Bayesian Reasoning and Machine Learning

David Barber

University College London



CONTENTS

Preface					3.3.1 Conditional independence	
List of notation					3.3.2 The impact of collisions	
BRMLTOOLBOX					3.3.3 Graphical path manipulations for independence	
ı	Inference in probabilistic models				3.3.4 d-separation	
_					3.3.5 Graphical and distributional	
1	Probabilistic reasoning	3			in/dependence	
•	1.1 Probability refresher				3.3.6 Markov equivalence in belief	
	1.1.1 Interpreting conditional				networks	
	probability				3.3.7 Belief networks have limited	
	1.1.2 Probability tables				expressibility	
	1.2 Probabilistic reasoning			3.4	Causality	
	1.3 Prior, likelihood and posterior				3.4.1 Simpson's paradox	
	1.3.1 Two dice: what were the				3.4.2 The do-calculus	
	individual scores?				3.4.3 Influence diagrams and the do-calculus	
	1.4 Summary			2.5	Summary	
	1.5 Code				Code	
	1.6 Exercises				Exercises	
				3.1	Exercises	
2	Basic graph concepts	22	4	Gra	phical models	58
2	2.1 Graphs	22	4	4.1	Graphical models	58
2	2.1 Graphs2.2 Numerically encoding graphs	22	4	4.1	-	58
2	2.1 Graphs2.2 Numerically encoding graphs2.2.1 Edge list	22	4	4.1	Graphical models	58
2	2.1 Graphs2.2 Numerically encoding graphs2.2.1 Edge list2.2.2 Adjacency matrix	22	4	4.1	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields	58
2	 2.1 Graphs 2.2 Numerically encoding graphs 2.2.1 Edge list 2.2.2 Adjacency matrix 2.2.3 Clique matrix 	22	4	4.1	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields 4.2.3 Hammersley-Clifford theorem	58
2	 2.1 Graphs 2.2 Numerically encoding graphs 2.2.1 Edge list 2.2.2 Adjacency matrix 2.2.3 Clique matrix 2.3 Summary 	22	4	4.1	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields 4.2.3 Hammersley-Clifford theorem 4.2.4 Conditional independence	58
2	 2.1 Graphs 2.2 Numerically encoding graphs 2.2.1 Edge list 2.2.2 Adjacency matrix 2.2.3 Clique matrix 2.3 Summary 2.4 Code 	22	4	4.1	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields 4.2.3 Hammersley–Clifford theorem 4.2.4 Conditional independence using Markov networks	58
2	 2.1 Graphs 2.2 Numerically encoding graphs 2.2.1 Edge list 2.2.2 Adjacency matrix 2.2.3 Clique matrix 2.3 Summary 	22	4	4.1 4.2	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields 4.2.3 Hammersley-Clifford theorem 4.2.4 Conditional independence using Markov networks 4.2.5 Lattice models	58
	 2.1 Graphs 2.2 Numerically encoding graphs 2.2.1 Edge list 2.2.2 Adjacency matrix 2.2.3 Clique matrix 2.3 Summary 2.4 Code 2.5 Exercises 		4	4.1 4.2 4.3	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields 4.2.3 Hammersley-Clifford theorem 4.2.4 Conditional independence using Markov networks 4.2.5 Lattice models Chain graphical models	58
3	 2.1 Graphs 2.2 Numerically encoding graphs 2.2.1 Edge list 2.2.2 Adjacency matrix 2.2.3 Clique matrix 2.3 Summary 2.4 Code 2.5 Exercises Belief networks	22	4	4.1 4.2 4.3	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields 4.2.3 Hammersley-Clifford theorem 4.2.4 Conditional independence using Markov networks 4.2.5 Lattice models Chain graphical models Factor graphs	58
	 2.1 Graphs 2.2 Numerically encoding graphs 2.2.1 Edge list 2.2.2 Adjacency matrix 2.2.3 Clique matrix 2.3 Summary 2.4 Code 2.5 Exercises Belief networks 3.1 The benefits of structure 		4	4.1 4.2 4.3	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields 4.2.3 Hammersley-Clifford theorem 4.2.4 Conditional independence using Markov networks 4.2.5 Lattice models Chain graphical models Factor graphs 4.4.1 Conditional independence in	58
	 2.1 Graphs 2.2 Numerically encoding graphs 2.2.1 Edge list 2.2.2 Adjacency matrix 2.2.3 Clique matrix 2.3 Summary 2.4 Code 2.5 Exercises Belief networks 3.1 The benefits of structure 3.1.1 Modelling independencies 		4	4.1 4.2 4.3 4.4	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields 4.2.3 Hammersley-Clifford theorem 4.2.4 Conditional independence using Markov networks 4.2.5 Lattice models Chain graphical models Factor graphs 4.4.1 Conditional independence in factor graphs	58
	 2.1 Graphs 2.2 Numerically encoding graphs 2.2.1 Edge list 2.2.2 Adjacency matrix 2.2.3 Clique matrix 2.3 Summary 2.4 Code 2.5 Exercises Belief networks 3.1 The benefits of structure 3.1.1 Modelling independencies 3.1.2 Reducing the burden of 		4	4.1 4.2 4.3 4.4	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields 4.2.3 Hammersley-Clifford theorem 4.2.4 Conditional independence using Markov networks 4.2.5 Lattice models Chain graphical models Factor graphs 4.4.1 Conditional independence in factor graphs Expressiveness of graphical	58
	 2.1 Graphs 2.2 Numerically encoding graphs 2.2.1 Edge list 2.2.2 Adjacency matrix 2.2.3 Clique matrix 2.3 Summary 2.4 Code 2.5 Exercises Belief networks 3.1 The benefits of structure 3.1.1 Modelling independencies 3.1.2 Reducing the burden of specification 		4	4.1 4.2 4.3 4.4	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields 4.2.3 Hammersley-Clifford theorem 4.2.4 Conditional independence using Markov networks 4.2.5 Lattice models Chain graphical models Factor graphs 4.4.1 Conditional independence in factor graphs Expressiveness of graphical models	58
	 2.1 Graphs 2.2 Numerically encoding graphs 2.2.1 Edge list 2.2.2 Adjacency matrix 2.2.3 Clique matrix 2.3 Summary 2.4 Code 2.5 Exercises Belief networks 3.1 The benefits of structure 3.1.1 Modelling independencies 3.1.2 Reducing the burden of specification 3.2 Uncertain and unreliable evidence		4	4.1 4.2 4.3 4.4 4.5 4.6	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields 4.2.3 Hammersley-Clifford theorem 4.2.4 Conditional independence using Markov networks 4.2.5 Lattice models Chain graphical models Factor graphs 4.4.1 Conditional independence in factor graphs Expressiveness of graphical models Summary	58
	 2.1 Graphs 2.2 Numerically encoding graphs 2.2.1 Edge list 2.2.2 Adjacency matrix 2.2.3 Clique matrix 2.3 Summary 2.4 Code 2.5 Exercises Belief networks 3.1 The benefits of structure 3.1.1 Modelling independencies 3.1.2 Reducing the burden of specification 		4	4.1 4.2 4.3 4.4 4.5 4.6 4.7	Graphical models Markov networks 4.2.1 Markov properties 4.2.2 Markov random fields 4.2.3 Hammersley-Clifford theorem 4.2.4 Conditional independence using Markov networks 4.2.5 Lattice models Chain graphical models Factor graphs 4.4.1 Conditional independence in factor graphs Expressiveness of graphical models	58

