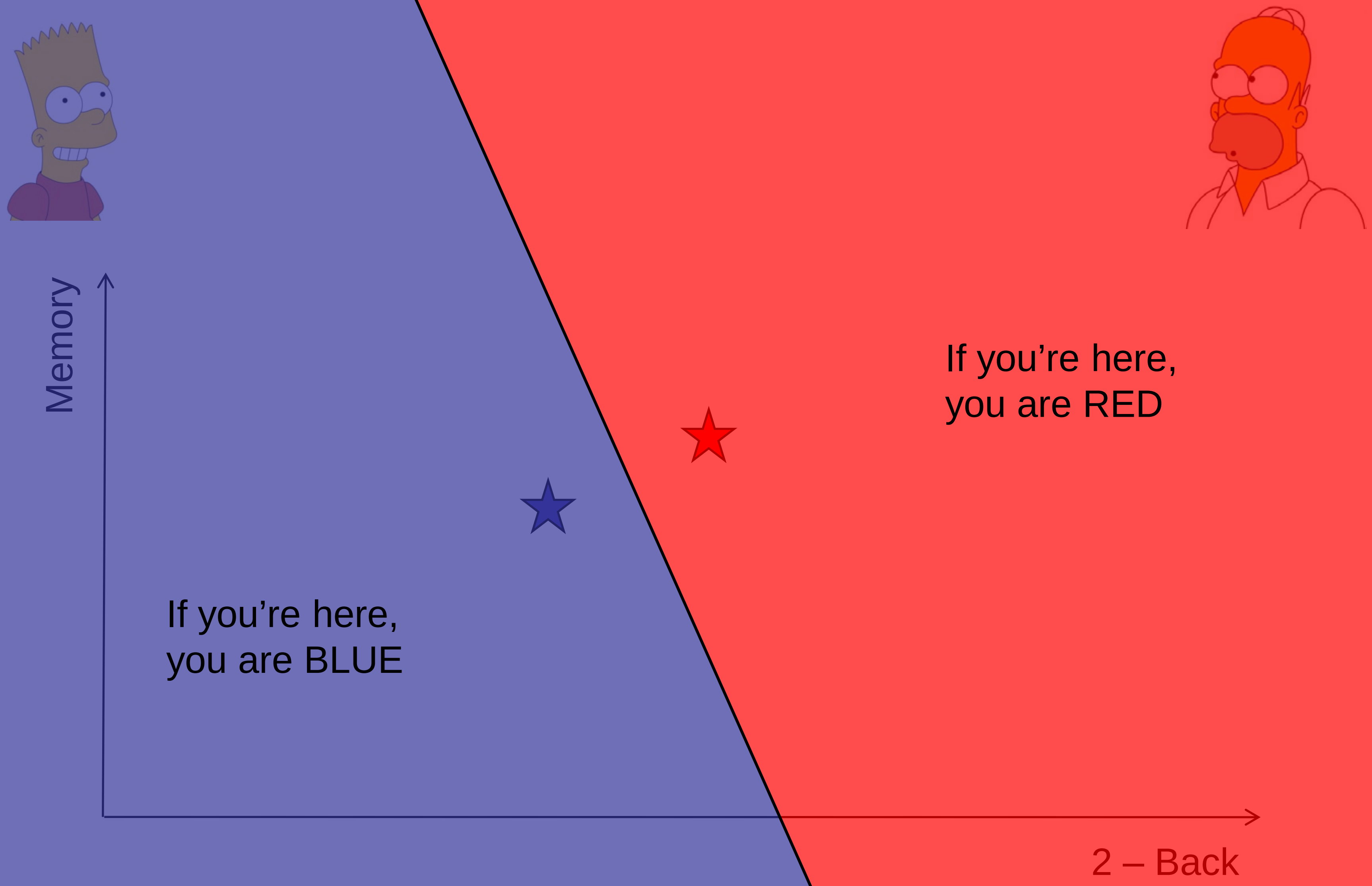
Data Classification

Linear Classifier II

Latent Differential Analysis

Mean Classification



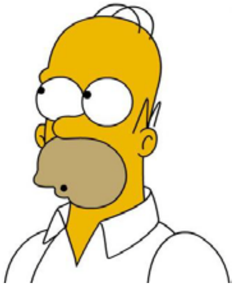
Linear Classifier

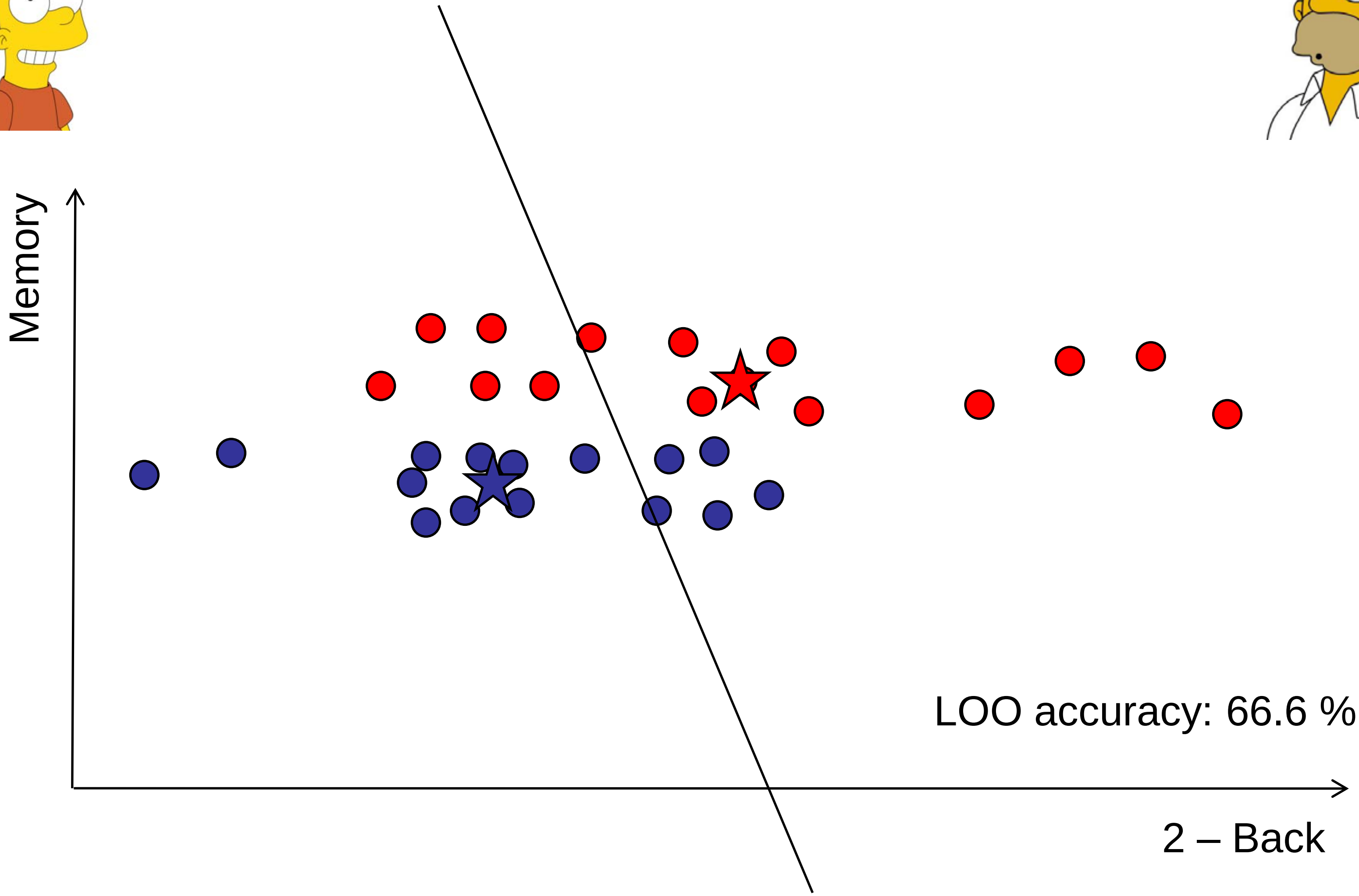
A classifier that assigns a class to a new point based on a separation hyperplane is called a *linear classifier*.

