Bayesian Canonical Correlation Analysis, (BCCA)

Submitted By

Prajal Badjatya

0801CS171055

Contents

Chapter 1

Introduction

1.1 Abstract

Canonical correlation analysis (CCA) is a classical method for seeking correlations between two multivariate data sets. During the last ten years, it has received more and more attention in the machine learning community in the form of novel computational formulations and a plethora of applications. We review recent developments in Bayesian models and inference methods for CCA which are attractive for their potential in hierarchical extensions and for coping with the combination of large dimensionalities and small sample sizes. The existing methods have not been particularly successful in fulfilling the promise yet; we introduce a novel efficient solution that imposes group-wise sparsity to estimate the posterior of an extended model which not only extracts the statistical dependencies (correlations) between data sets but also decomposes the data into shared and data set-specific components. In statistics literature the model is known as inter-battery factor analysis (IBFA), for which we now provide a Bayesian treatment

1.2 CCA

CCA is a method for finding linear correlation relationships between two or more multidimensional datasets. CCA finds a canonical coordinate space that maximizes correlations between projections of the datasets into that space.

Given two co-occurring random variables with N observations collected as matrices $X1 \in R^{D1\times N}$ and $X2 \in R^{D2\times N}$, the task is to find linear projections $U \in R^{D1\times K}$ and $V \in R^{D2\times K}$ so that the correlation between $u_k^T X^{(1)}$ and $v_k^{(T)} X^{(2)}$ is maximized for the components k, under the constraint that $u_k^{(T)} X^{(1)}$ and $u_{k'}^{(T)} X^{(1)}$ are uncorrelated for all $k \neq k'$ (and similarly for the other view).

The solution can be found analytically by solving the eigen value problems :

$$\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}\mathbf{u} = \rho^2\mathbf{u},$$
$$\mathbf{C}_{22}^{-1}\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{v} = \rho^2\mathbf{v},$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}$$

is the joint covariance matrix of $x^{(1)}$ and $x^{(2)}$ and ρ denotes the canonical correlation.

BCCA

For Bayesian Analysis, the model needs to be complemented with priors for the model parameters. In Bayesian treatment of CCA , inverse Wishart distribution is used as a prior for the covariance matrices and automatic relevance distribution ARD as a prior for linear mappings.

Chapter 2

Mathematical Formula

The Bayesian approach of CCA is based on latent variable models and linear projections. The full model is as follows :

$$
\begin{aligned}
\mathbf{z} &\sim \mathrm{N}(\mathbf{0}, \mathbf{I}), \\
\mathbf{x}^{(m)} &\sim \mathrm{N}(\mathbf{A}^{(m)}\mathbf{z}, \mathbf{\Psi}^{(m)}), \\
\mathbf{A}^{(m)} &\sim \mathrm{ARD}(\alpha_0, \beta_0), \\
\mathbf{\Psi}^{(m)} &\sim \mathrm{IW}(\mathbf{S}_0, \nu_0),
\end{aligned}
$$

For Bayesian Analysis , the model needs to be complemented with priors for the model parameters. In Bayesian treatment of CCA inverse-Wishart distribution is used as a prior for the covariance matrices and apply the automatic relevance determination (ARD) prior for the linear mappings $A^{(m)}$.

The ARD is a Normal-Gamma prior for the projection weights. For each component (column) $a_k^{(m)}$ the prior specifies a precision $\alpha_k^{(m)}$ that controls the scale of the values for that component:

$$\mathrm{ARD}(\mathbf{A}^{(m)}|\alpha_0,\beta_0) = \prod_{k=1}^{K} p(\mathbf{a}_k^{(m)}|\boldsymbol{\alpha}_k^{(m)})p(\boldsymbol{\alpha}_k^{(m)}|\alpha_0,\beta_0),$$

$$\boldsymbol{\alpha}_k^{(m)} \sim \mathrm{Gamma}(\alpha_0,\beta_0),$$

$$\mathbf{a}_k^{(m)} \sim \mathrm{N}(\mathbf{0},(\boldsymbol{\alpha}_k^{(m)})^{-1}\mathbf{I}).$$

