

CRIMINALS

Jan de Leeuw

May 9, 2026

TBD

Table of contents

1	Between/Within Decomposition	3
2	Matrix Approximation	5
3	Meet Loss	6
	References	7

Note: This is a working manuscript which will be expanded/updated frequently. All suggestions for improvement are welcome. All Rmd, tex, html, pdf, R, and C files are in the public domain. Attribution will be appreciated, but is not required. The files can be found at <https://github.com/deleeuw/lda>

1 Between/Within Decomposition

Suppose X is an $n \times m$ matrix of rank m and G is an $n \times k$ indicator matrix indicating group membership. Define $D := G'G$, the diagonal matrix with the group frequencies, and $T_X := X'X$. We suppose G and D are non-singular, i.e. all groups have at least one member.

We have the decomposition

$$X = X_B + X_W := (GD^{-1}G')X + (I - GD^{-1}G')X \quad (1)$$

into the between-component of group averages X_B and the within-component of deviations from group averages X_W .

The matrix $P_G := GD^{-1}G'$ is a symmetric projector of rank k , which is idempotent with $P_G^2 = P_G$. Thus also $P_G(I - P_G) = 0$, and the two components are orthogonal in the usual inner product on $\mathbb{R}^{n \times m}$. Consequently

$$T_X = B_X + W_X := X'P_GX + X'(I - P_G)X \quad (2)$$

Now apply this decomposition to p linear combinations XA , with A an $n \times p$ matrix. Here $p < m$ i.e. the number of “dimensions” is smaller than the number of groups. Then $A'T_XA = A'B_XA + A'W_XA$. In discriminant analysis we find A such that $A'T_XA = I$ and $\text{tr } A'B_XA$ is maximized. In words: we find orthogonal directions in which the between-group averages are as far apart as possible (and thus the within-group variation is as small as possible).

Use a symmetric matrix M of Lagrange multipliers of order p . The stationary equations are

$$B_XA = T_XAM, \quad (3a)$$

$$A'T_XA = I. \quad (3b)$$

Premultiplying (3a) by A' shows $M = A'B_XA$, and thus M is positive semi-definite of rank p . Its eigen-decomposition is $M = K\Lambda^2K'$. Now

$$B_X(AK) = T_X(AK)\Lambda^2, \quad (4)$$

and AK are eigenvalues of the generalized eigen-problem defined by the pair (B_X, T_X) . Premultiplying (4) by $(AK)'$ shows that the between-groups variation is maximized by choosing the p largest eigenvalues $\lambda_1 \geq \dots \geq \lambda_p$. If $\lambda_p = \lambda_{p+1}$ there is some indeterminacy, and we simply choose the appropriate number of A -orthonormal eigenvectors from the eigen subspace associated with λ_p . In any case the solution for A is only determined up to a rotation.

If $\text{rank}(X) = r < m$ then B_X and T_X have the same null-space as X and the eigen equation (3a) has $m - r$ linearly independent solutions for which both B_XA and T_XA are zero. The corresponding eigenvalues are $0/0$, i.e. undefined, and the solutions cannot be normalized by (3b).

[1] "Matrix T_X"

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	+102.168333	-6.322667	+189.873000	+76.924333
Sepal.Width	-6.322667	+28.306933	-49.118800	-18.124267
Petal.Length	+189.873000	-49.118800	+464.325400	+193.045800
Petal.Width	+76.924333	-18.124267	+193.045800	+86.569933

[1] "Matrix B_X"

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	63.21213	-19.95267	165.2484	71.27933
Sepal.Width	-19.95267	11.34493	-57.2396	-22.93267
Petal.Length	165.24840	-57.23960	437.1028	186.77400
Petal.Width	71.27933	-22.93267	186.7740	80.41333

[1] "Matrix W_X"

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	38.9562	13.6300	24.6246	5.6450
Sepal.Width	13.6300	16.9620	8.1208	4.8084
Petal.Length	24.6246	8.1208	27.2226	6.2718
Petal.Width	5.6450	4.8084	6.2718	6.1566

2 Matrix Approximation

In this section we take a different route to the same equations. Consider the matrix approximation problem of minimizing

$$\sigma(U, V) = \text{tr} (X'X)^{-1}(X'G - UV')D^{-1}(X'G - UV')'. \quad (5a)$$

Define $\tilde{U} := (X'X)^{-\frac{1}{2}}U$ and $\tilde{V} := D^{-\frac{1}{2}}V$. The loss function (5a) can also be written as

$$\sigma(\tilde{U}, \tilde{V}) = \|(X'X)^{-\frac{1}{2}}X'GD^{-\frac{1}{2}} - \tilde{U}\tilde{V}'\|^2, \quad (5b)$$

with $\|\cdot\|^2$ the unweighted sum of squares.

Suppose $(X'X)^{-\frac{1}{2}}X'GD^{-\frac{1}{2}} = K\Lambda L'$ is a singular value decomposition. By the Eckart-Young theorem (Eckart and Young (1936)) the solution minimizing (5b) is

$$\tilde{U}\tilde{V}' = K_p\Lambda_p L_p', \quad (6a)$$

with Λ_p the p largest singular values, and with K_p and L_p the corresponding left and right singular vectors.

It follows that the solution minimizing (5a) is

$$UV' = (X'X)^{-\frac{1}{2}}K_p\Lambda_p L_p' D^{-\frac{1}{2}} = \tilde{K}_p \tilde{L}_p'. \quad (6b)$$

with $\tilde{K}_p := (X'X)^{-\frac{1}{2}}K_p$ and $\tilde{L}_p := D^{-\frac{1}{2}}L_p\Lambda_p$. Equation (6b) only determines U and V up to a non-singular transformation. The obvious choice is $U = \tilde{K}_p$ and $V = \tilde{L}_p$.

We now relate this result to the previous section. From the singular value equations

$$(X'X)^{-\frac{1}{2}}X'P_G X(X'X)^{-\frac{1}{2}}K_p = K_p\Lambda_p^2, \quad (7a)$$

or

$$X'P_G X\tilde{K}_p = X'X\tilde{K}_p\Lambda_p^2. \quad (7b)$$

Also $\tilde{K}_p'X'X\tilde{K}_p = K_p'K_p = I$. Thus $U = \tilde{K}_p$ from this section is A from the previous section (up to a rotation). For $V = \tilde{L}_p$ we find from the singular value equations

$$\tilde{L}_p = D^{-1}G'XA. \quad (8)$$

3 Meet Loss

$$\sigma(Z, U, V) := \|Z - XU\|^2 + \|Z - GV\|^2$$

$$U = (X'X)^{-1}X'ZV = D^{-1}G'V$$

$$\sigma_*(Z) := \min_{U, V} \sigma(Z, U, V) = Z'(I - X(X'X)^{-1}X')Z + Z'(I - P_G)Z$$

$$\sigma^*(U, V) := \min_Z \sigma(Z, U, V) =$$

$$\min_{Z'Z=I} \sigma_*(Z) = \min_{U'X'XU+V'DV=I} \sigma^*(U, V)$$

References

Eckart, C., and G. Young. 1936. "The Approximation of One Matrix by Another of Lower Rank." *Psychometrika* 1 (3): 211–18.