The criterion for a linear classifier can be written as vector product, ie., there is a vector *w* and a number *c* such that a new data vector *x* is classified as being in group one exactly if

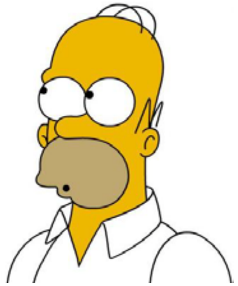
$$w^T x + c > 0$$

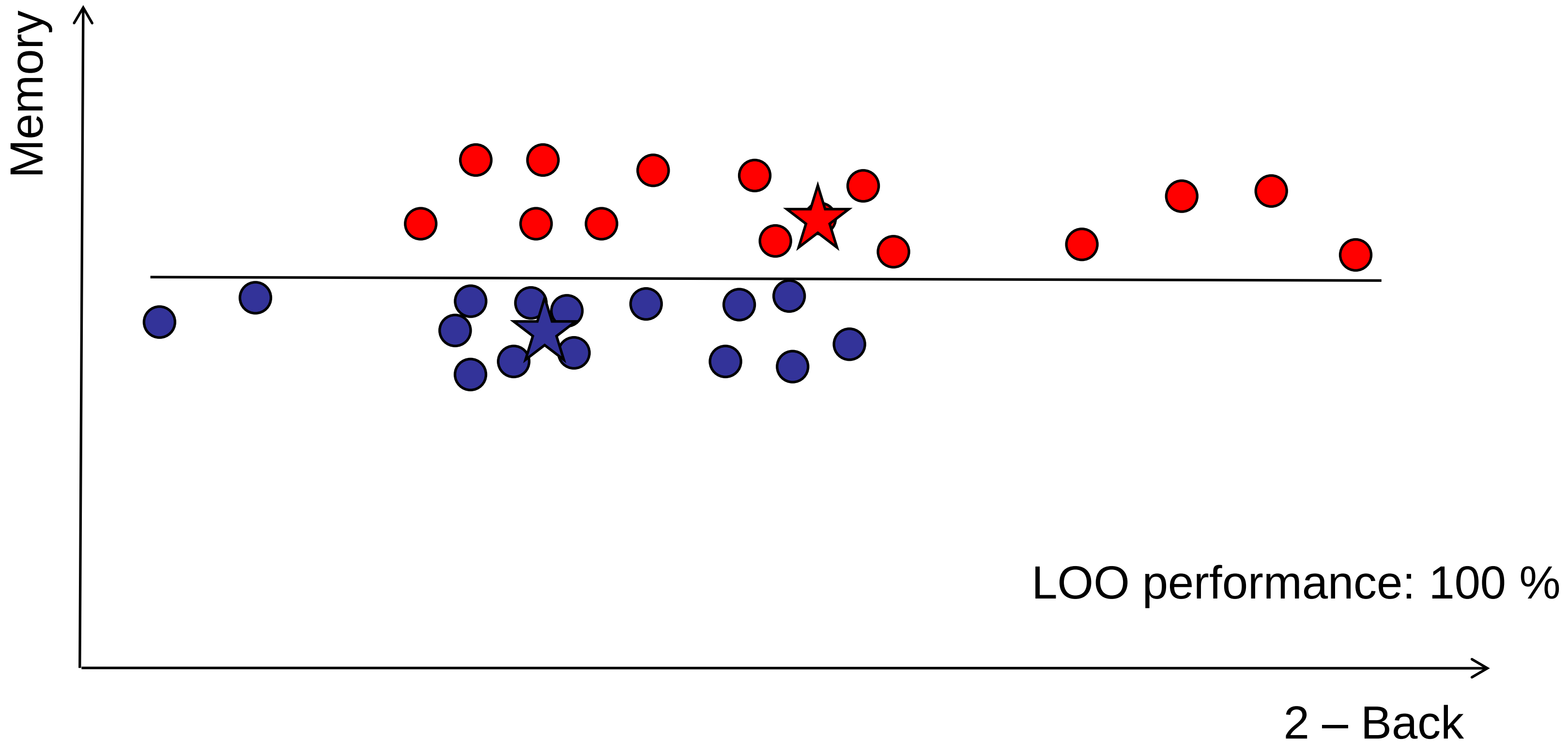
Limitations of Mean Classifier





Linear Classifier Works



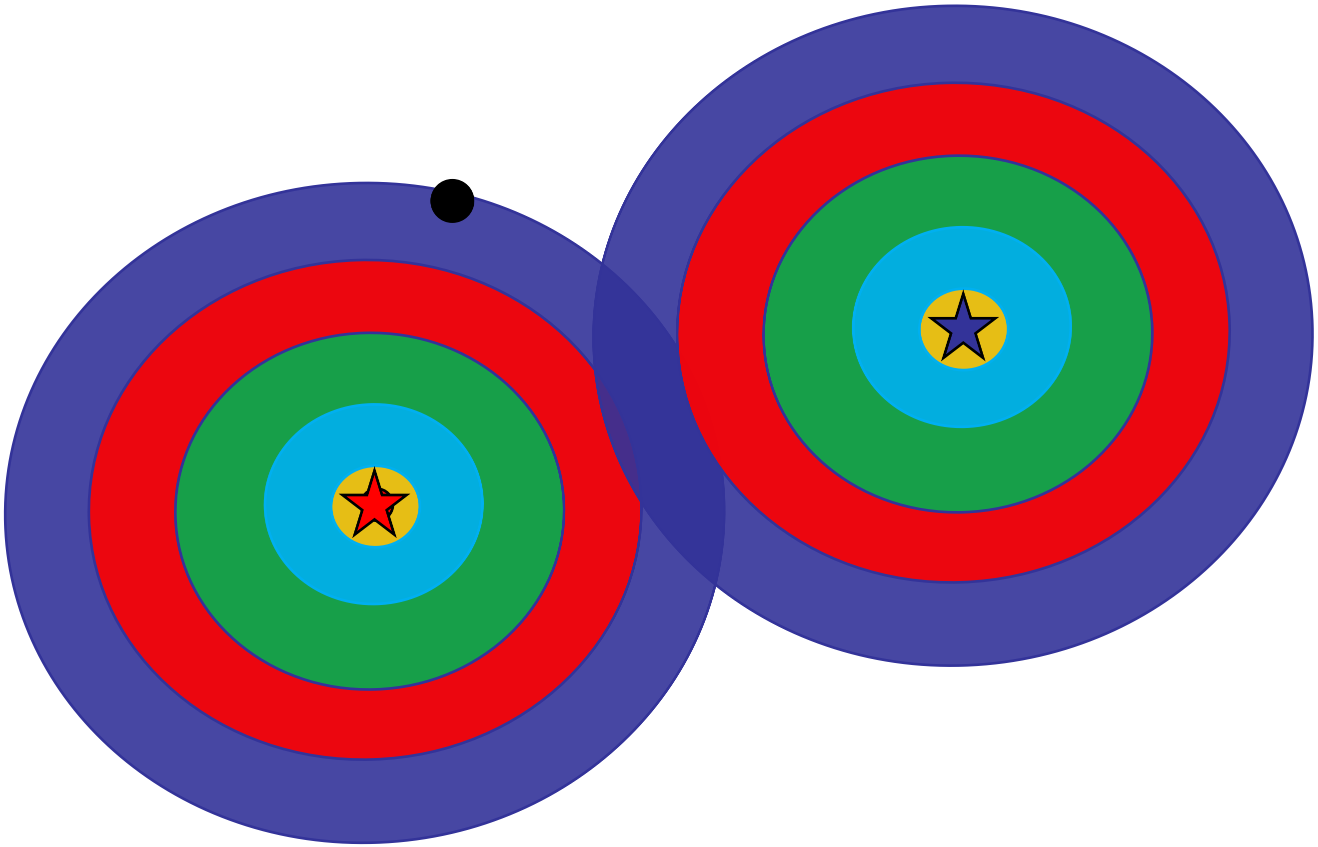


Linear Discriminant Analysis

Observation 1: Mean Classification is equivalent to classifying according to a Gaussian likelihood with identity as covariance matrix.

