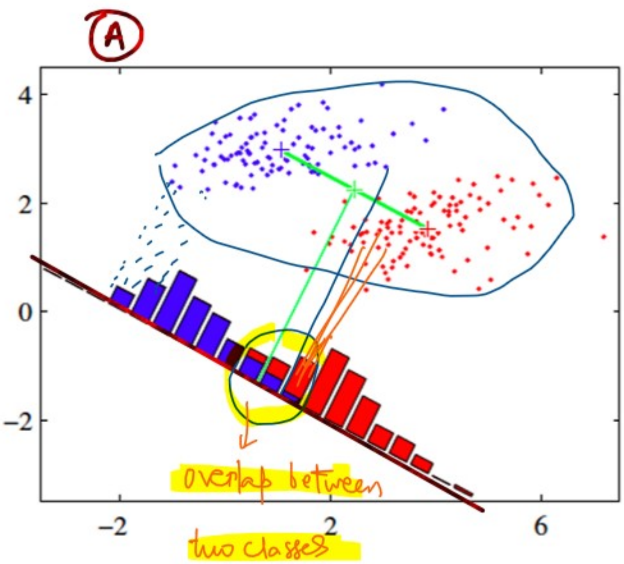
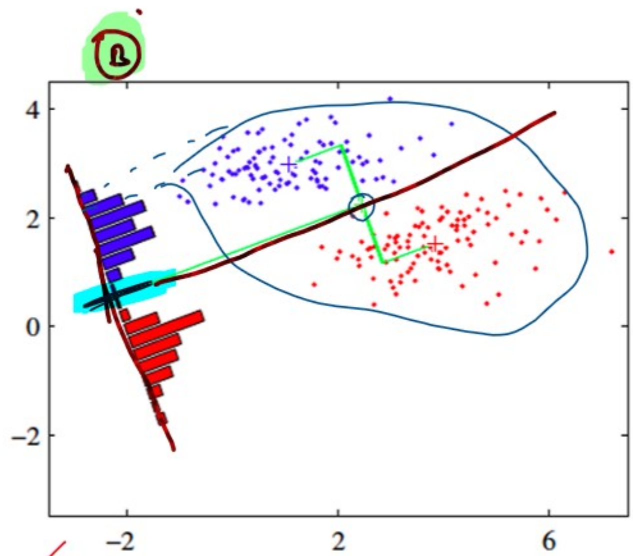
Linear Discriminant Analysis

28 April 2024 20:22

(LDA)





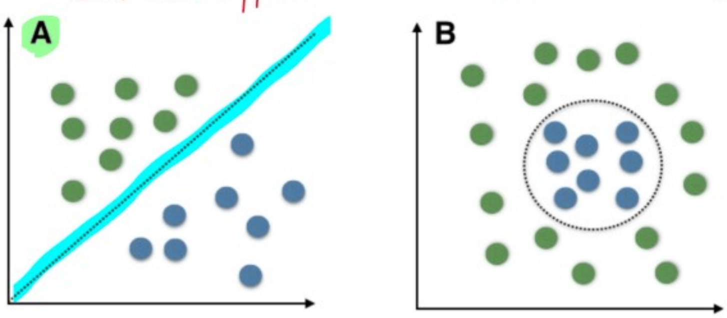
Note: The right plot shows the corresponding projection with improved class separation

\# Fisher Linear Discriminant Analysis

\# Linear classifier

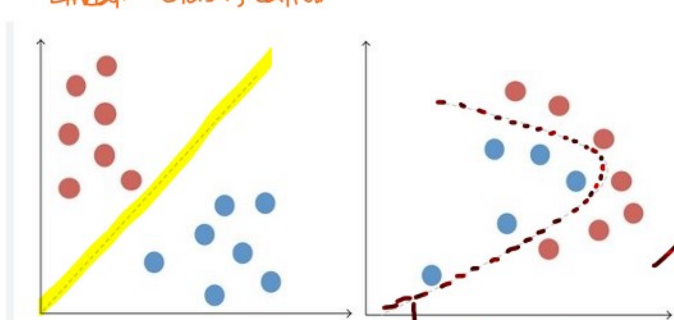
LDA can be applied

'LDA can't be applied'



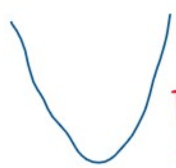
A: Linearly Separable Data B: Non-Linearly Separable Data