5

5	Effi	cient i	inference in trees	77			6.6.2	1 &	
	5.1	Margi	nal inference					normalisation constant of a	
		5.1.1	Variable elimination in a					distribution	
			Markov chain and message				6.6.3	The marginal likelihood	
			passing				6.6.4	*	
		5.1.2	The sum-product algorithm				6.6.5	Shafer-Shenoy propagation	
			on factor graphs			6.7	Findi	ng the most likely state	
		5.1.3	Dealing with evidence			6.8	Reabs	sorption: converting a	
		5.1.4	Computing the marginal				juncti	on tree to a directed network	
			likelihood			6.9	The n	need for approximations	
		5.1.5	The problem with loops				6.9.1	Bounded width junction trees	
	5.2	Other	forms of inference			6.10	Sumr	nary	
		5.2.1	Max-product				Code		
		5.2.2	Finding the N most probabl	le states			Exerc		
		5.2.3	Most probable path and			0.12	LACIC		
			shortest path		7	Mak	ina da	ecisions	127
		5.2.4	Mixed inference		•		-	cted utility	127
	5.3	Infere	ence in multiply connected	graphs		7.1	-	Utility of money	
		5.3.1	Bucket elimination			7.0		sion trees	
		5.3.2	Loop-cut conditioning						
	5.4	Mess	age passing for continuous	}		1.3		nding Bayesian networks	
		distri	butions					ecisions	
	5.5	Sumr	nary				7.3.1	Syntax of influence	
		Code	•				a	diagrams	
	5.7	Exerc	cises			7.4		ng influence diagrams	
								Messages on an ID	
6	The	iun <i>c</i> i	tion tree algorithm	102				Using a junction tree	
•		-	ering variables			7.5		ov decision processes	
	0.1		Reparameterisation				7.5.1	Maximising expected	
	6.2		ie graphs				7.50	utility by message passing	
	0.2		Absorption			7.		Bellman's equation	
			Absorption schedule on			7.6	-	orally unbounded MDPs	
		0,2,2	clique trees					Value iteration	
	63	Lunct	ion trees					Policy iteration	
	0.5	6.3.1	The running intersection			~ ~		A curse of dimensionality	
		0.511	property			7.7		tional inference and	
	6.4	Cons	tructing a junction tree for			_ ^	plann	•	
	0.1		y connected distributions			7.8		icial matters	
		6.4.1	Moralisation				7.8.1	Options pricing and	
		6.4.2	Forming the clique graph					expected utility	
		6.4.3		m			7.8.2	Binomial options pricing	
		0.4.5	a clique graph	11			~ ~ ~	model	
		6.4.4	Assigning potentials to			7.0	7.8.3	Optimal investment	
		0.1.1	cliques			7.9		er topics	
	6.5	Lunct	ion trees for multiply				7.9.1	Partially observable MDPs	
	0.5		ected distributions				7.9.2	Reinforcement learning	
			Triangulation algorithms				Sumr	-	
	6.6		unction tree algorithm				Code		
	0.0		Remarks on the ITA			7.12	Exerc	eises	

Contents

	Lea	rning	in probabilistic models				9.4.2	Learning binary variable	
8	Stat	istics	for machine learning	165			9.4.3	tables using a Beta prior Learning multivariate	
_			esenting data	.05				discrete tables using a	
		_	Categorical					Dirichlet prior	
		8.1.2	Ordinal			9.5	Struc	ture learning	
		8.1.3	Numerical				9.5.1	PC algorithm	
	8.2		butions				9.5.2	Empirical independence	
			The Kullback-Leibler				9.5.3	Network scoring	
			divergence $KL(q p)$				9.5.4	Chow-Liu trees	
		8.2.2				9.6	Maxi	mum likelihood for	
	8.3		ical distributions				undir	ected models	
			variate Gaussian				9.6.1	The likelihood gradient	
	0		Completing the square				9.6.2	General tabular clique	
			Conditioning as system					potentials	
		٠ـ	reversal				9.6.3	Decomposable Markov	
		8.4.3	Whitening and centring					networks	
	8.5		nential family				9.6.4	Exponential form potentials	
	0.0	8.5.1	Conjugate priors				9.6.5	Conditional random fields	
	8.6		ning distributions				9.6.6	Pseudo likelihood	
			erties of maximum likelihood				9.6.7	Learning the structure	
	0.7	8.7.1	Training assuming the			9.7	Sumn	nary	
		0.7.1	correct model class			9.8	Code		
		8.7.2				9.9	Exerc	ises	
		0.7.2	model is incorrect						
		8.7.3	Maximum likelihood and		10	Naiv	e Bay	es	243
			the empirical distribution				-	Bayes and conditional	
	8.8	Learr	ning a Gaussian					endence	
	-	8.8.1	_			10.2		ation using maximum	
		8.8.2	Bayesian inference of the			10.2	likelil		
			mean and variance					Binary attributes	
		8.8.3	Gauss-gamma distribution					Multi-state variables	
	8.9	Sumr	_					Text classification	
		Code				10.3		sian naive Bayes	
		Exerc	ricac					augmented naive Bayes	
	0.11	DACIC	1303			10.4		Learning tree augmented	
a	Loor	nina -	s informes	100			10.4.1	naive Bayes networks	
9		_	is inference	199		10.5	Sumn		
	9.1		ing as inference				Code	itat y	
			Learning the bias of a coin				Exerc	iaaa	
		9.1.2	Making decisions			10.7	Exerc	ases	
		9.1.3	A continuum of parameters Decisions based on		4.4	1	!	مما ما	25.0
		9.1.4	continuous intervals		11		_	vith hidden variables	256
	0.2	D				11.1		en variables and missing	
		_	sian methods and ML-II				data		
	9.3		mum likelihood training				11.1.1	Why hidden/missing	
	_		lief networks					variables can complicate	
	9.4	•	sian belief network training					proceedings	
		9.4.1	Global and local parameter				11.1.2	The missing at random	
			independence					assumption	