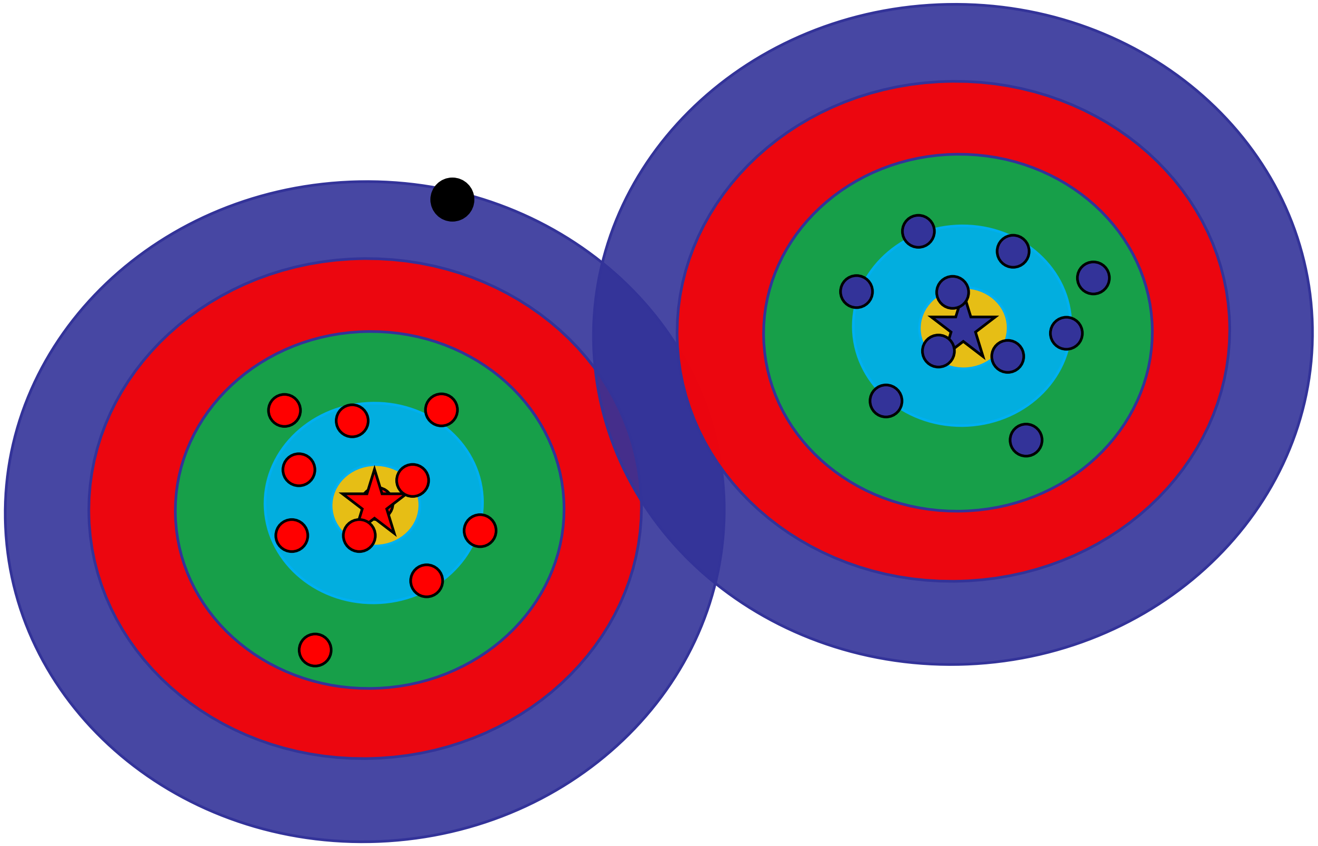
Linear Discriminant Analysis

Observation 1: Mean Classification is equivalent to classifying according to a Gaussian likelihood with identity as covariance matrix.



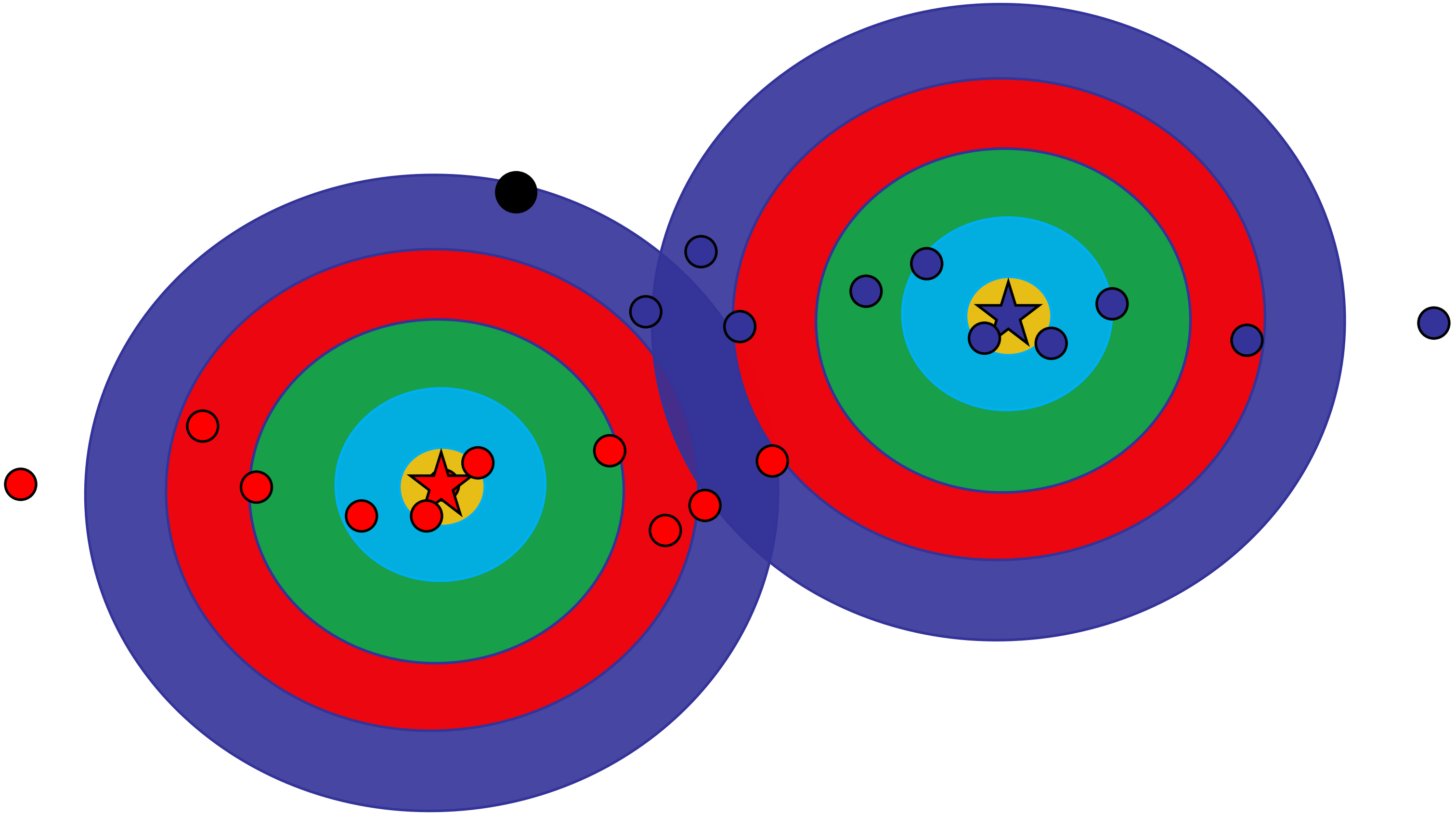
Linear Discriminant Analysis

Observation 1: Mean Classification is equivalent to classifying according to a Gaussian likelihood with identity as covariance matrix.



Linear Discriminant Analysis

Observation 1: Mean Classification is equivalent to classifying according to a Gaussian likelihood with identity as covariance matrix.



