Prediction Error Computation on a Grid

Wallace E. Larimore  
Adaptics, Inc.  
McLean, VA 22101 USA

Franklin T. Luk  
Department of Computer Science  
Rensselaer Polytechnic Institute  
Troy, New York 12180 USA

Sanzheng Qiao  
Department of Computing and Software  
McMaster University  
Hamilton, Ontario L8S4L7 Canada

Abstract

This paper presents a distributed algorithm for prediction error computation.

1 Introduction

In statistical computation, the inverse or pseudoinverse of a covariance matrix is often needed. When random variables are added, the order of the covariance matrix is increased. It is desirable to efficiently update the pseudoinverse when the order of the covariance matrix is increased. Instead of updating the pseudoinverse, Larimore [2, 3] proposed a method for updating a factor of the pseudoinverse. Let $y$ and $x_1$ be two vectors of random variables and $Y$ and $X_1$ the two corresponding data matrices such that the covariance matrices

$$S_{yx_1} = Y^T X_1 \quad \text{and} \quad S_{11} = X_1^T X_1.$$ 

Consider an estimation problem where the minimum variance unbiased estimate of $y$ from $x_1$ is given by

$$\hat{y} = S_{yx_1} S_{11}^{-1} x_1$$

and the prediction error variance is

$$E[(y - \hat{y})(y - \hat{y})^T] = S_{yy} - S_{yx_1} S_{11}^{-1} S_{x_1 y}.$$ 

1
The covariance matrix $S_{11}$ is symmetric and positive semidefinite. In general, $S_{11}$ may be singular. In this general case, Larimore [2, 3] has shown that if rank($S_{11}$) = $r_1$ and a matrix $M_1$ satisfies

$$M_1^T S_{11} M_1 = I_{r_1},$$

then the prediction error variance can be expressed by $S_{yy} - S_{yx} M_1 M_1^T S_{x,y}$. Actually, since $(M_1 M_1^T) S_{11} (M_1 M_1^T) = M_1 M_1^T$, $M_1$ is a factor in the $\{2\}$-inverse $M_1 M_1^T$ of $S_{11}$. How to find such $M_1$? Suppose that

$$X_1 = [U_1 \hat{U}_1] \begin{bmatrix} R_1 & 0 \\ 0 & 0 \end{bmatrix} [V_1 \hat{V}_1]^T$$

is the complete orthogonal decomposition [1] of $X_1$, we call $X_1 = U_1 R_1 V_1^T$ the economy version of the complete orthogonal decomposition. It then can be verified that $M_1 = V_1 R_1^{-1}$ satisfies (1).

Now, a vector $x_2$ of $n_2$ random variables is added to the $n_1$-vector $x_1$ to form $x = [x_1^T x_2^T]^T$. If the covariance matrix is $S = X^T X$, rank($S$) = $r$, and $M$ satisfies $M^T S M = I_r$, then the prediction error variance is $S_{yy} - S_{yx} M M^T S_{xy}$. Since the covariance matrices $S_{yx}$ and $S_{xy}$ can be easily obtained by updating $S_{yx_1}$ and $S_{x_1 y}$, the problem remains to compute $M$ from $M_1$. Specifically, given $M_1$ satisfying $M_1^T S_{11} M_1 = I_{r_1}$, find $M$ such that $M^T S M = I_r$. An updating method for the $M$ matrix is presented in Section 2. Then in Section 3, assuming the samples of random variables are collected over various locations, we give a distributed algorithm for the prediction error computation.

2 An Updating Scheme

Let the covariance matrix

$$S = [X_1 \ X_2]^T [X_1 \ X_2] = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix},$$

where $X_1$ is $m$-by-$n_1$ and $X_2$ is $m$-by-$n_2$, $m > n_1 + n_2$. Suppose that

$$X_2 = U_2 R_2 V_2^T$$

is the economy version of the complete orthogonal decomposition of $X_2$, where $R_2$ is $r_2$-by-$r_2$ upper triangular and nonsingular. Thus, the matrix $M_2 = V_2 R_2^{-1}$ satisfies $M_2^T S_{22} M_2 = I_{r_2}$. In practice, it is rare that a matrix

2
is exactly singular. In this case, we set the small blocks in the $R$ matrix in the complete orthogonal decomposition to zero and then get the economy version.