	1	1.1.3 Maximum likelihood			12.7	Summary	
	1	1.1.4 Identifiability issues			12.8	Code	
	11.2 E	Expectation maximisation			12.9	Exercises	
		1.2.1 Variational EM					
	1	1.2.2 Classical EM		Ш	Mad	chine learning	
	1	1.2.3 Application to belief networks					
	1	1.2.4 General case		13	Maci	hine learning concepts	305
	1	1.2.5 Convergence		, ,		Styles of learning	
	1	1.2.6 Application to Markov			13.1	13.1.1 Supervised learning	
		networks				13.1.2 Unsupervised learning	
	11.3 E	Extensions of EM				13.1.3 Anomaly detection	
	1	1.3.1 Partial M-step				13.1.4 Online (sequential) learning	
	1	1.3.2 Partial E-step				13.1.5 Interacting with the	
	11.4	A failure case for EM				environment	
	11.5 V	Variational Bayes				13.1.6 Semi-supervised learning	
		11.5.1 EM is a special case of			13.2	Supervised learning	
		variational Bayes			13.2	13.2.1 Utility and loss	
	1	11.5.2 An example: VB for the				13.2.2 Using the empirical	
		Asbestos-Smoking-Cancer				distribution	
		network				13.2.3 Bayesian decision approach	
	11.6 (Optimising the likelihood by			122	Bayes versus empirical decisions	
		gradient methods				Summary	
		11.6.1 Undirected models					
		Summary			13.5	Exercises	
	11.8	-					227
		Exercises		14		rest neighbour classification	322
	11.7	LACICISCS				Do as your neighbour does	
12	Dayor	sian model coloction	284			K-nearest neighbours	
12	-	sian model selection	204		14.3	A probabilistic interpretation of	
		Comparing models the				nearest neighbours	
		Bayesian way				14.3.1 When your nearest	
		Illustrations: coin tossing				neighbour is far away	
		12.2.1 A discrete parameter space			14.4	Summary	
		12.2.2 A continuous parameter			14.5	Code	
		space			14.6	Exercises	
		Occam's razor and Bayesian					
		complexity penalisation		15	Unsi	upervised linear dimension	
	12.4	A continuous example: curve				ıction	329
		fitting				High-dimensional spaces –	J-3
	12.5	Approximating the model			13.1	low-dimensional manifolds	
		likelihood			150		
		12.5.1 Laplace's method			13.2	Principal components analysis	
		12.5.2 Bayes information criterion				15.2.1 Deriving the optimal linear	
	12.6	Bayesian hypothesis testing for				reconstruction	
		outcome analysis				15.2.2 Maximum variance	
		12.6.1 Outcome analysis				criterion 15.2.3 PCA algorithm	
		12.6.2 H _{indep} : model likelihood				15.2.4 PCA and nearest	
		12.6.3 H_{same} : model likelihood					
		12.6.4 Dependent outcome analysis				neighbours classification 15.2.5 Comments on PCA	
		12.6.5 Is classifier A better than B?				15.2.5 Comments on FCA	

Contents

15.3	High-dimensional data				17.4.1	Logistic regression	
	15.3.1 Eigen-decomposition for					Beyond first-order	
	N < D					gradient ascent	
	15.3.2 PCA via singular value				17.4.3	Avoiding overconfident	
	decomposition					classification	
15.4	Latent semantic analysis				17.4.4	Multiple classes	
	15.4.1 Information retrieval				17.4.5	The kernel trick for	
15.5	PCA with missing data					classification	
	15.5.1 Finding the principal			17.5		ort vector machines	
	directions				17.5.1	Maximum margin linear	
	15.5.2 Collaborative filtering					classifier	
	using PCA with missing					Using kernels	
1 7 /	data					Performing the optimisation	
15.6	Matrix decomposition methods			177.		Probabilistic interpretation	
	15.6.1 Probabilistic latent			17.6		ero-one loss for outlier	
	semantic analysis			10.0	robust		
	15.6.2 Extensions and variations15.6.3 Applications of PLSA/NMF				Summ	iary	
157	Kernel PCA				Code		
	Canonical correlation analysis			17.9	Exerci	ses	
13.0	15.8.1 SVD formulation			_			
15 0	Summary		18	-		ear models	392
	Code			18.1	_	ssion with additive	
	Exercises					ian noise	
13.11	Exercises				18.1.1	Bayesian linear parameter	
Cunar	wised linear dimension				10 1 0	models	
reduc	vised linear dimension	359			18.1.2	Determining M. H.	
		223			1013	hyperparameters: ML-II Learning the	
	Supervised linear projections Fisher's linear discriminant				10.1.5	hyperparameters using EM	
					18.1.4	Hyperparameter Hyperparameter	
10.5	Canonical variates 16.3.1 Dealing with the nullspace					optimisation: using the	
16.4	Summary					gradient	
	Code				18.1.5	Validation likelihood	
	Exercises				18.1.6	Prediction and model	
10.0	Exercises					averaging	
Linon	r models	367				Sparse linear models	
		307		18.2		fication	
1 /.1	Introduction: fitting a straight					Hyperparameter optimisation	
170	line					Laplace approximation	
17.2	Linear parameter models for				18.2.3	Variational Gaussian	
	regression				1024	approximation	
	17.2.1 Vector outputs17.2.2 Regularisation				18.2.4	Local variational	
	17.2.2 Regularisation 17.2.3 Radial basis functions				1825	approximation Relevance vector	
173					10.4.5	machine for classification	
17.5	The dual representation and kernels				1826	Multi-class case	
	17.3.1 Regression in the dual space			183	Summ		
171	Linear parameter models for				Code	iui j	
1/.4	classification				Exerci	ises	
	CIGOGIIICGIICII			10.0	~~,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		