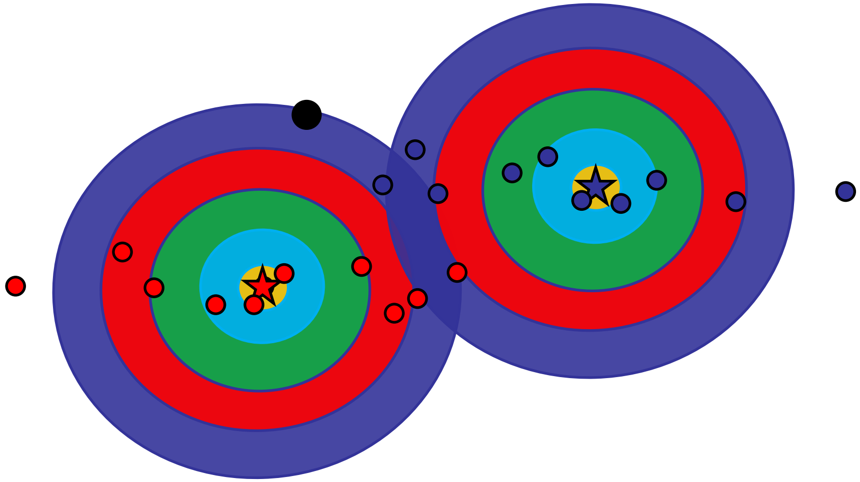
Linear Discriminant Analysis

Observation 1: Mean Classification is equivalent to classifying according to a Gaussian likelihood with identity as covariance matrix.

Q1 Why doesn't the mean classifier work here?

1 The points are not linearly separable.

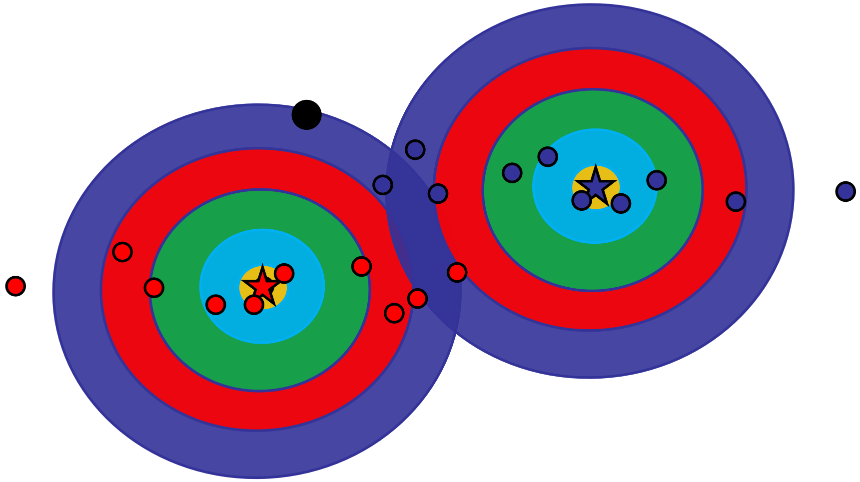
2 The covariance matrix is far from the identity matrix.



Linear Discriminant Analysis

Observation 1: Mean Classification is equivalent to classifying according to a Gaussian likelihood with identity as covariance matrix.

Observation 2: Mean Classification works great if the variables really are distributed with unit covariance matrix, but badly otherwise.

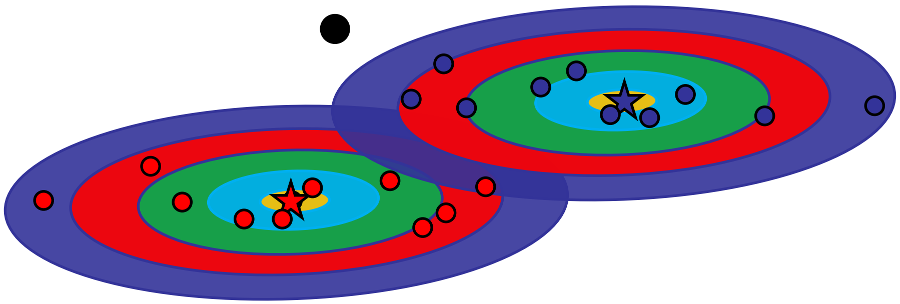


Linear Discriminant Analysis

Observation 1: Mean Classification is equivalent to classifying according to a Gaussian likelihood with identity as covariance matrix.

Observation 2: Mean Classification works great if the variables really are distributed with unit covariance matrix, but badly otherwise.

Linear Discriminant Analysis (LDA): Implement Observation 1, but using real data covariance matrix!



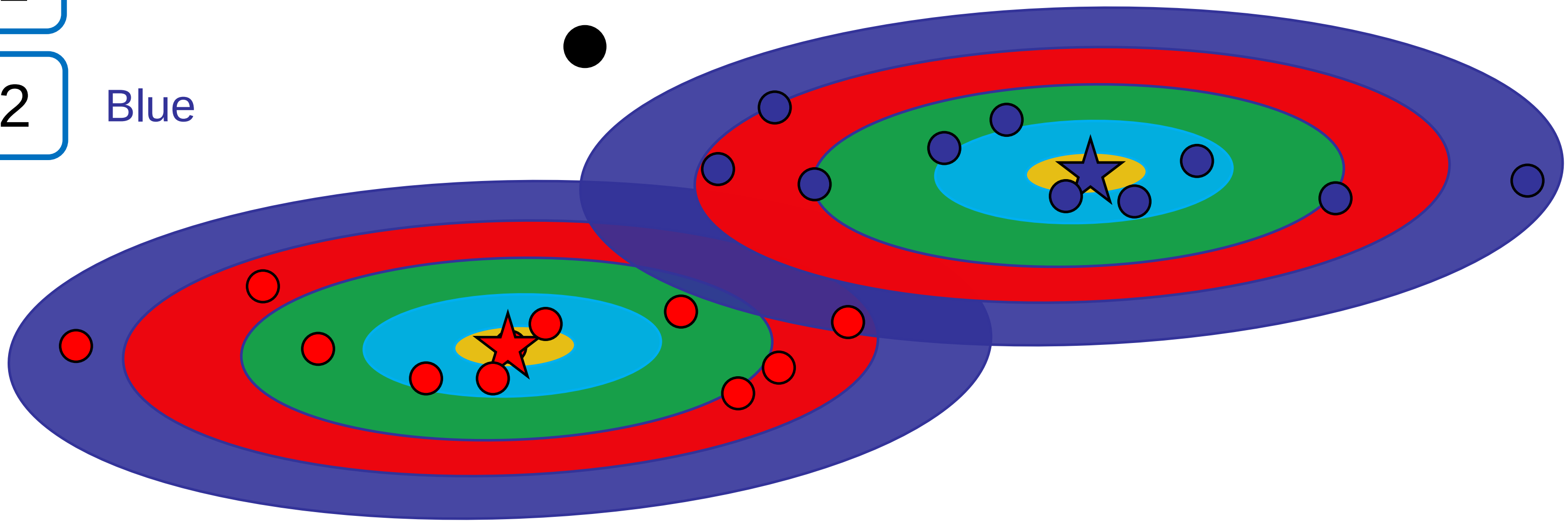
Linear Discriminant Analysis

Linear Discriminant Analysis (LDA): Classify according to a Gaussian likelihood with covariance matrix of the data.

Q2 Which color does LDA classify this point to?