Suppose that

$$(X_1M_1)^T(X_2M_2) = Q_1\Sigma Q_2^T$$

is the full SVD, where $Q_1$ and $Q_2$ are orthogonal and of orders $r_1$ and $r_2$ respectively and

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ \vdots & \vdots \\ 0 & \sigma_{r_2} \end{bmatrix},$$

assuming $r_1 \geq r_2$. Letting

$$J_1 = M_1Q_1 \quad \text{and} \quad J_2 = M_2Q_2,$$

we have the following equations

$$J_1^T S_{11} J_1 = Q_1^T M_1^T S_1 M_1 Q_1 = I_{r_1}$$
$$J_2^T S_{22} J_2 = Q_2^T M_2^T S_2 M_2 Q_2 = I_{r_2}$$
$$J_1^T S_{12} J_2 = Q_1^T M_1^T X_1 X_2 M_2 Q_2 = \Sigma.$$

Finally, the block upper triangular matrix

$$M = \begin{bmatrix} M_1 & -J_1\Sigma D \\ 0 & J_2 D \end{bmatrix},$$

where $D = \text{diag} (\sqrt{1 - \sigma_1^2}, ..., \sqrt{1 - \sigma_{r_2}^2})$ so that $D^T (I - \Sigma^T \Sigma) D = I_{r_2}$, satisfies

$$M^T S M = \begin{bmatrix} M_1 & -J_1\Sigma D \\ 0 & J_2 D \end{bmatrix}^T \begin{bmatrix} S_{11} & S_{12} \\ S_{12}^T & S_{22} \end{bmatrix} \begin{bmatrix} M_1 & -J_1\Sigma D \\ 0 & J_2 D \end{bmatrix} = \begin{bmatrix} Q_1^T & -\Sigma D \\ 0 & D \end{bmatrix} \begin{bmatrix} I_{r_1} & \Sigma \\ \Sigma^T & I_{r_2} \end{bmatrix} \begin{bmatrix} Q_1^T & -\Sigma D \\ 0 & D \end{bmatrix} = \begin{bmatrix} Q_1 & Q_1 \Sigma \\ 0 & D^T (I - \Sigma^T \Sigma) \end{bmatrix} \begin{bmatrix} Q_1^T & -\Sigma D \\ 0 & D \end{bmatrix}.$$
\[ 
\begin{bmatrix}
I_r \\
0
\end{bmatrix} = 
\begin{bmatrix}
0 \\
D^T(I - \Sigma^T \Sigma)D
\end{bmatrix} = I_r.
\]

As shown in [3], \( r = r_1 + r_2. \)

3 Distributing the Computation

Suppose that the data matrix \( X \) is distributed among \( k \) locations: \( X = [X_1, ..., X_k] \). Location \( i \) has a block \( X_i \) of columns and computes a column block \( M_i \) of \( M = [M_1, ..., M_k] \) and a column block \( Y_i \) of \( XM = [Y_1, ..., Y_k] \). Note that \( M \) is block upper triangular. The algorithm for location \( i \) is as follows.

**Algorithm**

1. Get \( Y_1, ..., Y_{i-1} \);
2. Get \( M_1, ..., M_{i-1} \);
3. Complete orthogonal decomposition \( X_i = U_i R_i V_i^T \);
4. SVD \( [Y_1, ..., Y_{i-1}]^T U_i = Q_1 \Sigma Q_2^T \), \( \Sigma = \text{diag}(\sigma_j) \);
5. \( J_1 = [M_1, ..., M_{i-1}] Q_1 \);
6. \( J_2 = V_i R_i^{-1} Q_2 \);
7. \( M_i = \begin{bmatrix} -J_1 \Sigma D \\ J_2 D \end{bmatrix} \), where \( D = \text{diag}(\sqrt{1 - \sigma_j^2}) \);
8. \( Y_i = -[X_1, ..., X_{i-1}] J_1 \Sigma D + X_i J_2 D = -[Y_1, ..., Y_{i-1}] Q_1 \Sigma D + X_i J_2 D \).

References