9	Gaussian processes					20.3.6 Bayesian mixture models	
		Non-parametric prediction				20.3.7 Semi-supervised learning	
		19.1.1 From parametric to			20.4	Mixture of experts	
		non-parametric			20.5	Indicator models	
		19.1.2 From Bayesian linear				20.5.1 Joint indicator approach:	
		models to Gaussian processes				factorised prior	
		19.1.3 A prior on functions				20.5.2 Polya prior	
		Gaussian process prediction			20.6	Mixed membership models	
		19.2.1 Regression with noisy				20.6.1 Latent Dirichlet allocation	
		training outputs				20.6.2 Graph-based	
	10.3	Covariance functions				representations of data	
		19.3.1 Making new covariance				20.6.3 Dyadic data	
		functions from old				20.6.4 Monadic data	
		19.3.2 Stationary covariance				20.6.5 Cliques and adjacency	
		functions				matrices for monadic	
		19.3.3 Non-stationary				binary data	
		covariance functions			20.7	Summary	
	10 /	Analysis of covariance				Code	
		functions				Exercises	
		19.4.1 Smoothness of the functions			20.5	Excicises	
		19.4.2 Mercer kernels		24	Laton	t linear models	462
		19.4.3 Fourier analysis for		21		it illical illoadis	+02
		stationary kernels			21.1	Factor analysis	
	10.5	Gaussian processes for				21.1.1 Finding the optimal bias	
	19.3	classification			21.2	Factor analysis: maximum	
						likelihood	
		19.5.1 Binary classification19.5.2 Laplace's approximation				21.2.1 Eigen-approach	
		19.5.3 Hyperparameter optimisation				likelihood optimisation	
		19.5.4 Multiple classes				21.2.2 Expectation maximisation	
	10.6	_				Interlude: modelling faces	
		Summary			21.4	Probabilistic principal	
		19.7 Code 19.8 Exercises				components analysis	
	19.8				21.5	Canonical correlation analysis	
			422			and factor analysis	
20		re models	432		21.6	Independent components	
	20.1	Density estimation using				analysis	
		mixtures			21.7	Summary	
	20.2	Expectation maximisation for				Code	
		mixture models				Exercises	
		20.2.1 Unconstrained discrete			2	Likeleises	
		tables		22	Lator	nt ability models	479
		20.2.2 Mixture of product of		22		The Rasch model	413
		Bernoulli distributions			22.1		
	20.3	The Gaussian mixture model				22.1.1 Maximum likelihood	
		20.3.1 EM algorithm				training	
		20.3.2 Practical issues			22.2	22.1.2 Bayesian Rasch models	
		20.3.3 Classification using			22.2	Competition models	
		Gaussian mixture models				22.2.1 Bradley-Terry-Luce model	
		20.3.4 The Parzen estimator				22.2.2 Elo ranking model	
		20.3.5 K-means				22.2.3 Glicko and TrueSkill	

