
$\boldsymbol{C}_i$ — Clique, 346
$\boldsymbol{x} \sim \boldsymbol{c}$ — Compatability of values , 20
$\mathrm{cont}(\gamma)$ — Joint factor contraction, 1102
$\boldsymbol{Cov}[X; Y]$ — Covariance of X and Y, 248

$\boldsymbol{D}$ — A subclique, 104
Δ — Discrete variables (hybrid models), 605
$\boldsymbol{d}$ — Value of a subclique, 104
$\mathcal{D}^+$ — Complete data, 871
$\mathcal{D}$ — Empirical samples (data), 698
$\mathcal{D}$ — Sampled data, 489
$\mathcal{D}^*$ — Complete data, 912
$\mathcal{D}$ — Decisions, 1087
$\mathrm{Descendants}_X$ — Descendants of X (in graph), 36
$\tilde{\delta}_{i \rightarrow j}$ — Approximate sum-product message, 435
$\delta_{i \rightarrow j}$ — Sum-product message, 352
$\mathrm{Dim}[\mathcal{G}]$ — Dimension of a graph, 801
$\mathit{Dirichlet}(\alpha_1, \ldots, \alpha_K)$ — Dirichlet distribution, 738
$\boldsymbol{D}(P \| Q)$ — Relative entropy, 1139
$\boldsymbol{D}_{\mathit{var}}(P; Q)$ — Variational distance, 1141
$\mathbf{Down}^*(r)$ — Downward closure, 422
$\mathbf{Down}^+(r)$ — Extended downward closure, 422
$\mathbf{Down}(r)$ — Downward regions, 422
$\mathit{do}(Z := z), \mathit{do}(z)$ — Intervention, 1010
$\mathit{d\text{-}sep}_{\mathcal{G}}(\boldsymbol{X}; \boldsymbol{Y} \mid \boldsymbol{Z})$ — d-separation, 71

$\mathcal{E}$ — Edges in MRF, 127