Cross Regression
for
Multi-view Feature Extraction
(CRMvFE)

Utkarsh Shrivastava
0801CS171090

Sachin Motwani
0801CS171065

Contents

Chapter 1

Introduction

1.1 Multi-view Data

In today's world, data is often collected from diverse feature extractors. For example, an image can be represented by different types of papers, and a web page can be represented by hyperlinks and content texts. These different types of samples are called multi-view data.

1.2 Multi-view learning

Multi-view learning is an emerging direction in machine learning which considers learning with multiple views to improve the generalization performance. Multi-view learning is also known as data fusion or data integration from multiple feature sets. Multiview learning (MVL) [1–6] is proposed to integrate compatible and complementary information among different views, which has better performance than traditional single-view learning.

1.3 Multi-view Feature Extraction (MvFE)

Traditional feature extraction methods include Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). They are only suitable for single-view data and not for multi-view data.

For multi-view data, various MvFE methods are there which exploit the correlation information from multiple views when extracting features. Two of them typical methods are Canonical Correlation Analysis (CCA) and Partial Least Squares (PLS) which aim at maximising between-view correlation and covariance respectively.

Several other methods have been proposed but they are only suitable for two-view scenario. In order to deal with multi-view scenario, Multi-view CCA (MCCA) was proposed which tries to find a common space by maximizing the total canonical correlation coefficients between any two views.

Most MvFE methods have two characteristics:

1. They mainly explore the cross-view correlations in the projected subspace without considering the correlation information in original high dimensional space.
2. They are sensitive to the outliers due to using L2-norm or F-norm.

Recently, ridge regression (RR) has made a breakthrough, which uses the previous information (original data or label information) directly in the regression models.

Inspired by the regression strategy for feature extraction, this document constructs a novel cross-regression regularization term to discover the relationship between multiple views in original high-dimensional space. Firstly, minimizing the proposed regularization term aims at seeking a set of projection matrices to transform the samples from different views into a common low dimensional subspace. Then, another set of projection matrices is introduced to transform the low-dimensional samples back to the original high-dimensional space. Minimizing our proposed cross-regression regularization term is to minimize the distance between original data and the projected high-dimensional data.

Finally, two methods namely cross-regression for multi-view feature extraction (CRMvFE) and robust cross-regression for multi-view feature extraction (RCRMvFE) are proposed.

Chapter 2

Mathematical Formulation

2.1 Formulation

Given the data matrices:

$$X_A = [X_1^A, X_2^A, \ldots, X_n^A]^T \in R^{N \times DA},\ A \in \{1, \ldots, V\}$$

where X_A represents the sample matrix from the A^{th} view, x_i^A, $i = 1, \ldots, N$ represents the ith sample of the A^{th} view, D^A is the feature dimension in A^{th} view.

Given $d < D^A$, $A = 1, \ldots, V$, multi-view feature extraction aims to find the projection matrices $P_A \in R^{DA \times d}$, $A = 1, \ldots, V$, such that the samples from D^A dimensional space can be projected to d dimensional space.

2.2 Model of CRMvFE

CRMvFE preserves the correlation between multiple views directly by introducing a novel cross-regression regularization term and explores the correlation in single-view itself.

CRMvFE finds the projection matrices $P_A \in R^{D_A \times d}$, $A = 1, \ldots, V$, by solving the following optimization problem:

$$\begin{aligned} \min_{\substack{P_B, F_A \\ A,B=1,\ldots,V}} \quad & \sum_{A,B=1}^{V} \sum_{i=1}^{N} \left\| x_i^A - x_i^B P_B F_A \right\|_2^2 + \gamma \sum_{A=1}^{V} \left\| F_A \right\|_F^2 \\ \text{s.t.} \quad & P^{\mathrm{T}} P = I \end{aligned} \tag{9}$$

where $P = \left[P_1^{\mathrm{T}}, \ldots, P_V^{\mathrm{T}}\right]^{\mathrm{T}}$, $P_B \in R^{D_B \times d}$ and $F_A \in R^{d \times D_A}$ represent two sets of projection matrices. γ is the parameter, and $I \in R^{d \times d}$ is the identity matrix. A sample x_i^B in high-dimensional space of view-B is projected to a new sample $x_i^B P_B$ in the low-dimensional common subspace, and the sample $x_i^B P_B$ is projected back to $x_i^B P_B F_A$ in original high-dimensional space of view-A. The first term in problem (9) means minimizing the loss between original sample x_i^A and the projected sample $x_i^B P_B F_A$, the second term is the

regularization term. Let $X_A = \left[x_1^A, \ldots, x_N^A\right]^{\mathrm{T}}$, $F = \left[F_1^{\mathrm{T}}, \ldots, F_V^{\mathrm{T}}\right]^{\mathrm{T}}$, the optimization problem (9) can be rewritten as:

$$\begin{aligned} \min_{P,F} \quad & J(P, F) = \sum_{A,B=1}^{V} \| X_A - X_B P_B F_A \|_F^2 + \gamma \sum_{A=1}^{V} \| F_A \|_F^2 \\ \text{s.t.} \quad & P^{\mathrm{T}} P = I \end{aligned} \tag{10}$$

Note that the first term in the objective function in (10) indicates that the sample matrix X_A in the Ath view is expected to be reconstructed by the sample matrix X_B in the Bth view by two projection matrices P_B and F_A. If $A = B$, the first term explores the correlation in single-view itself; if $A \neq B$, it explores the correlations from different views. Therefore, CRMvFE learns a unified subspace by combining compatible and complementary information among single-view and multiple views simultaneously. And the second term $\alpha \sum_{A=1}^{V} \|F_A\|_F^2$ is applied to strengthen the model's stability in computing the optimal solution.