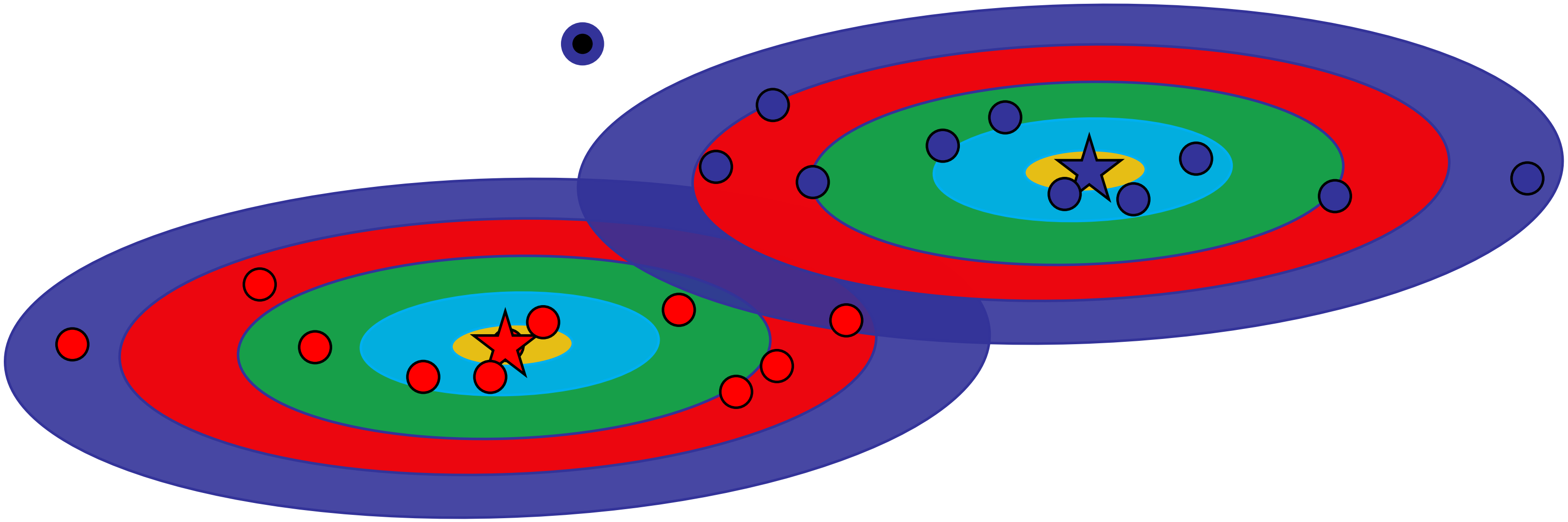
1 RED

2 Blue

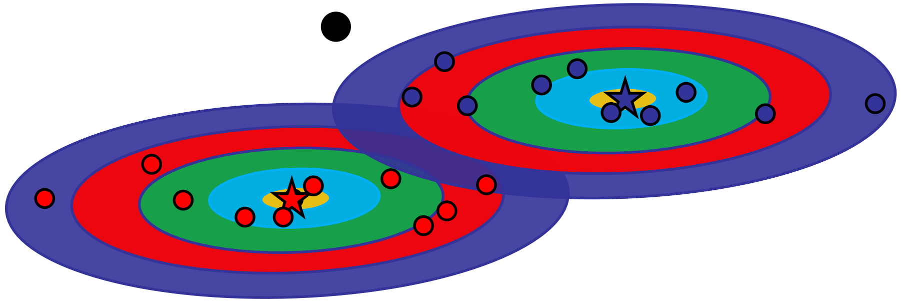


Linear Discriminant Analysis

Linear Discriminant Analysis (LDA): Classify according to a Gaussian likelihood with covariance matrix of the data.



Linear Discriminant Analysis



Linear Discriminant Analysis: Classify according to Gaussian,

That is: Classify *x* as blue if

$$\frac{1}{\sqrt{2\pi|\Sigma|}} e^{-\frac{1}{2}(x-\mu_{Blue})^T \Sigma^{-1} (x-\mu_{Blue})}$$

$$> \quad \frac{1}{\sqrt{2\pi|\Sigma|}} e^{-\frac{1}{2}(x-\mu_{Red})^T \Sigma^{-1} (x-\mu_{Red})}$$

Linear Discriminant Analysis

$$\frac{1}{\sqrt{2\pi|\Sigma|}} e^{-\frac{1}{2}(x-\mu_{Blue})^T \Sigma^{-1}(x-\mu_{Blue})} > \frac{1}{\sqrt{2\pi|\Sigma|}} e^{-\frac{1}{2}(x-\mu_{Red})^T \Sigma^{-1}(x-\mu_{Red})}$$

Q3 $$e^{-\frac{1}{2}(x-\mu_{Blue})^T \Sigma^{-1}(x-\mu_{Blue})} > e^{-\frac{1}{2}(x-\mu_{Red})^T \Sigma^{-1}(x-\mu_{Red})}$$

Q4 $$-\frac{1}{2}(x-\mu_{Blue})^T \Sigma^{-1}(x-\mu_{Blue}) > -\frac{1}{2}(x-\mu_{Red})^T \Sigma^{-1}(x-\mu_{Red})$$

Q5 $$(x-\mu_{Blue})^T \Sigma^{-1}(x-\mu_{Blue}) < (x-\mu_{Red})^T \Sigma^{-1}(x-\mu_{Red})$$

Linear Discriminant Analysis

$$(x - \mu_{Blue})^T \Sigma^{-1} (x - \mu_{Blue}) < (x - \mu_{Red})^T \Sigma^{-1} (x - \mu_{Red})$$

$$x^T \Sigma^{-1} x - 2x^T \Sigma^{-1} \mu_{Blue} + \mu_{Blue}^T \Sigma^{-1} \mu_{Blue}$$

$$< x^T \Sigma^{-1} x - 2x^T \Sigma^{-1} \mu_{Red} + \mu_{Red}^T \Sigma^{-1} \mu_{Red}$$

$$x^T \underbrace{2\Sigma^{-1}(\mu_{Blue} - \mu_{Red})}_{w} + \underbrace{\mu_{Red}^T \Sigma^{-1} \mu_{Red} - \mu_{Blue}^T \Sigma^{-1} \mu_{Blue}}_{c} > 0$$

Linear Discriminant Analysis

Let μ_1 and μ_2 be the two group means in the training set, and Σ the covariance matrix. The linear classifier that classifies each item *x* to the group with higher Gaussian likelihood under these means and the common covariance matrix,

$$x^T 2\Sigma^{-1}(\mu_1 - \mu_2) + \mu_2^T \Sigma^{-1} \mu_2 - \mu_1^T \Sigma^{-1} \mu_1 > 0$$

is called *Linear Discriminant Analysis*.

Note: The common covariance matrix is the average squared distance from the *mean in each group*, not from the total mean!

Example for LDA

Control Group

ID 1 = (1,2,1)
ID 2 = (2,1,0)
ID 3 = (1,1,1)
ID 4 = (0,0,2)

$$\mu_1 = (1, 1, 1)$$

Treatment Group

ID 5 = (3,1,1)
ID 6 = (4,1,1)
ID 7 = (0,2,0)
ID 8 = (1,0,2)

$$\mu_2 = (2, 1, 1)$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Example for LDA

Control Group

ID 1 = (1,2,1)
ID 2 = (2,1,0)
ID 3 = (1,1,1)
ID 4 = (0,0,2)

Treatment Group

ID 5 = (3,1,1)
ID 6 = (4,1,1)
ID 7 = (0,2,0)
ID 8 = (1,0,2)

$$\mu_1 = (1, 1, 1)$$

$$\mu_2 = (2, 1, 1)$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

1

$$\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Q6

2

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

Example for LDA

Control Group

ID 1 = (1,2,1)
ID 2 = (2,1,0)
ID 3 = (1,1,1)
ID 4 = (0,0,2)

$$\mu_1 = (1, 1, 1)$$

Treatment Group

ID 5 = (3,1,1)
ID 6 = (4,1,1)
ID 7 = (0,2,0)
ID 8 = (1,0,2)

$$\mu_2 = (2, 1, 1)$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

Example for LDA

Control Group

ID 1 = (1,2,1)
ID 2 = (2,1,0)
ID 3 = (1,1,1)
ID 4 = (0,0,2)

$$\mu_1 = (1, 1, 1)$$

$$cov_1 = \frac{1}{4} \begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & -1 \\ -2 & -1 & 2 \end{pmatrix}$$

Treatment Group

ID 5 = (3,1,1)
ID 6 = (4,1,1)
ID 7 = (0,2,0)
ID 8 = (1,0,2)

$$\mu_2 = (2, 1, 1)$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 4 & -2 & 2 \\ -2 & 1 & -1 \\ 2 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

Example for LDA

Control Group

ID 1 = (1,2,1)
ID 2 = (2,1,0)
ID 3 = (1,1,1)
ID 4 = (0,0,2)

$$\mu_1 = (1, 1, 1)$$

$$cov_1 = \frac{1}{4}\begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & -1 \\ -2 & -1 & 2 \end{pmatrix}$$

Treatment Group

ID 5 = (3,1,1)
ID 6 = (4,1,1)
ID 7 = (0,2,0)
ID 8 = (1,0,2)

$$\mu_2 = (2, 1, 1)$$

$$cov_2 = \frac{1}{4}\begin{pmatrix} 10 & -1 & 1 \\ -1 & 2 & -2 \\ 1 & -2 & 2 \end{pmatrix}$$