Linear Classification



~~LDA~~

LDA



Function

$\boxed{ax^2+bx+c=0}$ (Quadratic Equation)

$(a \neq 0)$

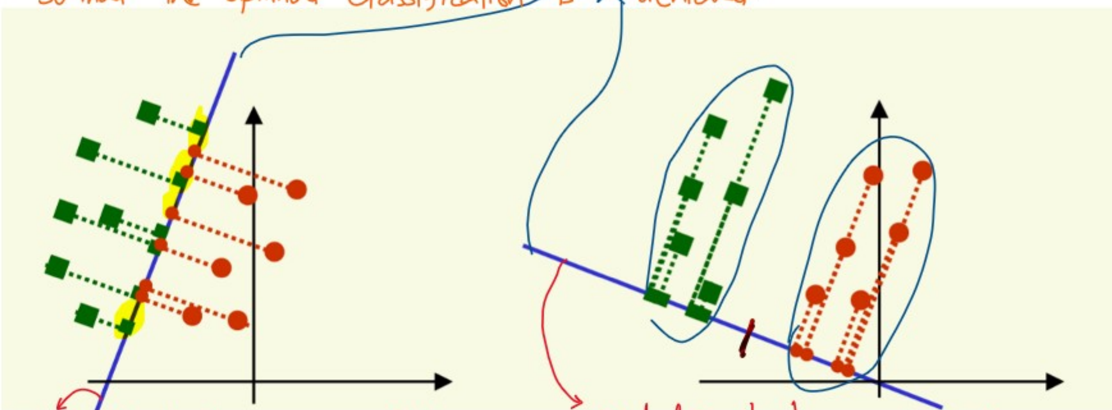
$y^2 = 4ax$ Latus = 4a

LDA

$y^2 = 4ax$ Latus = 4a

$x^2 = 4ay$

LDA algorithm tries to find the optimal blue line so that the optimal classification is achieved.



bad line to project the data having two classes → are mixed up.

good line to project the data having two classes → are well separated.

LDA is a supervised technique (ML- supervised learning) used for classification.

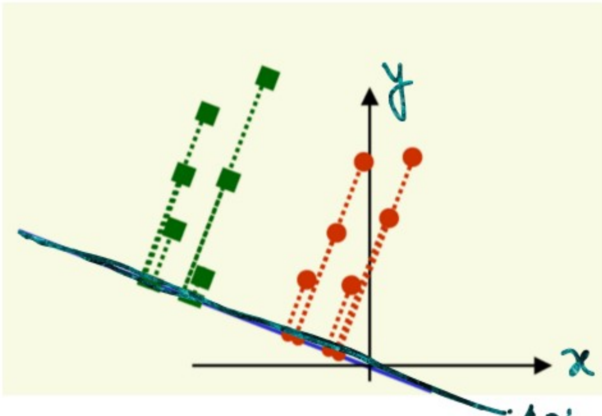
It is often employed when the objective is to find a linear combination of features that best separates two or more classes in a dataset.

Linear classifier

Application

Dimensionality Reduction

LDA reduces the number of features while preserving most of the class discriminatory information.



1 Dimension.

⇓

Two classes well separated.

Why do we need dimensionality reduction?

Memory efficient: space reqd. to store/process the data is reduced as the no. of dimensions come down.

Lesser training time: less dimensions lead to lesser computation/training time.

2-3 weeks → reduce → 2-3 days

- model training time

⇓

(less compute resource)

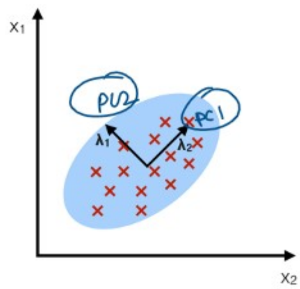
Multicollinearity: LDA takes care of multicollinearity by removing the redundant features.

Optimal separation: LDA maximizes the separation between classes and also it is computationally efficient and relatively simple to implement.