22.4 Code 22.5 Exercises IV Dynamical models 23 Discrete-state Markov models 23.1 Markov models 23.1.1 Equilibrium and stationary distribution of a Markov chain 23.1.2 Fitting Markov models 23.2.1 Hidden Markov models 23.2.2 Filtering $p(h_t v _{1:T})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v _{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 Earming HMMs 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.4 Inference 24.4.1 Filtering 24.4.2 Smoothing: Rauch-Tung-S correction emit 24.4.3 Time independe Riccati equation 24.5.1 Identifiability is 24.5.2 EM algorithm 24.5.3 Subspace methe 24.5.4 Structured LDS 24.5.5 Bayesian LDSs 24.6.1 Inference 24.6.2 Maximum likelifearning using E 24.7 Summary 24.8 Code 24.9 Exercises 25 Switching linear dynamic systems 25.1.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.2 Discrete filterin 25.3.3 The likelihood 25.3.4 Collapsing Gau 25.3.4 Collapsing Gau 25.3.5 Relation to othe	lels
IV Dynamical models 23 Discrete-state Markov models 23.1 Markov models 23.1.1 Equilibrium and stationary distribution of a Markov chain 23.1.2 Fitting Markov models 23.1.3 Mixture of Markov models 23.2.1 Flitering $p(h_t v_{1:T})$ 23.2.2 Filtering $p(h_t v_{1:T})$ 23.2.3 Parallel smoothing $p(h_t v_{1:T})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.2 Input-output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.3 Applications 23.5.3 Applications 23.5.4 Vattomatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises	model
V Dynamical models 24.2.4 Time-varying v AR models 23.1 Markov models 23.1.1 Equilibrium and stationary distribution of a Markov chain 24.4.1 Filtering 24.4.2 Smoothing: Rauch-Tung-S correction method 23.1.3 Mixture of Markov models 23.2.1 The classical inference problems 23.2.2 Filtering p(h _t v _{1:t}) 23.2.4 Correction smoothing 23.2.5 Sampling from p(h _{1:T} v _{1:T}) 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 24.5 Switching auto-regree models 23.4.1 Explicit duration model 23.4.2 Input-output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.4 Part-of-speech tagging 23.5.4 Part-of-speech tagging 23.5.3 Relation to othe 25.3.5 Relation to the 25.3.5 Relatio	ı OLDS
AR models 23.1 Markov models 23.1.1 Equilibrium and stationary distribution of a Markov chain 23.1.2 Fitting Markov models 23.1.3 Mixture of Markov models 23.2.1 The classical inference problems 23.2.2 Filtering $p(h_t v_{1:T})$ 23.2.3 Parallel smoothing $p(h_t v_{1:T})$ 23.2.4 Correction smoothing $p(h_t v_{1:T})$ 23.2.5 Sampling from $p(h_{t:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input—output HMM 23.4.3 Linear chain CRFs 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5 Code 23.6 Code 23.7 Code 23.7 Code 23.7 Code 23.8 Exercises AR models 24.4.1 Filtering 24.4.2 Smoothing: Rauch—Tung—S correction meth 24.4.3 The likelihood 24.4.5 Time independe Riccati equation 24.5.1 Learning linear dynamic expection meth 24.5.2 EM algorithm 24.5.3 Subspace methe 24.5.3 Subspace methe 24.5.4 Structured LDS 24.5.5 Bayesian LDSs 24.6.1 Inference 24.4.2 Smoothing: Rauch—Tung—S correction meth 24.4.3 The likelihood 24.4.4 Most likely stat 24.5.1 Identifiability is 24.5.2 EM algorithm 24.5.5 Bayesian LDSs 24.6.1 Inference 24.6.2 Maximum likeli learning using E 24.6.2 Maximum likeli learning using E 25.3 Switching linear dynamic systems 25.1 Introduction 25.2 The switching LDS 25.2 The switching LDS 25.3 Gaussian sum filterin 25.3.1 Continuous filte 25.3.2 Discrete filterin	R model
23.1 Markov models 23.1.1 Equilibrium and stationary distribution of a Markov chain 23.1.2 Fitting Markov models 23.1.3 Mixture of Markov models 23.2.1 The classical inference problems 23.2.2 Filtering $p(h_t v_{1:T})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.2 Mixture emission 23.3.4 Discriminative training 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.6 Summary 23.7 Code 23.8 Exercises 24.4 Inference 24.4.1 Filtering 24.4.2 Smoothing: Rauch—Tung—S correction meth 24.4.5 The likelihood 24.4.5 Time independe Riccati equation 24.5.1 Identifiability is 24.5.2 EM algorithm 24.5.2 EM algorithm 24.5.3 Subspace meth 24.5.5 Bayesian LDSs 24.5.5 Bayesian LDSs 24.5.6 Switching auto-regres models 24.6.1 Inference 24.4.2 Smoothing: 24.4.2 Smoothing: 24.4.2 Smoothing: 24.4.3 The likelihood 24.4.5 Time independe Riccati equation 24.5.1 Identifiability is 24.5.2 EM algorithm 24.5.2 EM algorithm 24.5.3 Subspace meth 24.5.5 Bayesian LDSs 24.5.1 Switching auto-regres models 24.6.2 Maximum likeling learning using Emission 24.6.2 Maximum likeling learning using Emission 24.6.2 Switching linear dynamica systems 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference Riccati equation 25.3.1 Continuous filte 24.5.3 The likelihood 25.3.4 Collapsing Gau	ariance
23.1 Markov models 23.1.1 Equilibrium and stationary distribution of a Markov chain 23.1.2 Fitting Markov models 23.1.3 Mixture of Markov models 23.2.1 The classical inference problems 23.2.2 Filtering $p(h_t v_{1:T})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.1 Explicit duration model 23.4.1 Explicit duration model 23.4.2 Input-output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.6 Summary 23.7 Code 23.8 Exercises	cal systems
stationary distribution of a Markov chain 23.1.2 Fitting Markov models 23.1.3 Mixture of Markov models 23.2.1 The classical inference problems 23.2.2 Filtering $p(h_t v_{1:T})$ 23.2.3 Parallel smoothing $p(h_t v_{1:T})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input-output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.6 Summary 23.7 Code 23.8 Exercises 24.4.2 Smoothing: Rauch-Tung-S correction medels 24.4.3 The likelihood 24.4.5 Time independe Riccati equation 24.5.1 Learning linear dynamic subspace mether 24.5.2 EM algorithm 24.5.3 Subspace mether 24.5.4 Structured LDS 24.6.1 Inference 24.6.2 Maximum likeling 24.6.2 Maximum likeling 24.6.3 Switching auto-regrest models 24.6.1 Inference 24.6.2 Maximum likeling 24.6.3 Switching auto-regrest models 24.6.1 Inference 24.6.2 Smoothing: 24.4.4 Mot likely state 24.5.3 Time independe Riccati equation 24.5.5 EM algorithm 24.5.5 Bayesian LDSs 24.6.1 Inference 24.6.2 Maximum likeling 24.6.2 Vasummary 24.6.3 Switching auto-regrest models 24.6.3 Switching auto-regrest models 25 Switching linear dynamics 25.3 Introduction 25.3.1 Introduction 25.3.1 Exact inference computationally 25.3.3 The likelihood 25.3.1 Continuous filtering 25.3.3 The likelihood 25.3.3 The likelihood 25.3.3 The likelihood 25.3.3 The likelihood 25.3.4 Collapsing Gau 25.3.3 Relation to other	
stationary distribution of a Markov chain 23.1.2 Fitting Markov models 23.2.3 Mixture of Markov models 23.2.1 The classical inference problems 23.2.2 Filtering $p(h_t v_{1:t})$ 23.2.3 Parallel smoothing $p(h_t v_{1:t})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.2 Input–output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 24.4 Most likely stat 24.4.5 Time independe Riccati equation 24.5.1 Identifiability in 24.5.3 Subspace methe 24.5.2 EM algorithm 24.5.2 EM algorithm 24.5.3 Subspace methe 24.5.4 Structured LDS 24.5.5 Bayesian LDSs 24.6.1 Inference 24.6.2 Maximum likeling learning using Email of the subspace	