Example for LDA

Control Group

ID 1 = (1,2,1)
ID 2 = (2,1,0)
ID 3 = (1,1,1)
ID 4 = (0,0,2)

Treatment Group

ID 5 = (3,1,1)
ID 6 = (4,1,1)
ID 7 = (0,2,0)
ID 8 = (1,0,2)

$$\mu_1 = (1, 1, 1)$$

$$\mu_2 = (2, 1, 1)$$

$$cov_1 = \frac{1}{4}\begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & -1 \\ -2 & -1 & 2 \end{pmatrix}$$

$$cov_2 = \frac{1}{4}\begin{pmatrix} 10 & -1 & 1 \\ -1 & 2 & -2 \\ 1 & -2 & 2 \end{pmatrix}$$

$$cov_{corrected} = \frac{1}{8}\begin{pmatrix} 12 & 0 & -1 \\ 0 & 4 & -3 \\ -1 & -3 & 4 \end{pmatrix}$$

Example for LDA

Control Group

$$\mu_1 = (1, 1, 1)$$

Treatment Group

$$\mu_2 = (2, 1, 1)$$

$$cov_{corrected} = \frac{1}{8} \begin{pmatrix} 12 & 0 & -1 \\ 0 & 4 & -3 \\ -1 & -3 & 4 \end{pmatrix}$$

$$w^T x + c > 0$$

$$w = 2\Sigma^{-1}(\mu_1 - \mu_2) = 2 \begin{pmatrix} 0.7 & 0.3 & -0.4 \\ 0.3 & 4.7 & -3.6 \\ -0.4 & -3.6 & 4.8 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1.4 \\ 0.6 \\ -0.8 \end{pmatrix}$$

$$c = \mu_2^T \Sigma^{-1} \mu_2 - \mu_1^T \Sigma^{-1} \mu_1 = -1.9$$

Example for LDA

Control Group

ID 1 = (1,2,1)
ID 2 = (2,1,0)
ID 3 = (1,1,1)
ID 4 = (0,0,2)

Treatment Group

ID 5 = (3,1,1)
ID 6 = (4,1,1)
ID 7 = (0,2,0)
ID 8 = (1,0,2)

$$(1.4, 0.6, -0.8)x - 1.9 > 0$$

Filter

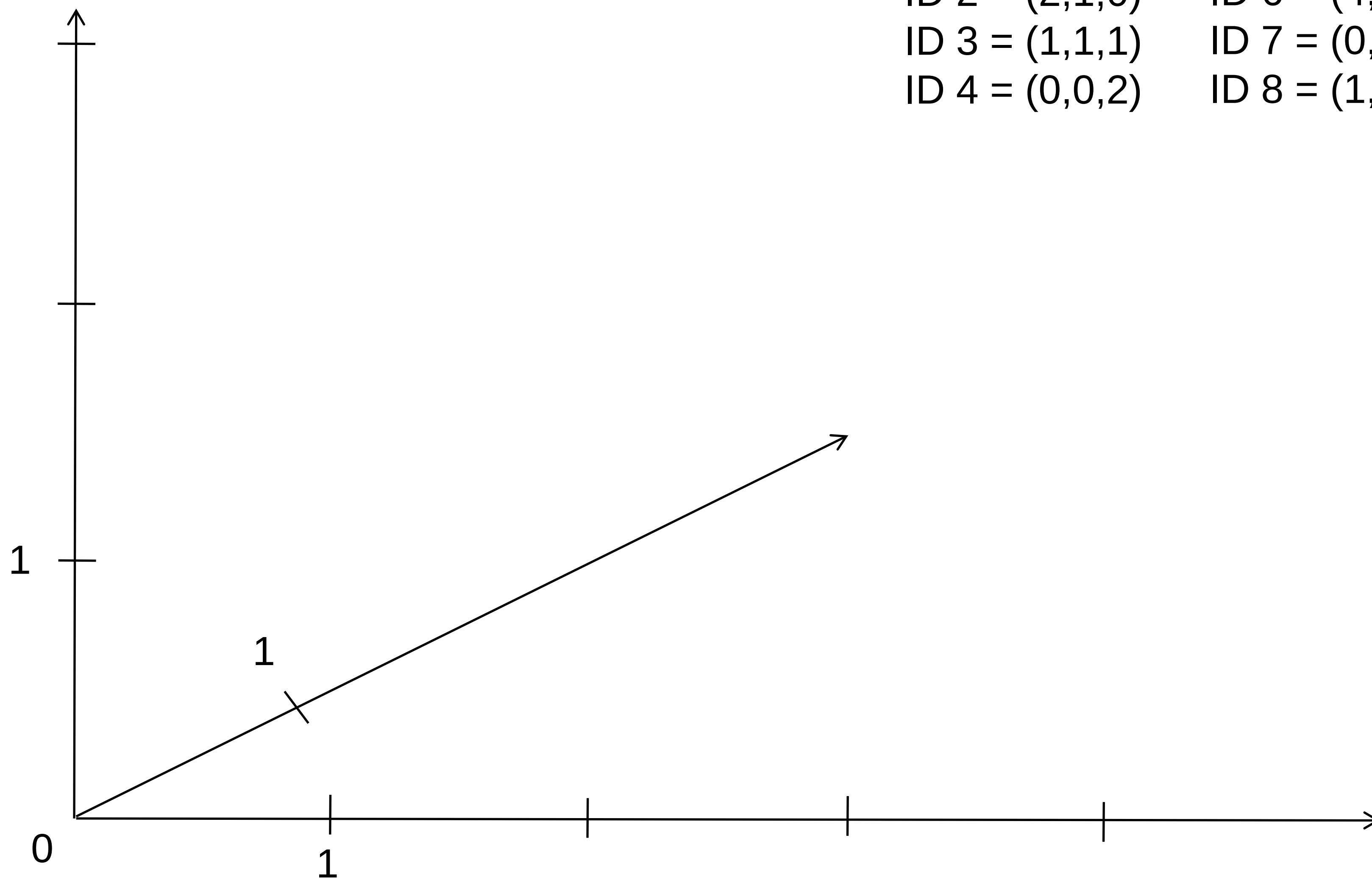
Example for LDA

Control Group		Treatment Group	
ID 1 = (1,2,1)	-0.1	ID 5 = (3,1,1)	2.1
ID 2 = (2,1,0)	1.5	ID 6 = (4,1,1)	5.4
ID 3 = (1,1,1)	-0.7	ID 7 = (0,2,0)	-0.7
ID 4 = (0,0,2)	-3.5	ID 8 = (1,0,2)	-2.1