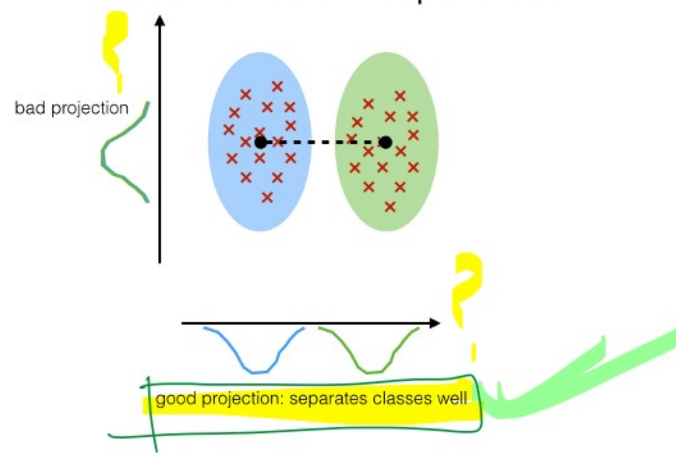
1936 → Fisher

PCA vs LDA

PCA:
component axes that maximize the variance

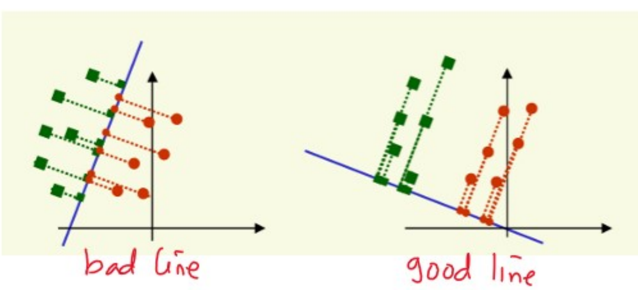


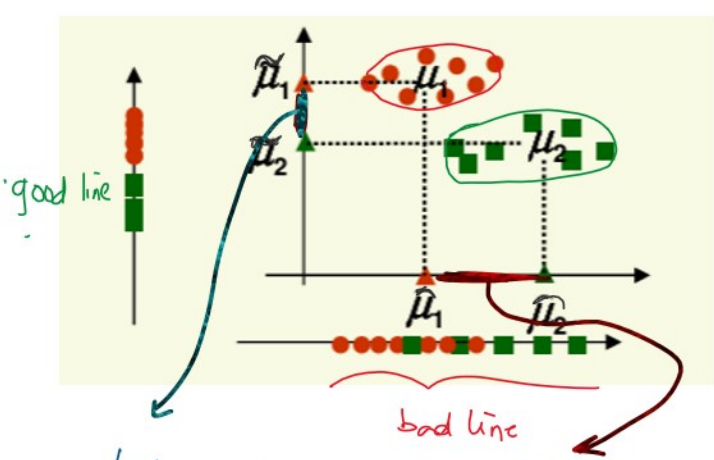
LDA:
maximizing the component axes for class-separation



How does LDA work?

Fisher LDA projects to a line which preserves direction useful for data classification



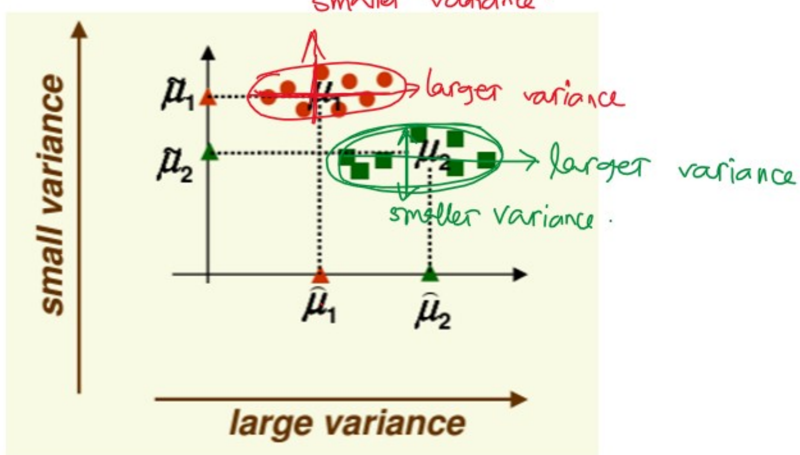


$$|\tilde{\mu}_1 - \tilde{\mu}_2| < |\hat{\mu}_1 - \hat{\mu}_2|$$

Larger $|\tilde{\mu}_1 - \tilde{\mu}_2|$, the better is the expected separation

however, $|\hat{\mu}_1 - \hat{\mu}_2| > |\tilde{\mu}_1 - \tilde{\mu}_2|$

smaller variance.