23.1.2 Fitting Markov models 23.1.3 Mixture of Markov models 23.2.4 Hidden Markov models 23.2.1 The classical inference problems 23.2.2 Filtering $p(h_t v_{1:T})$ 23.2.3 Parallel smoothing $p(h_t v_{1:T})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.2 Input–output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises correction meth 24.4.3 The likelihood 24.4.5 Time independe Riccati equation 24.5.1 Identifiability is 24.5.2 EM algorithm 24.5.3 Subspace methor 24.5.3 Subspace methor 24.5.4 Structured LDS 24.6.1 Inference 24.6.2 Maximum likeli learning using E 24.6.2 Maximum likeli learning using E 24.6.3 Summary 24.5.1 Identifiability is 24.5.4 Structured LDS 24.5.4 Structured LDS 24.5.5 Bayesian LDSs 24.6 Switching auto-regret models 24.6.2 Maximum likeli learning using E 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.1 Continuous filte in 25.3.1 Continuous filte in 25.3.2 Discrete filterin model 25.3.3 Element depution and independence Riccati equation and 24.5.5 Identifiability is 24.5.1 Identifiability is 24.5.2 EM algorithm 24.5.3 Subspace methor 24.5.3 Subspace methor 24.5.4 Structured LDS 24.5. Switching auto-regret models 25.5 Saysian LDSs 25.6 Switching auto-regret models 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.1 Explications and 24.6.2 Maximum likeli learning using E 25.3 Gaussian sum filterin 25.3.1 Continuous filte 25.3.2 Discrete filterin 25.3.3 The likelihood 25.3.4 Collapsing Gau	
23.1.3 Mixture of Markov models 23.2.1 Hidden Markov models 23.2.2 Filtering $p(h_t v_{1:T})$ 23.2.3 Parallel smoothing $p(h_t v_{1:T})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMMs 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.2 Input–output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.5 Applications 23.5.6 Summary 23.6 Summary 23.7 Code 23.7 Code 24.4.3 The likelihood 24.4.4 Most likely state 24.4.5 Time independe Riccati equation 24.5.1 Identifiability is 24.5.2 EM algorithm 24.5.3 Subspace methe 24.5.4 Structured LDS 24.6 Switching auto-regres models 24.6.1 Inference 24.6.2 Maximum likeli learning using E 24.7 Summary 24.8 Code 24.9 Exercises 25 Switching linear dynamics systems 25.1 Introduction 25.2.1 Exact inference Riccati equation 24.5.5 Emalogorithm 24.5.2 EM algorithm 24.5.3 Subspace methe 24.6.5 Witching auto-regres models 24.6.1 Inference 24.6.2 Maximum likeli learning using E 24.5. Semilarions 24.5.5 Bayesian LDSs 24.6.1 Inference 24.6.2 Maximum likeli learning using E 25.5 Switching linear dynamics systems 25.1 Introduction 25.2.1 Exact inference computationally systems 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.3 The likelihood graph of the point of the production of the product	
23.2 Hidden Markov models 23.2.1 The classical inference problems 23.2.2 Filtering $p(h_t v_{1:T})$ 23.2.3 Parallel smoothing $p(h_t v_{1:T})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.2 Input—output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.5 Applications 23.5.6 Summary 23.7 Code 23.8 Exercises 24.4.4 Most likely state 24.4.5 Time independe Riccati equation 24.5.1 Identifiability is 24.5.2 EM algorithm 24.5.3 Subspace methor 24.5.5 Bayesian LDSs 24.6 Switching auto-regree models 24.6.1 Inference 24.6.2 Maximum likeli learning using E 24.7 Summary 24.8 Code 24.9 Exercises 25 Switching linear dynamica systems 25.1 Introduction 25.2.1 Exact inference computationally 25.3.2 Discrete filterin 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.3 The likelihood, 25.3.4 Collapsing Gau 25.3.5 Relation to other	od
23.2.1 The classical inference problems 23.2.2 Filtering $p(h_t v_{1:T})$ 23.2.3 Parallel smoothing $p(h_t v_{1:T})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.2 Input-output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 24.5 Learning linear dynam 24.5.1 Identifiability is 24.5.2 EM algorithm 24.5.3 Subspace metho 24.5.4 Structured LDS 24.5.5 Bayesian LDSs 24.6.1 Inference 24.6.2 Maximum likeli learning using E 24.6 Switching linear dynamical systems 25.3 Introduction 25.2 The switching LDS 25.3.1 Continuous filtering 25.3.2 Discrete filtering 25.3.3 The likelihood gate of the problems 25.3.4 Collapsing Gau 25.3.5 Relation to other	
Problems 23.2.2 Filtering $p(h_t v_{1:T})$ 23.2.3 Parallel smoothing $p(h_t v_{1:T})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3 Learning HMMs 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input-output HMM 23.4.3 Linear chain CRFs 23.4 A Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 24.5 Learning linear dynamical speech recognition 24.5.5 Bayesian LDSs 24.6.1 Inference 24.6.2 Maximum likeling 24.7 Summary 24.8 Code 24.9 Exercises 25 Switching linear dynamical systems 25.1 Introduction 25.2.1 Exact inference computationally computation computationally computationally computationally computationally computation computationally computation computationally computatio	
23.2.2 Filtering $p(h_t v_{1:t})$ 23.2.3 Parallel smoothing $p(h_t v_{1:T})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input—output HMM 23.4.3 Linear chain CRFs 23.4 A Dynamic Bayesian networks 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5 Summary 23.6 Summary 23.7 Code 24.5 Learning linear dynamic 24.5.1 Identifiability is 24.5.2 EM algorithm 24.5.2 EM algorithm 24.5.3 Subspace method 24.5.5 Bayesian LDSs 24.6.1 Inference 24.6.2 Maximum likelinear dynamic 324.7 Summary 24.8 Code 24.9 Exercises 25 Switching linear dynamica 25.2.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.3 The likelihood 23.3.5 Relation to other	
23.2.3 Parallel smoothing $p(h_t v_{1:T})$ 23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.2 Input—output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5 Summary 23.6 Summary 23.7 Code 23.8 Exercises 24.5 EM algorithm 24.5.5 Bayesian LDSs 24.6.1 Inference 24.6.2 Maximum likeli learning using E 24.7 Summary 24.8 Code 24.9 Exercises 25 Switching linear dynamica systems 25.1 Introduction 25.2 The switching LDS 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.3 The likelihood of 25.3.4 Collapsing Gau 25.3.5 Relation to othe	
23.2.4 Correction smoothing 23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input—output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.6 Summary 23.7 Code 23.8 Exercises 24.5.3 Subspace method 24.5.5 Bayesian LDSs 24.6.1 Inference 24.6.2 Maximum likelidiearning using E 24.7 Summary 24.8 Code 24.9 Exercises 25 Switching linear dynamical systems 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.1 Continuous filtering 25.3.2 Discrete filtering 25.3.3 The likelihood of 25.3.4 Collapsing Gau 25.3.5 Relation to other	•
23.2.5 Sampling from $p(h_{1:T} v_{1:T})$ 23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3 Learning HMMs 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input-output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Subspace method 24.5.4 Structured LDS 24.6.5 Switching auto-regres models 24.6.1 Inference 24.6.2 Maximum likeling 24.7 Summary 24.8 Code 24.9 Exercises 25 Switching linear dynamical systems 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.3 Gaussian sum filtering 25.3.4 Continuous filte 25.3.5 Gaussian sum filtering 25.3.5 Point on the	sues
23.2.6 Most likely joint state 23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3 Learning HMMs 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input-output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 24.6 Switching auto-regremodels 24.6.2 Maximum likeling 24.8 Code 24.8 Code 24.9 Exercises 25 Switching linear dynamical systems 25.1 Introduction 25.2.1 Exact inference computationally 25.3.1 Continuous filtering 25.3.2 Discrete filtering 25.3.3 The likelihood of the service of the selection of	