Chapter 3

Algorithm

3.1 Algorithm of CRMvFE

Although there are two variables, P and F in the optimization problem (10), we are only interested in the first part P^* of its solution (P^*, F^*). So we have the following theorem:

Theorem 1. *Let*

$$L = \begin{bmatrix} X_1^{\mathrm{T}}X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_V^{\mathrm{T}}X_V \end{bmatrix} \tag{11}$$

Consider the eigenvalue problem:

$$(L + \gamma I)^{-1}\left(X^{\mathrm{T}}\sum_{A=1}^{V} X_A X_A^{\mathrm{T}} X\right) p = \lambda p \tag{12}$$

where $X = [X_1, \ldots, X_V]$, $I \in R^{\sum_A D_A \times \sum_A D_A}$ *is the identity matrix. If* $p_1, \ldots, p_d$ *are the eigenvectors associated with the* d *largest eigenvalues, then the first part* P^* *of the solution* (P^*, F^*) *of problem* (10) *can be obtained by:*

$$P^* = \begin{pmatrix} P_1^* \\ \vdots \\ P_V^* \end{pmatrix} = (p_1, \ldots, p_d) \tag{13}$$

3.2 Proof

Proof. Suppose P in problem (10) is known and consider the following optimization problem with only one variable.

$$\begin{aligned} \min_{F} \quad & J(P, F) \\ \text{s.t.} \quad & P^{\mathrm{T}}P = I \end{aligned} \tag{14}$$

Since the constraint is not related to the variable F. (14) is actually an unconstrained optimization problem for the variable F.

Therefore, let $\partial J(P, F)/\partial F_A = 0$, we have:

$$\sum_{B=1}^{V} \mathrm{tr}\left(-2P_B^{\mathrm{T}}X_B^{\mathrm{T}}X_A + 2P_B^{\mathrm{T}}X_B^{\mathrm{T}}X_B P_B F_A\right) + 2\gamma\,\mathrm{tr}(F_A) = 0 \tag{15}$$

Thus, the solution

$$F = F(P) = \left[F_1^{\mathrm{T}}, \ldots, F_V^{\mathrm{T}}\right]^{\mathrm{T}} \tag{16}$$

can be given by:

$$F_A = H^{-1}\left(\sum_{B=1}^{V} P_B^{\mathrm{T}}X_B^{\mathrm{T}}X_A\right), A = 1, \ldots, V. \tag{17}$$

where

$$H = \sum_{B=1}^{V} P_B^{\mathrm{T}}X_B^{\mathrm{T}}X_B P_B + \gamma I \tag{18}$$

Substituting (16), (17) back to (10) yields the optimization problem with the variable P,

$$\begin{aligned} \min_{P} \quad & J(P, F(P)) \\ \text{s.t.} \quad & P^{\mathrm{T}}P = I \end{aligned} \tag{19}$$

where $F(P)$ is given by (16), (17). It is easy to see that the solution P^* of (19) should be the first part of the solution (P^*, F^*). The optimization problem (19) can be written as:

$$
\begin{aligned}
&\min_{\substack{P^{\mathrm{T}}P=I, F_A\\ A,B=1,\ldots,V}} && \sum_{A,B=1}^{V} \|X_A - X_B P_B F_A\|_F^2 + \gamma \sum_{A=1}^{V} \|F_A\|_F^2 \\
\Leftrightarrow &\min_{\substack{P^{\mathrm{T}}P=I, F_A\\ A,B=1,\ldots,V}} && \operatorname{tr}\left(\sum_{A,B=1}^{V} \left(-2F_A^{\mathrm{T}} P_B^{\mathrm{T}} X_B^{\mathrm{T}} X_A + F_A^{\mathrm{T}} P_B^{\mathrm{T}} X_B^{\mathrm{T}} X_B P_B F_A\right) + \gamma \sum_{A=1}^{V} F_A^{\mathrm{T}} F_A \right) \\
\Leftrightarrow &\min_{P^{\mathrm{T}}P=I} && \operatorname{tr}\left(\sum_{A,B=1}^{V} \left(-2 \left(\sum_{c=1}^{V} X_A^{\mathrm{T}} X_c P_c H^{-1} \right) P_B^{\mathrm{T}} X_B^{\mathrm{T}} X_A \right.\right. \\
& && + \left(\sum_{c=1}^{V} X_A^{\mathrm{T}} X_c P_c H^{-1} \right) P_B^{\mathrm{T}} X_B^{\mathrm{T}} X_B P_B \left(H^{-1} \sum_{c=1}^{V} P_c^{\mathrm{T}} X_c^{\mathrm{T}} X_A \right) \\
& && \left.\left. + \gamma \sum_{A=1}^{V} \left(\sum_{c=1}^{V} X_A^{\mathrm{T}} X_c P_c H^{-1} \right) \left(H^{-1} \sum_{c=1}^{V} P_c^{\mathrm{T}} X_c^{\mathrm{T}} X_A \right) \right) \right) \\
\Leftrightarrow &\min_{P^{\mathrm{T}}P=I} && \operatorname{tr}\left(\sum_{A=1}^{V} X_A^{\mathrm{T}} \left(\sum_{c=1}^{V} X_c P_c \right) H^{-1} \left(-2H + \sum_{B=1}^{V} P_B^{\mathrm{T}} X_B^{\mathrm{T}} X_B P_B \right.\right. \\
& && \left.\left. + \gamma I \right) H^{-1} \left(\sum_{c=1}^{V} P_c^{\mathrm{T}} X_c^{\mathrm{T}} \right) X_A \right) \\
\Leftrightarrow &\min_{P^{\mathrm{T}}P=I} && \operatorname{tr}\left(\sum_{A=1}^{V} X_A^{\mathrm{T}} \left(\sum_{c=1}^{V} X_c P_c \right) H^{-1} (-H) H^{-1} \left(\sum_{c=1}^{V} P_c^{\mathrm{T}} X_c^{\mathrm{T}} \right) X_A \right) \\
\Leftrightarrow &\max_{P^{\mathrm{T}}P=I} && \operatorname{tr}\left(H^{-1} P^{\mathrm{T}} X^{\mathrm{T}} \sum_{A=1}^{V} X_A X_A^{\mathrm{T}} X P \right) \\
\Leftrightarrow &\max_{P^{\mathrm{T}}P=I} && \operatorname{tr}\left(\left(P^{\mathrm{T}}(L+\gamma I)P\right)^{-1} \times \left(P^{\mathrm{T}} X^{\mathrm{T}} \sum_{A=1}^{V} X_A X_A^{\mathrm{T}} X P \right) \right)
\end{aligned}
\tag{20}
$$

where $P = \left[P_1^{\mathrm{T}}, \ldots, P_V^{\mathrm{T}}\right]^{\mathrm{T}}$, $X = [X_1, \ldots, X_V]$ and $H = \sum_{B=1}^{V} P_B^{\mathrm{T}} X_B^{\mathrm{T}} X_B P_B + \gamma I$.

It is easy to find that the optimization problem (20) can be transformed into the eigenvalue problem (12) [37]. Therefore, the optimal solution P^*, F^* of problem (10) can be obtained by (12) and (13). □

The subspace feature Z can be extracted by : $Z = X_1P_1 + \cdots + X_VP_V$. The CRMvFE algorithm is shown in Algorithm 1.

Algorithm for CRMvFE

Input: The dataset from multiple views $X_A = \left[x_1^A, x_2^A, \cdots, x_N^A\right]^{\mathrm{T}} \in R^{N\times D_A}$, $A \in \{1, \cdots, V\}$ and the parameter γ .

1: Construct $X = [X_1, \cdots, X_A]$. Compute $O = (L + \gamma I)^{-1} \left(X^{\mathrm{T}} \sum_{A=1}^{V} X_A X_A^{\mathrm{T}} X\right)$ by (12);

2: Solve the eigenvalue problem $Op = \lambda p$, select the eigenvectors corresponding to the d largest eigenvalues to construct $P = \left[P_1^{\mathrm{T}}, \cdots P_A^{\mathrm{T}}, \cdots, P_V^{\mathrm{T}}\right]^{\mathrm{T}} = (p_1, \cdots, p_d)$;

3: Obtain the final projection matrices $P_A, A = 1, \cdots, V$. Compute the low-dimensional projection $Z = \sum_{A=1}^{V} X_A P_A$.

Output: The low dimensional embedding Z .

Chapter 4

Documentation of API

To be done since the implementation is not completed yet

Chapter 5

Example

To be done since the implementation is not completed yet

Chapter 6

Learning Outcomes

- Regression based methods for feature extraction using a single view
- Understand and deal with multi-view data

    - ❖ Understand and apply the concepts of cross regression for multi-view data

    - ❖ Understand the relationship between multiple views and also obtain low dimensional matrices for different views

Chapter 7

References

- J. Zhang, P. Zhang, L. Liu, N. Deng, L. Jing, Collaborative weighted multi-view feature extraction, Eng. Appl. Artif. Intell. 90 (2020) 1–10.
- He, J. Shao, J. Zhang, X. Zhou, Clustering based multiple instance learning with multiview feature, Expert Syst. Appl. (2019) 113027.
- Jinxin Zhang, LingJing, JunyanTan, Cross-regression for multi-view feature extraction, Knowledge-Based Systems 200 (2020) 105997