Discriminant Analysis and Classification

- Background and Motivation
- PCA v.s. LDA
- Mathematical Formulation for Fisher's LDA
- General Framework for LDA (next lecture)

Linear Discriminant Analysis

- This method was formulated by R. A. Fisher in 1936 for two population/classes/groups.
- It is known as *Fisher's linear discriminant*.
  - The basic idea is to maximize the variability between groups and minimize the variability within each group
- Linear discriminant analysis (LDA) or discriminant function analysis is a generalization of Fisher's linear discriminant for multiple groups.
- The terms *Fisher's linear discriminant* and *linear discriminant analysis* are often used interchangeably.

PCA v.s. LDA

- Recall that PCA is a method to find the linear combinations that account for as much variability as possible

$$PC = \alpha_1 X_1 + \alpha_2 X_2, \alpha_1^2 + \alpha_2^2 = 1$$

- LDA is a method that aims to maximize the separation between two or more groups/categories

$$LD = v_1 X_1 + v_2 X_2$$

Motivating Example

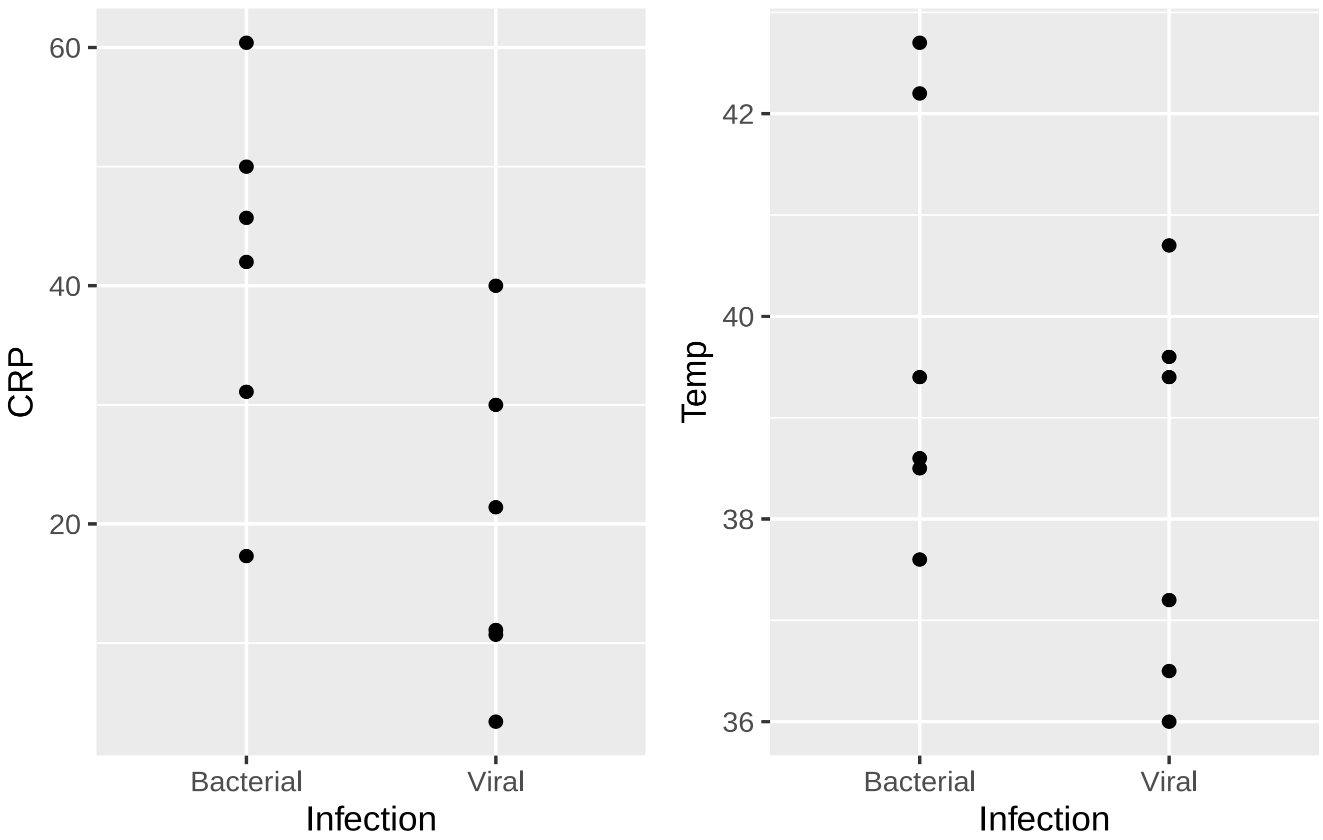
- How can one quickly determine if a patient has a viral infection or a bacterial infection with blood samples?
  - Problem: we have to wait about several days to know if antibiotic treatment is appropriate or not.
  - Maybe we could use information (e.g., CRP and body temperature) in the blood samples to tell viral or bacterial infection because blood samples can be measured within just an hour.