The hyperpriors α_0,β_0 are set to small values to $\alpha_0 = \beta_0 = 10^{-14}$
For the covariance matrices $\Psi(m)$ a natural choice is to use a conjugate inverse-Wishart prior

$$\boldsymbol{\Psi}^{(\mathbf{m})} \sim \mathbf{IW}(\mathbf{S_0},\boldsymbol{\nu})$$

with ν_0 degrees of freedom and scale matrix S_0

Chapter 3
Documentation of API

4.1 Package Organization

class BCCA (n_comp=2, reg_param=0.1)

Parameters:

n_comp represents the number of components
reg_param represents the regularization parameter

Attributes :

C_{aa} : covariance matrix of view a w.r.t view a
C_{bb} : covariance matrix of view b w.r.t view b
C_{ab} : covariance matrix of view a w.r.t view b

U,S,V : output of singular value decomposition
Covariance_matrix : covariance matrix for different views
View_len = number of views

4.2 Methods

__init__(self, n_comp, reg_param):

To initialize the class

Parameters:

n_comp: the number of components (Default=2)
reg_param: regularization parameter (Default=0.1)

fit(self , *x_list):
To fit the standardized data to BCCA so as to calculate weights and correlation of

transform(self,data):
To get the reduced data with the weights associated with it by returning dot product of the standardized data and weights.

calculate_covariance_matrix(self, x_list) :
To calculate covariance matrix of matrices

normalize (matrix) :
To normalize the matrix

Calculate_eigen :
To calculate eigen values and eigen vectors and sort them

Chapter 4

Example 1

A = [[0.4577126 0.51627482 0.0234379]
[0.78917247 0.10042776 0.24761369]
[0.54171431 0.58731726 0.93418298]]

B = [[0.45572318 0.95339909 0.27143393]
[0.56724832 0.89562746 0.0735327]
[0.073516   0.91303813 0.15807052]]

C_{aa} = [[0.37524947 0.21123949 0.23739934]
[0.21123949 0.207189   0.19520983]
[0.23739934 0.19520983 0.31151991]]

C_{bb} = [[0.17828629 0.33655072 0.05901025]
[0.33655072 0.84825233 0.15632239]
[0.05901025 0.15632239 0.03468991]]

C_{ab} = [[0.23202389 0.54593104 0.08929926]
[0.1118077 0.37280162 0.08011892]
[0.07327235 0.36568666 0.05741211]]

W_a = [[-3.40066422 -0.21503678 -0.02027057]
[-3.82515651 0.12630814 0.18812587]
[-3.52873808 0.24058348 -0.18761529]]

W_b = [[-5.04216015e+00 -3.21043685e-01 6.12069830e-03]
[-3.24749422e+00 2.64100402e-02 -1.82991067e-02]
[-7.20924014e+00 4.34157830e-01 9.18519788e-02]]

Chapter 5

Learning Outcomes

- Learnt how the classical CCA problem is solved through single value decomposition (SVD)

- Used the methods in the package on our locally generated data and got satisfactory results as expected from the analysis of the mathematical equations.

- Successfully analysed the concept of latent variable and how it is useful in Bayesian analysis of CCA

References:

1. Bayesian Canonical Correlation Analysis

   Research Paper by Arto Klami , Seppo Virtanen and Samuel Kaski

   https://jmlr.org/papers/volume14/klami13a/klami13a.pdf

2. Canonical Correlation Analysis (CCA) Based Multi-View Learning: An Overview by Chenfeng Guo, Dongrui Wu, https://arxiv.org/abs/1907.01693

3. A tutorial on Canonical correlation methods https://arxiv.org/pdf/1711.02391.pdf