23.2.7 Prediction 23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input—output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.6 Summary 23.7 Code 23.8 Exercises 24.6 Switching auto-regrest models 24.6.1 Inference 24.6.2 Maximum likelit learning using E 24.7 Summary 24.8 Code 24.9 Exercises 25 Switching linear dynamics systems 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally computationally 25.3.1 Continuous filte 25.3.2 Discrete filterin 25.3.3 The likelihood of 25.3.4 Collapsing Gau 25.3.5 Relation to other	
23.2.8 Self-localisation and kidnapped robots 23.2.9 Natural language models 23.3.1 Learning HMMs 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input—output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 24.6 Switching auto-regremodels 24.6.2 Maximum likeli 24.6.2 Maximum likeli 24.6 Summary 24.6 Switching linear dynamical systems 25.1 Introduction 25.2 The switching LDS 25.2 The switching LDS 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.3 The likelihood of 25.3.4 Collapsing Gau 25.3.5 Relation to other	3
kidnapped robots 23.2.9 Natural language models 23.3 Learning HMMs 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input–output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises models 24.6.1 Inference 24.6.2 Maximum likeli learning using E 24.7 Summary 24.8 Code 24.9 Exercises 25 Switching linear dynamica systems 25.1 Introduction 25.2.1 Exact inference computationally 25.2 The switching LDS 25.3.1 Continuous filte 25.3.2 Discrete filterin 25.3.2 Discrete filterin 25.3.3 The likelihood page 25.3.4 Collapsing Gau 25.3.5 Relation to other	. • .
23.2.9 Natural language models 23.2.9 Natural language models 23.2.9 Natural language models 24.6.1 Inference 24.6.2 Maximum likeli learning using E 24.7 Summary 24.8 Code 23.3.4 Discriminative training 24.9 Exercises 23.4 Related models 23.4.1 Explicit duration model 23.4.2 Input—output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 24.6.2 Maximum likeli learning using E 24.7 Summary 24.8 Code 25.9 Exercises 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.1 Continuous filte 25.3.2 Discrete filterin 25.3.3 The likelihood of the control of the con	ssive
23.3 Learning HMMs 23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input—output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 24.7 Summary 24.8 Code 24.9 Exercises 25 Switching linear dynamica systems 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.3 The likelihood of 25.3.4 Collapsing Gau 25.3.5 Relation to other	
23.3.1 EM algorithm 23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input-output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 24.7 Summary 24.8 Code 24.9 Exercises 25 Switching linear dynamica systems 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.2 Discrete filterin 25.3.3 The likelihood 25.3.4 Collapsing Gau 25.3.5 Relation to other	hood
23.3.2 Mixture emission 23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4.1 Explicit duration model 23.4.2 Input–output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 24.8 Code 24.9 Exercises 25 Switching linear dynamica systems 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.1 Continuous filterin 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.2 Discrete filterin 25.3.3 The likelihood 25.3.4 Collapsing Gau 25.3.5 Relation to other	
23.3.3 The HMM-GMM 23.3.4 Discriminative training 23.4 Related models 23.4.1 Explicit duration model 23.4.2 Input-output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 24.9 Exercises 25 Switching linear dynamica systems 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.1 Continuous filterin 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.3 The likelihood 25.3.4 Collapsing Gau 25.3.5 Relation to other	·1V1
23.3.4 Discriminative training 23.3.4 Discriminative training 23.4.2 Related models 23.4.2 Input–output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 24.9 Exercises 25 Switching linear dynamica systems 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.3 The likelihood part of the systems 25.3.4 Collapsing Gau 25.3.5 Relation to other	
23.4 Related models 23.4.1 Explicit duration model 23.4.2 Input-output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 25 Switching linear dynamica systems 25.1 Introduction 25.2.1 Exact inference computationally 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.3 The likelihood part of the systems 25.4 Collapsing Gau 25.3.5 Relation to other	
23.4.1 Explicit duration model 23.4.2 Input–output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 25 Switching linear dynamica systems 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.3 The likelihood part of the systems 25.4 Part-of-speech tagging 25.3.4 Collapsing Gau 25.3.5 Relation to other	
23.4.2 Input-output HMM 23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.1 Gaussian sum filterin 25.3.2 Discrete filterin 25.3.3 The likelihood part of the summary 25.3.4 Collapsing Gau 25.3.5 Relation to other	al.
23.4.3 Linear chain CRFs 23.4.4 Dynamic Bayesian networks 23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 25.1 Introduction 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3.2 Gaussian sum filterin 25.3.1 Continuous filte 25.3.2 Discrete filterin 25.3.3 The likelihood page of the properties of the prope	547
23.4.4 Dynamic Bayesian networks 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 25.2 The switching LDS 25.2.1 Exact inference computationally 25.3 Gaussian sum filterin 25.3.1 Continuous filte 25.3.2 Discrete filterin 25.3.3 The likelihood part of the computational sum filterin 25.3.4 Collapsing Gau 25.3.5 Relation to other	547
23.5 Applications 23.5.1 Object tracking 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 25.2 The swhething EDS 25.2.1 Exact inference computationally 25.3.2 Gaussian sum filterin 25.3.1 Continuous filte 25.3.2 Discrete filterin 25.3.3 The likelihood page 25.3.4 Collapsing Gau 25.3.5 Relation to other	
23.5.1 Object tracking computationally 23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 25.3.1 Continuous filterin 25.3.2 Discrete filterin 25.3.3 The likelihood part of the part of	
23.5.2 Automatic speech recognition 23.5.3 Bioinformatics 23.5.4 Part-of-speech tagging 23.6 Summary 23.7 Code 23.8 Exercises 25.3 Gaussian sum filterin 25.3.1 Continuous filte 25.3.2 Discrete filterin 25.3.3 The likelihood 25.3.4 Collapsing Gau 25.3.5 Relation to othe	
23.5.3 Bioinformatics 25.3.1 Continuous filte 23.5.4 Part-of-speech tagging 25.3.2 Discrete filterin 25.3.3 The likelihood of the likeliho	
23.5.4 Part-of-speech tagging 25.3.2 Discrete filterin 25.3.3 The likelihood of the	•
23.6 Summary 25.3.3 The likelihood 23.7 Code 25.3.4 Collapsing Gau 23.8 Exercises 25.3.5 Relation to other	_
23.7 Code 25.3.4 Collapsing Gau 23.8 Exercises 25.3.5 Relation to other	*
23.8 Exercises 25.3.5 Relation to other	
24 Continuous-state Markov models 520 25.4.1 Continuous smo	_
24.1 Observed linear dynamical 25.4.2 Discrete smooth	-
systems 25.4.3 Collapsing the r	_
24.1.1 Stationary distribution 25.4.4 Using mixtures	
with noise 25.4.5 Relation to other	

