Let's say we have two mean-centered matrices, X and Y, correspond to training data and targets with n x m and n x d dimensions, respectively. Each of which may have noises (small compared to data) or collinearity, so we need a dimension-reduced data as representative to the original ones as possible. To achieve this, we search for a pair of rotation matrices, P and Q, to map X and Y respectively to new coordinates that contribute to the maximum of cross covariance between new X and Y, which is

$$\max_{P,Q} \mathrm{Cov}[XP, YQ] \qquad (1)$$

With the help of singular value decomposition (s.v.d.), we can find the suitable coordinate pair not only maximize the cross covariance, but also uncorrelate new X and Y. Then eq. (1) becomes:

$$\max_{P,Q} \mathrm{Cov}[XP, YQ] \propto (XP)^T(YQ) = P^TX^TYQ = P^TU\Sigma V^TQ \qquad (2)$$

The s.v.d. of X^TQ in eq. (2) is left eigenvector matrix U, singular value matrix Σ, and right eigenvector matrix V. Since P, U, V, and Q are unit vector matrices, every element in each of them ranges from 0 to 1. To maximize eq. (2),

$$\begin{cases} P^TU = I_m \\ V^TQ = I_d \end{cases} \rightarrow \begin{cases} P = U \\ V = Q \end{cases} \qquad (3)$$

and the new cross correlation Σ is a diagonal matrix. So, the transformed X and Y are

$$\begin{cases} X_T = XU \\ Y_T = YV \end{cases} \rightarrow \begin{cases} X = X_TU^T \\ Y = Y_TV^T \end{cases} \qquad (4)$$

With the maximum cross correlation on each axis provided by Σ, we choose the first k columns of U to be U_k (hence X to be X_k) and V to be V_k (hence Y to be Y_k) correspond to the largest k eigenvalues in Σ, which are sufficient to refer to principal components in both X and Y to eliminate noises and collinearities. Replace with dimension-reduced matrices, eq. (4) becomes

$$\begin{cases} X_{T_k} = XU_k \\ Y_{T_k} = YV_k \end{cases} \rightarrow \begin{cases} X \approx X_k = X_{T_k}U_k^T \\ Y \approx Y_k = Y_{T_k}V_k^T \end{cases} \qquad (5)$$

Since we have dimension-reduced training data and responses X_{T_k} and Y_{T_k}, we use them to do least square regression to find coefficient matrix β_{T_k} so that the prediction of Y_{T_k}

$$\hat{Y}_{T_k} = X_{T_k}\beta_{T_k} \qquad (6)$$

With the help of linear algebra, we get

$$\beta_{T_k} = (X_{T_k}^TX_{T_k})^{-1}X_{T_k}^TY_{T_k} \qquad (7)$$

Then, the prediction of Y given X by eq. (5) – (6)

$$\hat{Y} \approx \hat{Y}_k = \hat{Y}_{T_k}V_k^T = X_{T_k}\beta_{T_k}V_k^T = XU_k\beta_{T_k}V_k^T = X\beta$$

where

$$\beta = U_k\beta_{T_k}V_k^T$$