Problem with $|\tilde{\mu}_1 - \tilde{\mu}_2|$ is that it doesn't consider the variance of the classes

But we need to normalize $|\tilde{\mu}_1 - \tilde{\mu}_2|$ by a factor which is proportional to variance

sample #1 (class#1) Mean.

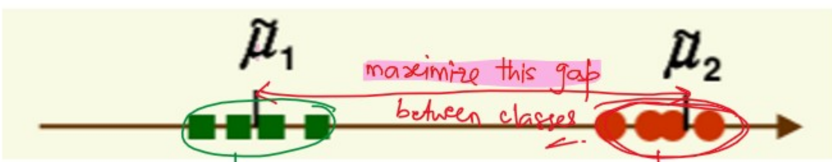
$$\boxed{\underbrace{J(v)}_{\text{Maximize}} = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2 \rightarrow \text{Maximize}}{(\tilde{s}_1^2 + \tilde{s}_2^2) \rightarrow \text{Minimize}}}$$

sample #2 (Class#2) Mean

sample variance for class#1

sample variance for class#2

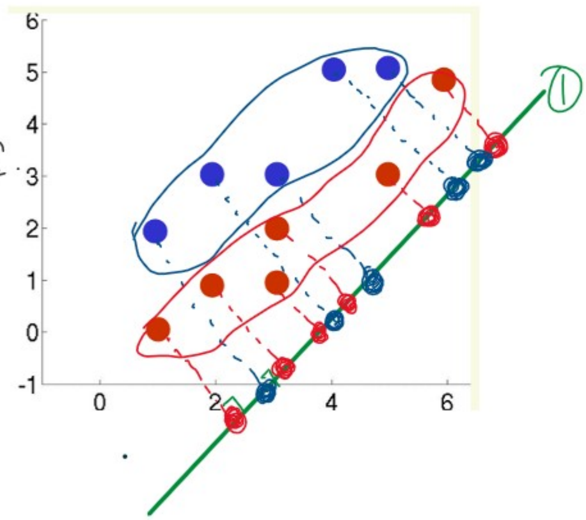
We want to find 'v' which makes J(v) large so that the classes are well separated we want to find v which makes J(v) large so that the classes are well separated

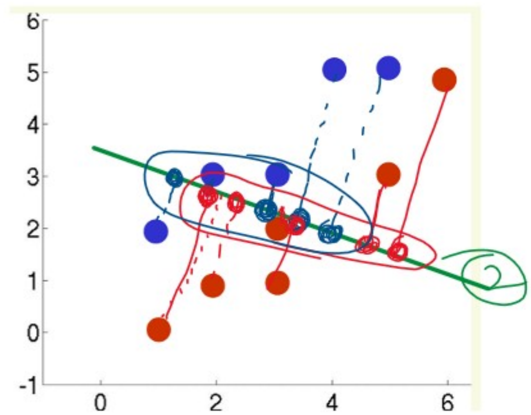


minimize variance within Class

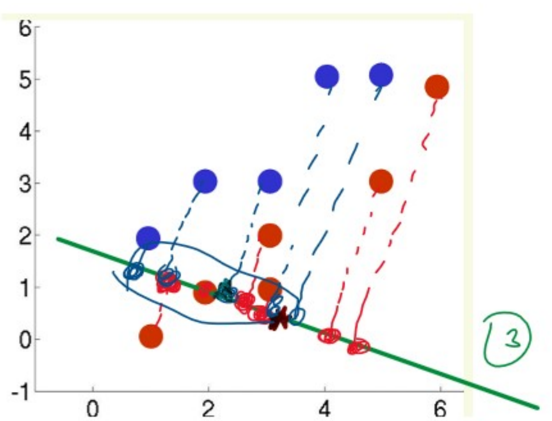
minimize variance within class

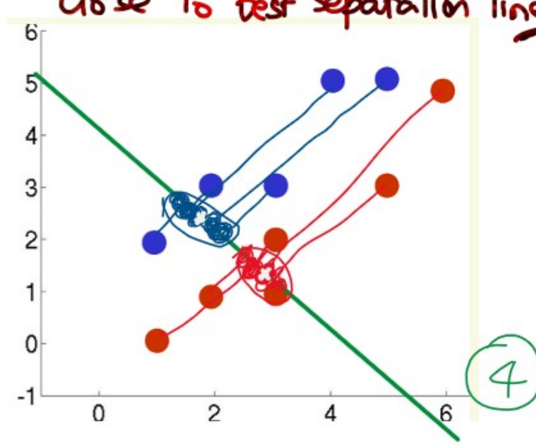
worst Project Line ever.





close to best separation line.