	25.5	Reset models				27.4.1	Markov chains	
		25.5.1 A Poisson reset model				27.4.2	Metropolis-Hastings	
		25.5.2 Reset-HMM-LDS					sampling	
	25.6	Summary			27.5	Auxili	ary variable methods	
		Code				27.5.1	Hybrid Monte Carlo (HMC)	
		Exercises				27.5.2	Swendson-Wang (SW)	
	22. 0	E.ROTOTOOS				27.5.3	Slice sampling	
26	Distri	buted computation	568		27.6	Impor	tance sampling	
20		Introduction	500				Sequential importance	
							sampling	
		Stochastic Hopfield networks				27.6.2	Particle filtering as an	
	26.3	Learning sequences					approximate forward pass	
		26.3.1 A single sequence			27.7	Summ		
		26.3.2 Multiple sequences				Code	-	
		26.3.3 Boolean networks				Exerc	ises	
	26.4	26.3.4 Sequence disambiguation			27.5	23,1010	.500	
	26.4	Tractable continuous latent		28	Datar	minist	ic approximate	
		variable models		20	infere		ic approximate	617
		26.4.1 Deterministic latent variables					nation	017
		26.4.2 An augmented Hopfield				Introd		
		network					aplace approximation	
	26.5	Neural models			28.3		rties of Kullback-	
		26.5.1 Stochastically spiking					er variational inference	
		neurons				28.3.1	Bounding the	
		26.5.2 Hopfield membrane					normalisation constant	
		potential				28.3.2	Bounding the marginal	
		26.5.3 Dynamic synapses					likelihood	
		26.5.4 Leaky integrate and fire				28.3.3	Bounding marginal	
	0//	models				20.2.4	quantities	
		Summary				28.3.4	Gaussian approximations	
	-	Code				20.2 5	using KL divergence	
	26.8	Exercises				28.3.3	Marginal and moment	
	_						matching properties of	
V	Appr	oximate inference			20.4	1 7	minimising $KL(p q)$	I \
					28.4		ional bounding using $KL(q)$ Pairwise Markov random field	
27	Samp	lling	587					1
	27.1	Introduction					General mean-field equations Asynchronous updating	
		27.1.1 Univariate sampling				20.4.3	guarantees approximation	
		27.1.2 Rejection sampling					improvement	
		27.1.3 Multivariate sampling				28 4 4	Structured variational	
	27.2	Ancestral sampling				20.4.4	approximation	
		27.2.1 Dealing with evidence			28.5	Local	and KL variational	
		27.2.2 Perfect sampling for a			20.5			
		Markov network					kimations	
	27.3	Gibbs sampling					Local approximation	
		27.3.1 Gibbs sampling as a				28.3.2	KL variational	
		Markov chain			40.7	1 (approximation	
		27.3.2 Structured Gibbs sampling			28.6		l information	
		27.3.3 Remarks					nisation: a KL	
	27.4	Markov chain Monte Carlo (MCM	C)			variati	onal approach	

	28.6.1 The information	28.12 Code	
	maximisation algorithm	28.13 Exercises	
	28.6.2 Linear Gaussian decoder		
28.7	Loopy belief propagation	Appendix A: Background mathematics	655
	28.7.1 Classical BP on an	A.1 Linear algebra	
	undirected graph	A.2 Multivariate calculus	
	28.7.2 Loopy BP as a variational procedure	A.3 Inequalities	
28.8	Expectation propagation	A.4 Optimisation	
	MAP for Markov networks	A.5 Multivariate optimisation	
20.9	28.9.1 Pairwise Markov networks	A.6 Constrained optimisation	
	28.9.2 Attractive binary Markov	using Lagrange multipliers	
	networks	References	675
	28.9.3 Potts model	Index	689
28.10	Further reading		
28.11	Summary	Colour plate section between pp. 360 and 361	