$$(1.4, 0.6, -0.8)x - 1.9 > 0$$

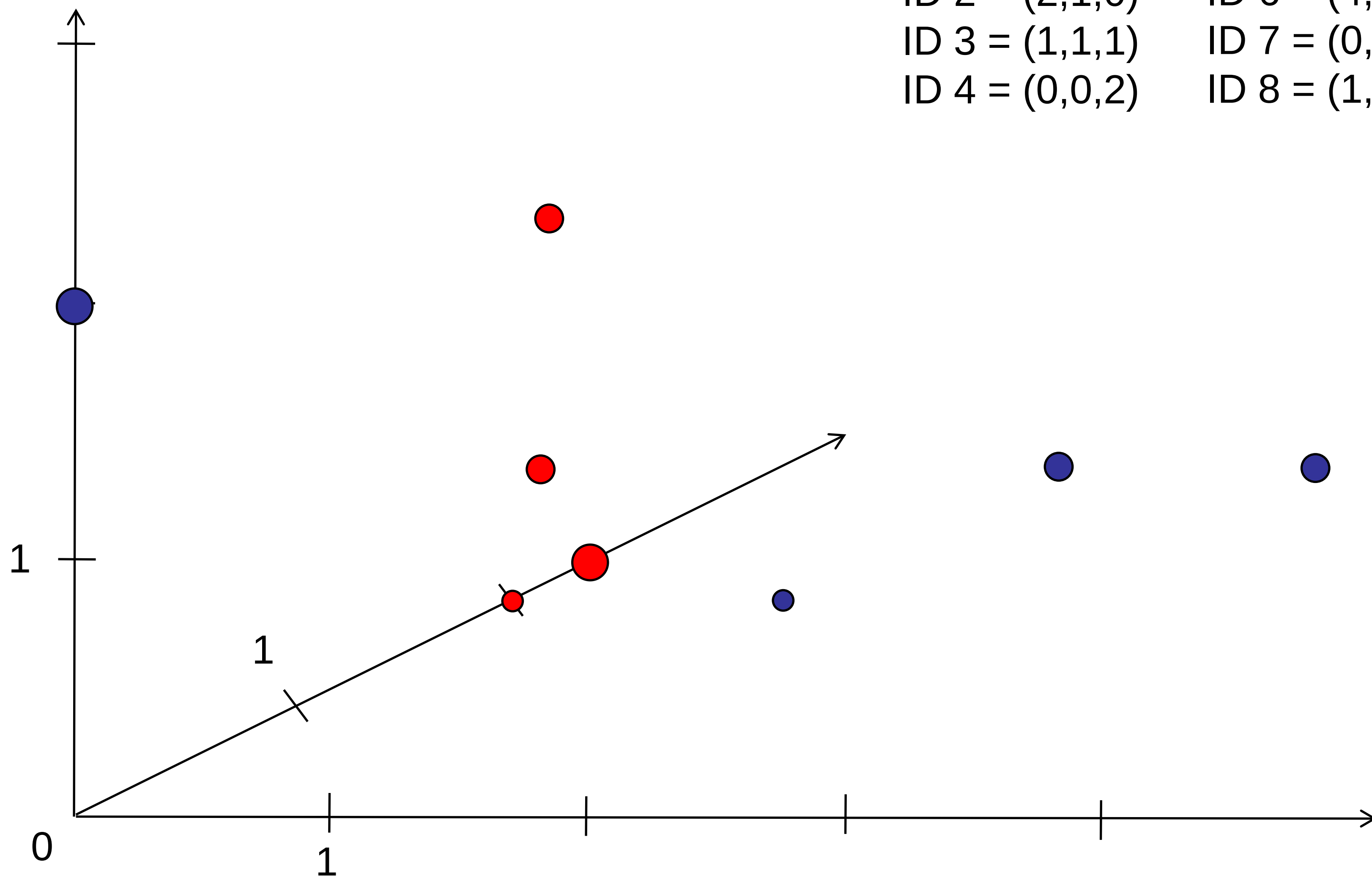
Example for LDA

ID 1 = (1,2,1)
ID 2 = (2,1,0)
ID 3 = (1,1,1)
ID 4 = (0,0,2)
ID 5 = (3,1,1)
ID 6 = (4,1,1)
ID 7 = (0,2,0)
ID 8 = (1,0,2)



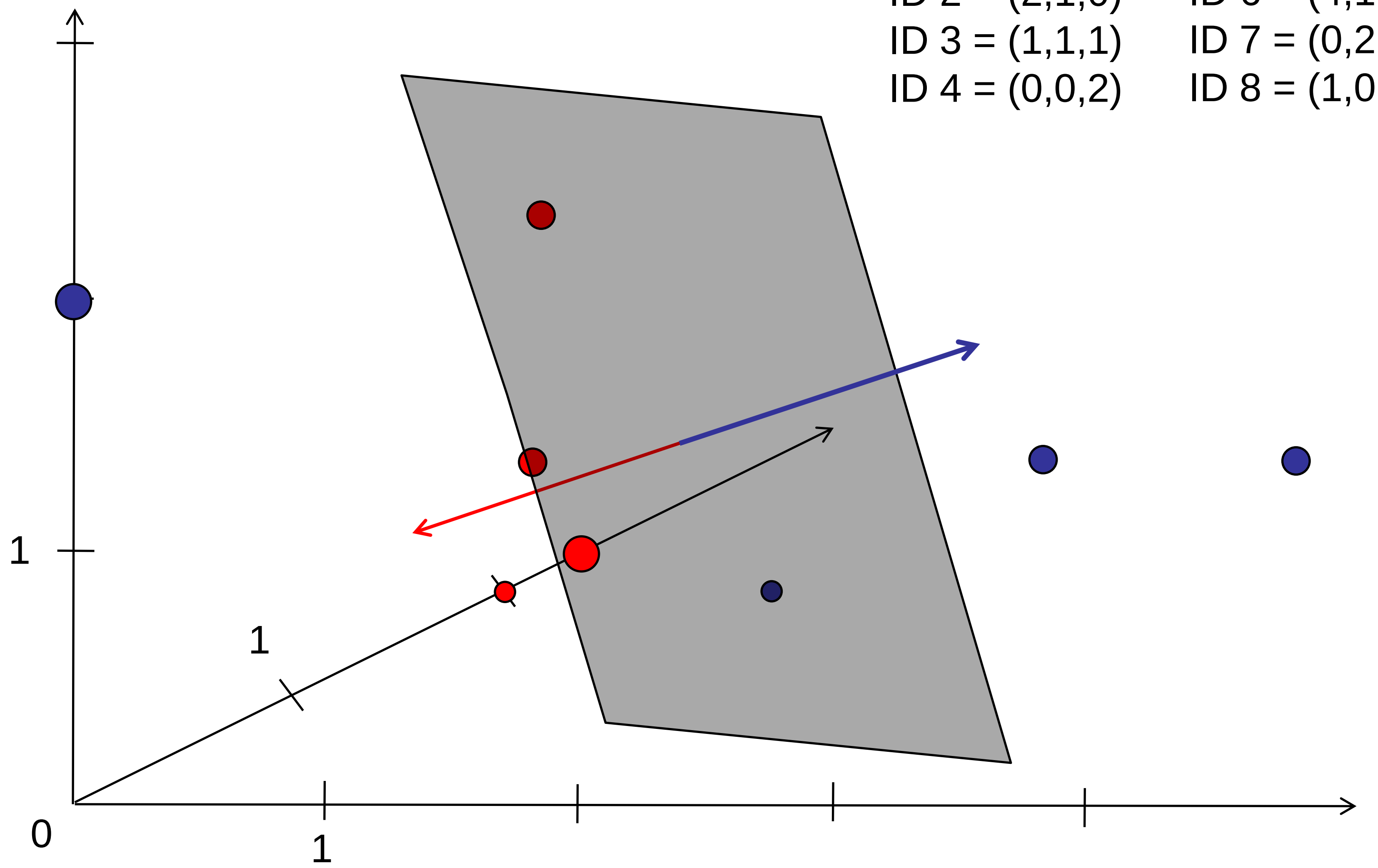
Example for LDA

ID 1 = (1,2,1)
ID 2 = (2,1,0)
ID 3 = (1,1,1)
ID 4 = (0,0,2)
ID 5 = (3,1,1)
ID 6 = (4,1,1)
ID 7 = (0,2,0)
ID 8 = (1,0,2)