Linear discriminant line (2D)

⇓

linear discriminant hyperplane (3D+)

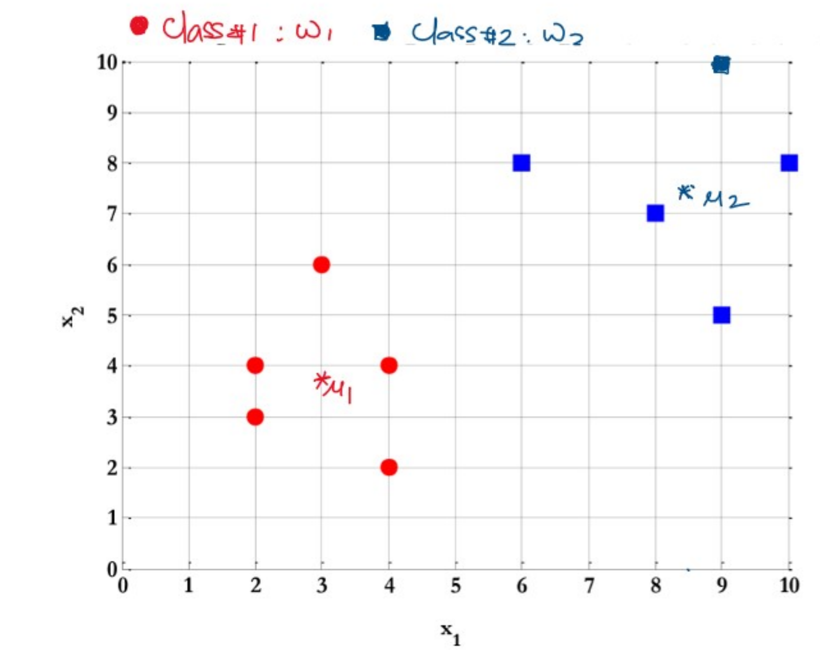
Q # Compute the LDA projection for following two-dimensional dataset:

C=2 (Two classes):

samples for class ω_1: $X_1 = (x_1, x_2) = \{(4,2), (2,4), (2,3), (3,6), (4,4)\}$

Two variables

samples for class ω_2: $X_2 = (x_1, x_2) = \{(9,10), (6,8), (9,5), (8,7), (10,8)\}$



Step # 1 Classes' Means

$$X_1 = \begin{array}{c} \begin{array}{cc} x_1 & x_2 \end{array} \\ \begin{bmatrix} 4 & 2 \\ 2 & 4 \\ 2 & 3 \\ 3 & 6 \\ 4 & 4 \end{bmatrix} \end{array} \qquad X_2 = \begin{array}{c} \begin{array}{cc} x_1 & x_2 \end{array} \\ \begin{bmatrix} 9 & 10 \\ 6 & 8 \\ 9 & 5 \\ 8 & 7 \\ 10 & 8 \end{bmatrix} \end{array}$$

$$\mu_1 = \begin{bmatrix} 3 \\ 3.8 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix}$$

Step # 2

Covariance matrix (scatter matrix) of 1st class: ω_1

$$S_1 = \sum_{x \in \omega_1} (x - \mu_1)(x - \mu_1)^T$$

$$(x - \mu_1)(x - \mu_1) = \boxed{(x - \mu_1)^2}$$

$$\mu_1 = \begin{bmatrix} 3 \\ 3.8 \end{bmatrix}$$

$$x_1 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

$$x_1 - \mu_1 = \begin{bmatrix} 4 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ 3.8 \end{bmatrix} = \begin{bmatrix} 1 \\ -1.8 \end{bmatrix}$$

$$var(x) = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{\;}$$

$$var(x) = \frac{\sum_{i=1} (x_i - \mu)^2}{n}$$

$$S_1 = \left[\begin{pmatrix}4\\2\end{pmatrix} - \begin{pmatrix}3\\3.8\end{pmatrix}\right]^2 + \left[\begin{pmatrix}2\\4\end{pmatrix} - \begin{pmatrix}3\\3.8\end{pmatrix}\right]^2 + \cdots + \left[\begin{pmatrix}4\\4\end{pmatrix} - \begin{pmatrix}3\\3.8\end{pmatrix}\right]^2$$

