

University of Exeter
Department of Computer Science

Robust spatio-temporal latent variable models

Jacqueline Christmas

January 2011

Submitted by Jacqueline Christmas, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Computer Science , January 2011.

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

(signature)

Abstract

Principal Component Analysis (PCA) and Canonical Correlation Analysis (CCA) are widely-used mathematical models for decomposing multivariate data. They capture spatial relationships between variables, but ignore any temporal relationships that might exist between observations. Probabilistic PCA (PPCA) and Probabilistic CCA (ProbCCA) are versions of these two models that explain the statistical properties of the observed variables as linear mixtures of an alternative, hypothetical set of hidden, or latent, variables and explicitly model noise. Both the noise and the latent variables are assumed to be Gaussian distributed.

This thesis introduces two new models, named PPCA-AR and ProbCCA-AR, that augment PPCA and ProbCCA respectively with autoregressive processes over the latent variables to additionally capture temporal relationships between the observations. To make PPCA-AR and ProbCCA-AR robust to outliers and able to model leptokurtic data, the Gaussian assumptions are replaced with infinite scale mixtures of Gaussians, using the Student-t distribution.

Bayesian inference calculates posterior probability distributions for each of the parameter variables, from which we obtain a measure of confidence in the inference. It avoids the pitfalls associated with the maximum likelihood method: integrating over all possible values of the parameter variables guards against overfitting. For these new models the integrals required for exact Bayesian inference are intractable; instead a method of approximation, the variational Bayesian approach, is used. This enables the use of automatic relevance determination to estimate the model orders.

PPCA-AR and ProbCCA-AR can be viewed as linear dynamical systems, so the forward-backward algorithm, also known as the Baum-Welch algorithm, is used as an efficient method for inferring the posterior distributions of the latent variables. The exact algorithm is tractable because Gaussian assumptions are made regarding the distribution of the latent variables. This thesis introduces a variational Bayesian forward-backward algorithm based on Student-t assumptions.

The new models are demonstrated on synthetic datasets and on real remote sensing and EEG data.

Contents

List of tables	8
List of figures	9
Nomenclature and Abbreviations	17
Publications	18
1 Introduction	19
1.1 Real data examples	21
ONS data	21
NASA data	22
EEG data	23
1.2 Summary of new models	24
1.3 Aims	25
1.4 Thesis structure	25
1.5 Novel contributions	26
2 Background	27
2.1 Bayesian probability theory	27
2.1.1 Bayesian inference	29
2.1.2 Bayesian approximation techniques	31
2.1.3 Variational Bayesian approximation	32
2.2 Latent Variable Models	34
2.2.1 Principal Component Analysis	36
2.2.2 Factor Analysis	37
2.2.3 Probabilistic PCA	38
2.2.4 Canonical Correlation Analysis	40
2.2.5 Probabilistic CCA	41
2.3 Model order selection	41
2.4 Robustness	44
2.5 Assessing model quality	47
2.5.1 Ambiguous transformations	48
2.5.2 Missing data estimation for PCA-like models	49
2.5.3 Error measures	51
2.6 Conclusions	51
3 Maximum likelihood PPCA with autoregressive latent variables	55

3.1	Autoregression	56
3.2	PPCA-AR: a linear dynamical system	57
	Forward-backward algorithm	58
	EM algorithm: E-step	59
	EM algorithm: M-step	60
	Imputing missing values	62
3.3	Illustration: synthetic data	64
	3.3.1 Illustration 1: estimating noise precisions	64
	3.3.2 Illustration 2: estimating \mathbf{W} and $\boldsymbol{\theta}$	68
	3.3.3 Illustration 3: imputing missing values	68
3.4	Results: real data	70
3.5	Conclusions	71
4	Robust variational PPCA	74
4.1	Gaussian variational PPCA	75
4.2	Robust variational PPCA	75
4.3	New robust variational PPCA	78
	4.3.1 Priors	79
	4.3.2 Factorised variational Bayesian method	80
	Mixing matrix, \mathbf{W}	81
	Latent variables, \mathbf{x}_n	82
	ARD parameters, $\boldsymbol{\delta}$	82
	Latent variables, \mathbf{z}_n	83
	Degrees of freedom, \mathbf{d}	84
	Latent variables, \mathbf{u}	84
	Observation noise degrees of freedom, \mathbf{v}	85
	Observation noise precision, λ	85
	Observation mean, $\bar{\mathbf{t}}$	86
4.4	Summary	86
4.5	Convergence criterion	89
4.6	Illustration: synthetic data	89
	4.6.1 Illustration 1: parameter estimation	90
	4.6.2 Illustration 2: model order estimation	95
	4.6.3 Illustration 3: robustness to outliers	96
4.7	Robust variational FA	99
4.8	Conclusions	100
5	Robust variational autoregression	102
5.1	Robust variational AR	102
	5.1.1 Priors	103
	5.1.2 Factorised variational Bayesian method	104
	AR coefficients, $\boldsymbol{\theta}$	105
	Excitation noise precision, κ	105
	Latent variables, \mathbf{z}	106
	Degrees of freedom, d	106

	ARD precisions, γ	106
5.2	Summary	107
5.3	Illustration: synthetic data	108
	5.3.1 Illustration 1: parameter estimation	109
	5.3.2 Illustration 2: model order estimation	110
	5.3.3 Illustration 3: robustness to outliers	111
5.4	Results: real data	113
5.5	Conclusions	114
6	Robust variational PPCA with autoregression	116
6.1	Priors	117
	6.1.1 Robust variational forward-backward algorithm	120
	Forward sweep	120
	Backward sweep	122
	6.1.2 AR coefficients, θ_q	123
	6.1.3 Latent variables, \mathbf{z}_n	124
6.2	Summary	125
6.3	Illustration: synthetic data	129
	6.3.1 Illustration 1: estimating noise parameters	129
	6.3.2 Illustration 2: estimating $\bar{\mathbf{t}}$, \mathbf{W} and $\boldsymbol{\theta}$	130
	6.3.3 Illustration 3: variances of approximate posterior distributions	135
	6.3.4 Illustration 4: estimating model orders	137
	6.3.5 Illustration 5: imputing missing values	138
6.4	Results: real data	142
6.5	Robust variational FA-AR	149
6.6	Conclusions	149
7	Robust variational CCA with autoregression	152
7.1	Priors	153
7.2	Factorised variational Bayesian method	155
	7.2.1 Mean observation, $\bar{\mathbf{t}}$	155
	7.2.2 Mixing matrix, \mathbf{W}	156
	7.2.3 Precision matrix, \mathbf{B}	157
	7.2.4 Latent variables, \mathbf{x}_n	158
	Forward sweep	159
	Backward sweep	160
	7.2.5 Summary	160
7.3	Illustration: synthetic data	163
	7.3.1 Illustration 1: parameter estimation	164
	7.3.2 Illustration 2: model order estimation	170
7.4	Results: real data	171
7.5	Extension to the ProbCCA-AR model	175
7.6	Conclusions	178
8	Conclusions	180

8.1	Future work	183
8.2	Summary	185
Appendices		186
A	Probability Distributions	187
A.1	Gamma	187
A.2	Gaussian	187
A.3	Student-t	188
A.4	Wishart	188
B	Maximum Likelihood estimation for Gaussian and Student-t distribution parameters	189
B.1	Gaussian maximum likelihood estimation	189
B.2	Student-t maximum likelihood estimation	190
Bibliography		192