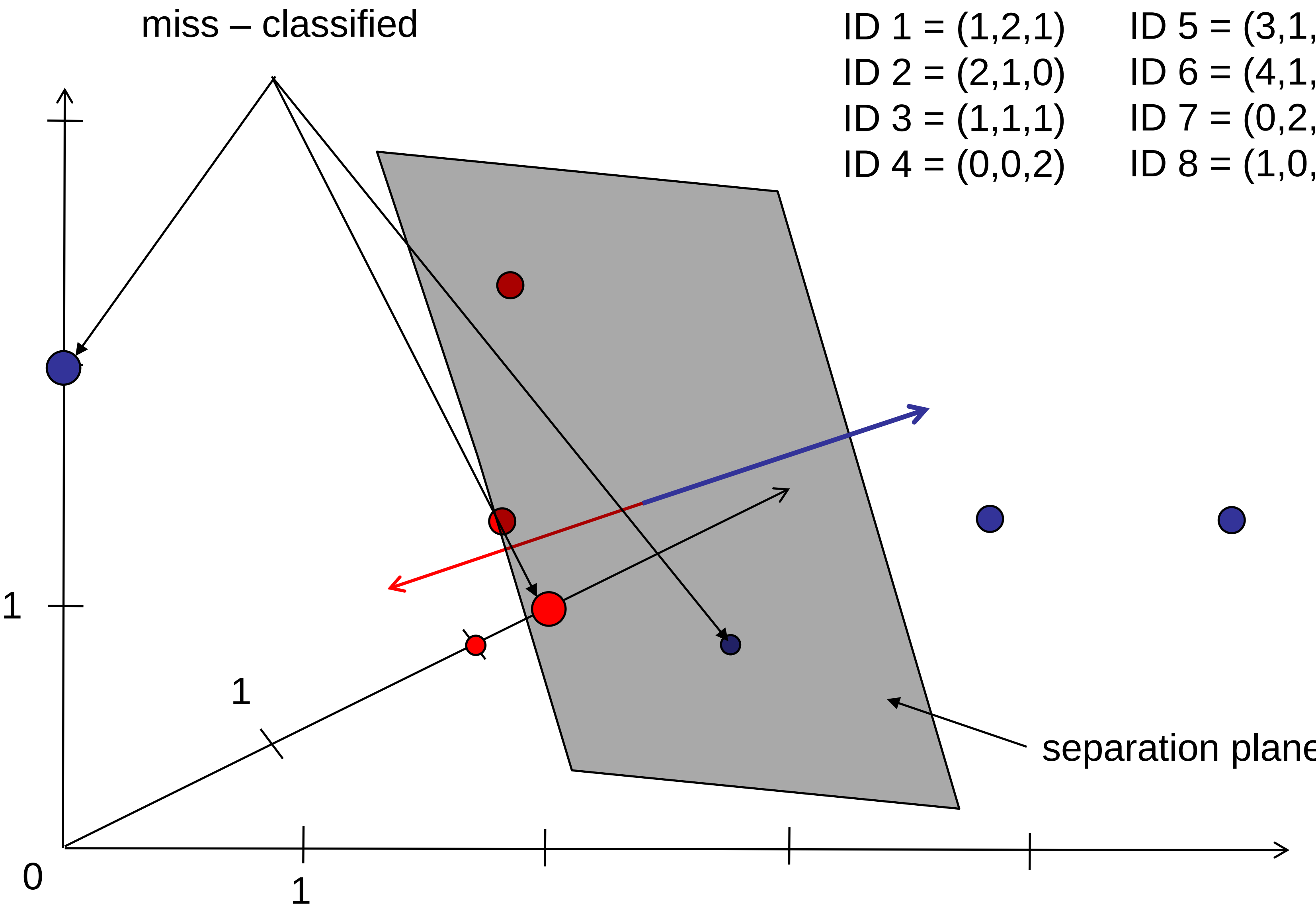


Example for LDA

ID 1 = (1,2,1)	ID 5 = (3,1,1)
ID 2 = (2,1,0)	ID 6 = (4,1,1)
ID 3 = (1,1,1)	ID 7 = (0,2,2)
ID 4 = (0,0,2)	ID 8 = (1,0,0)

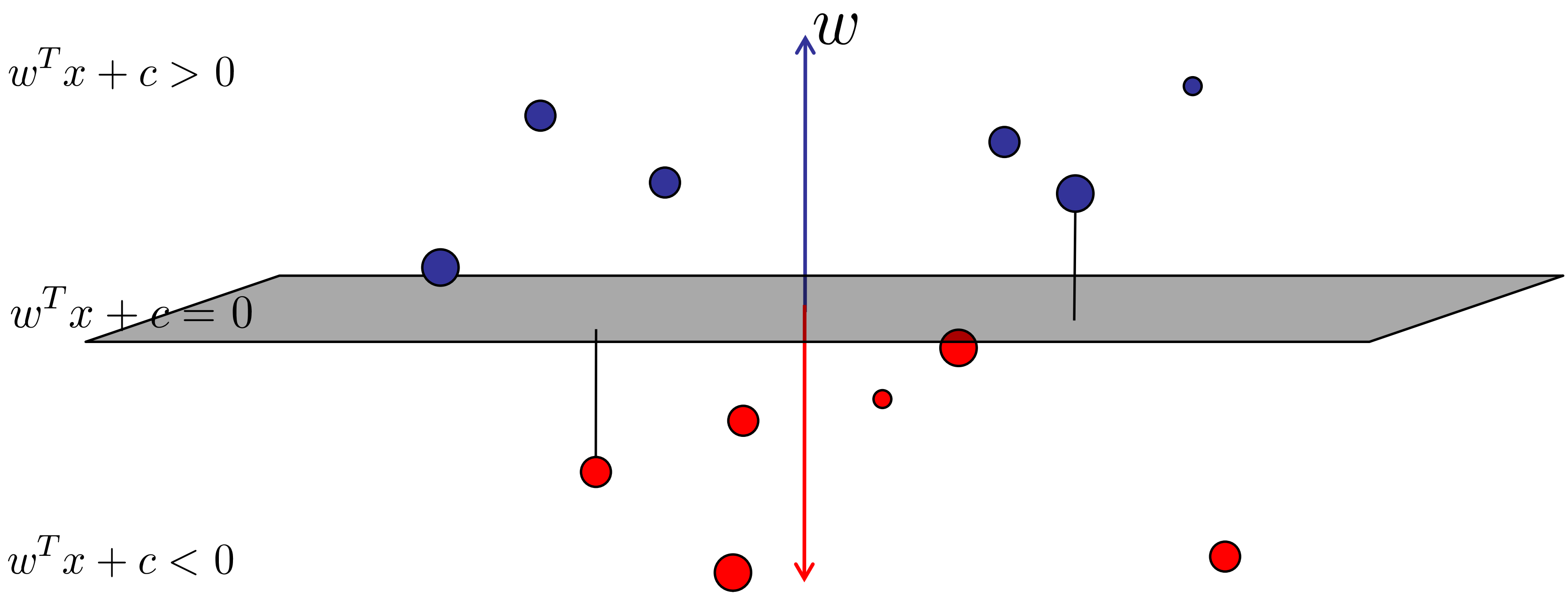


Example for LDA



ID 1 = (1,2,1)
ID 2 = (2,1,0)
ID 3 = (1,1,1)
ID 4 = (0,0,2)
ID 5 = (3,1,1)
ID 6 = (4,1,1)
ID 7 = (0,2,2)
ID 8 = (1,0,0)

Geometry of Linear Classifier



Linear Regression

x *t*

ID 1 = (1,2,1), -1
ID 2 = (2,1,0), -1
ID 3 = (1,1,1), -1
ID 4 = (0,0,2), -1
ID 5 = (3,1,1), 1
ID 6 = (4,1,1), 1
ID 7 = (0,2,2), 1
ID 8 = (1,0,0), 1

$$w^T x_i + c = t_i$$

Make it as close as possible

minimize: $$\sum_{i=1}^{N} (w^T x_i + c - t_i)^2$$

Linear Regression

x *t*

ID 1 = (1,1,2,1), -1
ID 2 = (1,2,1,0), -1
ID 3 = (1,1,1,1), -1
ID 4 = (1,0,0,2), -1
ID 5 = (1,3,1,1), 1
ID 6 = (1,4,1,1), 1
ID 7 = (1,0,2,2), 1
ID 8 = (1,1,0,0), 1

$$(c, w_1, w_2, w_3) \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} = (w_1, w_2, w_3) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + c$$

$$w^T x_i = t_i$$

minimize: $$\sum_{i=1}^{N} (w^T x_i - t_i)^2$$

Linear Regression

minimize: $$\sum_{i=1}^{N}(w^T x_i - t_i)^2$$

Minimization is the same as setting the 1st derivative to zero:

$$\frac{\partial}{\partial w}\sum_{i=1}^{N}(w^T x_i - t_i)^2 \quad = \quad 2\sum_{i=1}^{N}(w^T x_i - t_i)x_i^T$$

Answer Form
Working with Linear Classifier II

Q1	[ ]	2
Q2	1	[ ]

Q3: $e^{-\frac{1}{2}(x-\mu_{Blue})^T\Sigma^{-1}(x-\mu_{Blue})} > e^{-\frac{1}{2}(x-\mu_{Red})^T\Sigma^{-1}(x-\mu_{Red})}$

Q4: $-\frac{1}{2}(x-\mu_{Blue})^T\Sigma^{-1}(x-\mu_{Blue}) > -\frac{1}{2}(x-\mu_{Red})^T\Sigma^{-1}(x-\mu_{Red})$

Q5: $(x-\mu_{Blue})^T\Sigma^{-1}(x-\mu_{Blue}) < (x-\mu_{Red})^T\Sigma^{-1}(x-\mu_{Red})$

Q6	1	[ ]