$(x_1, x_2)^1$ $(x_1, x_2)^2$

$$S_1 = \underbrace{\underset{2\times1}{\begin{bmatrix}1\\-1.8\end{bmatrix}} \underset{1\times2}{\begin{bmatrix}1 & -1.8\end{bmatrix}}}_{\left(\quad\right)_{2\times2}} +$$

$$S_1 = \begin{bmatrix}1 & -0.25\\-0.25 & 2.2\end{bmatrix}_{2\times2}$$ Symmetric Matrix

Ily covariance of 2nd class : ω_2

$$S_2 = \begin{bmatrix}2.3 & -0.05\\-0.05 & 3.3\end{bmatrix}$$ symmetric matrix

Within-class scatter Matrix

$$\boxed{S_W = (S_1 + S_2)} = \begin{bmatrix}1 & -0.25\\-0.25 & 2.2\end{bmatrix} + \begin{bmatrix}2.3 & -0.05\\-0.05 & 3.3\end{bmatrix}$$

$$S_W = \begin{bmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{bmatrix}$$

Between class Scatter Matrix

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

$$= \left[\begin{pmatrix} 3 \\ 3.8 \end{pmatrix} - \begin{pmatrix} 8.4 \\ 7.6 \end{pmatrix}\right] * \left[\begin{pmatrix} 3 \\ 3.8 \end{pmatrix} - \begin{pmatrix} 8.4 \\ 7.6 \end{pmatrix}\right]^T$$

$$= \underset{2\times1}{\begin{bmatrix} -5.4 \\ -3.8 \end{bmatrix}} \underset{1\times2}{\begin{bmatrix} -5.4 & -3.8 \end{bmatrix}}$$

$$S_B = \underset{2\times2}{\begin{bmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{bmatrix}}$$

LDA projection is obtained as the solution of the generalized eigen value problem:

$$AX = \lambda X$$

$$A\vartheta = \lambda\vartheta$$

$$\boxed{\underbrace{S_W^{-1} \cdot S_B}_{A} \underbrace{W}_{\vartheta} = \underbrace{\lambda}_{\lambda} \underbrace{W}_{\vartheta}}$$

eigen vector

$$|A-\lambda I| = 0$$

$$S_W = \begin{bmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{bmatrix}$$: within class scatter Matrix

$$S_W^{-1} = \frac{\begin{bmatrix} 5.5 & 0.3 \\ 0.3 & 3.3 \end{bmatrix}}{(3.3 \times 5.5) - (-0.3 \times -0.3)}$$

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}_{2\times 2}$$

$$\boxed{A^{-1} = \frac{Adj(A)}{|A|}}$$

$$A^{-1} = \frac{\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}}{|A|}$$

3.3*5.5=18.15
18.15-0.09=18.06

$$\boxed{S_W^{-1} = \frac{\begin{bmatrix} 5.5 & 0.3 \\ 0.3 & 3.3 \end{bmatrix}}{18.06}}$$

$$S_B = \begin{bmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{bmatrix}$$

$$\begin{bmatrix} 9.2 & 6.4 \\ 4.2 & 2.9 \end{bmatrix}$$

$$S_W^{-1} * S_B = \left| \boxed{[\quad] * [\quad]} - \lambda I \right| = 0$$

$$\lambda_1 = 0 \qquad \lambda_2 = 12.2007$$

Trace On/Off a= 9.2 b= 6.4

Show Eigenvalues c= 4.2 d= 2.9

Eigenvalues: {-0.02, 12.12}

0

$$\left(\begin{bmatrix}9.2 & 6.4\\ 4.2 & 2.9\end{bmatrix} - 12.12\right)\begin{bmatrix}w_1\\ w_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$$

$A_{2\times 2}$ $\quad \lambda \quad w^{2\times 1}$

$\lambda_1 = 12.12$

$$w = \begin{array}{c} \begin{bmatrix}-0.57\\ 0.817\end{bmatrix} \end{array}\begin{array}{l} w_1 \\ w_2 \end{array}$$

$\lambda_2 = \underline{-0.2}$

$$w = \begin{bmatrix}0.91\\ 0.42\end{bmatrix}$$