Example Data

- CRP: Concentration of the c-reactive protein in blood from the time when the patients entered the hospital.
- Temp: Body temperature of the same patients at the same time point.
- Can we use CRP or body temperature to tell if a patient has a bacterial or viral infection?

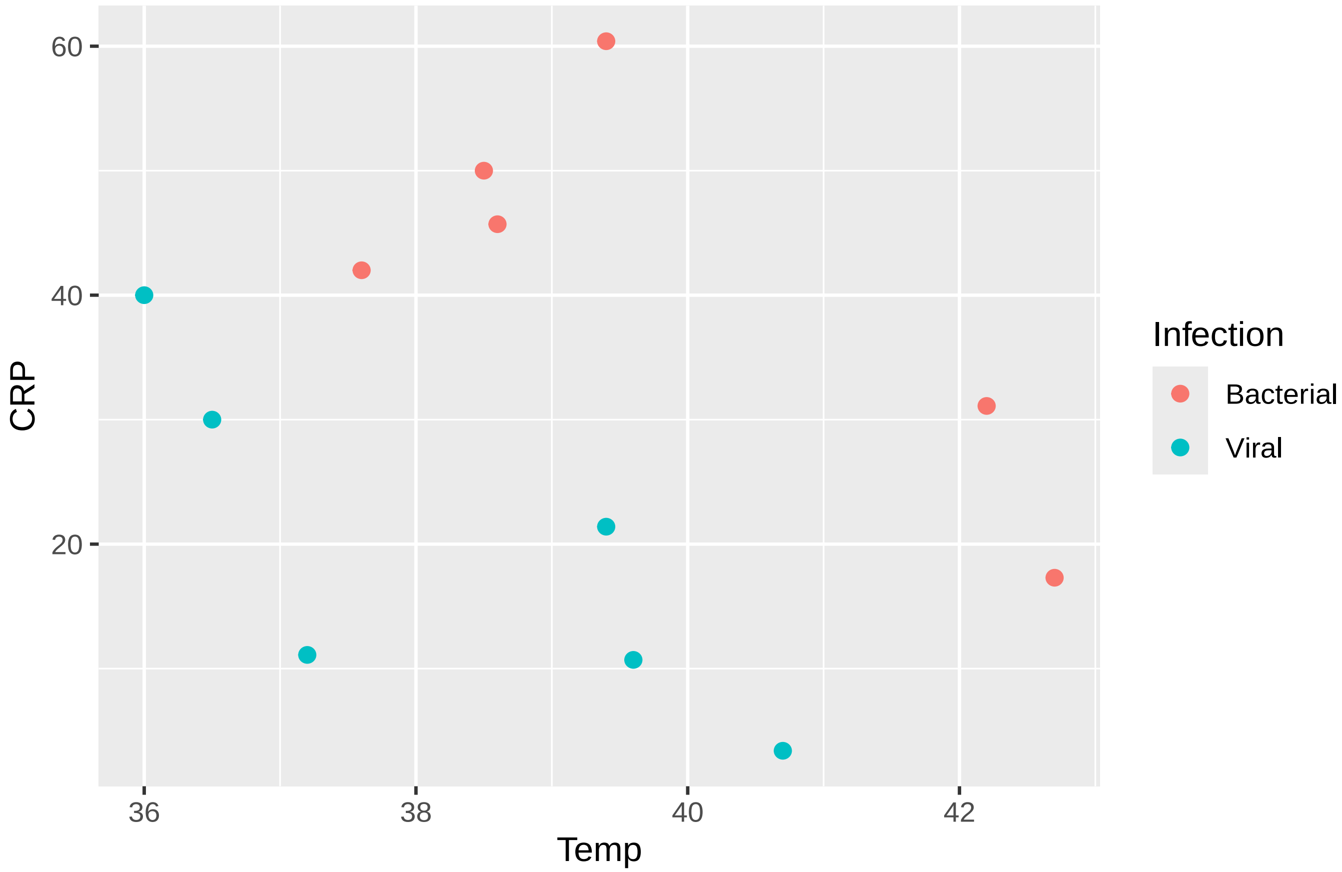
	Infection	CRP	Temp
1	Viral	40.0	36.0
2	Viral	11.1	37.2
3	Viral	30.0	36.5
4	Viral	21.4	39.4
5	Viral	10.7	39.6
6	Viral	3.4	40.7
7	Bacterial	42.0	37.6
8	Bacterial	31.1	42.2
9	Bacterial	50.0	38.5
10	Bacterial	60.4	39.4
11	Bacterial	45.7	38.6
12	Bacterial	17.3	42.7

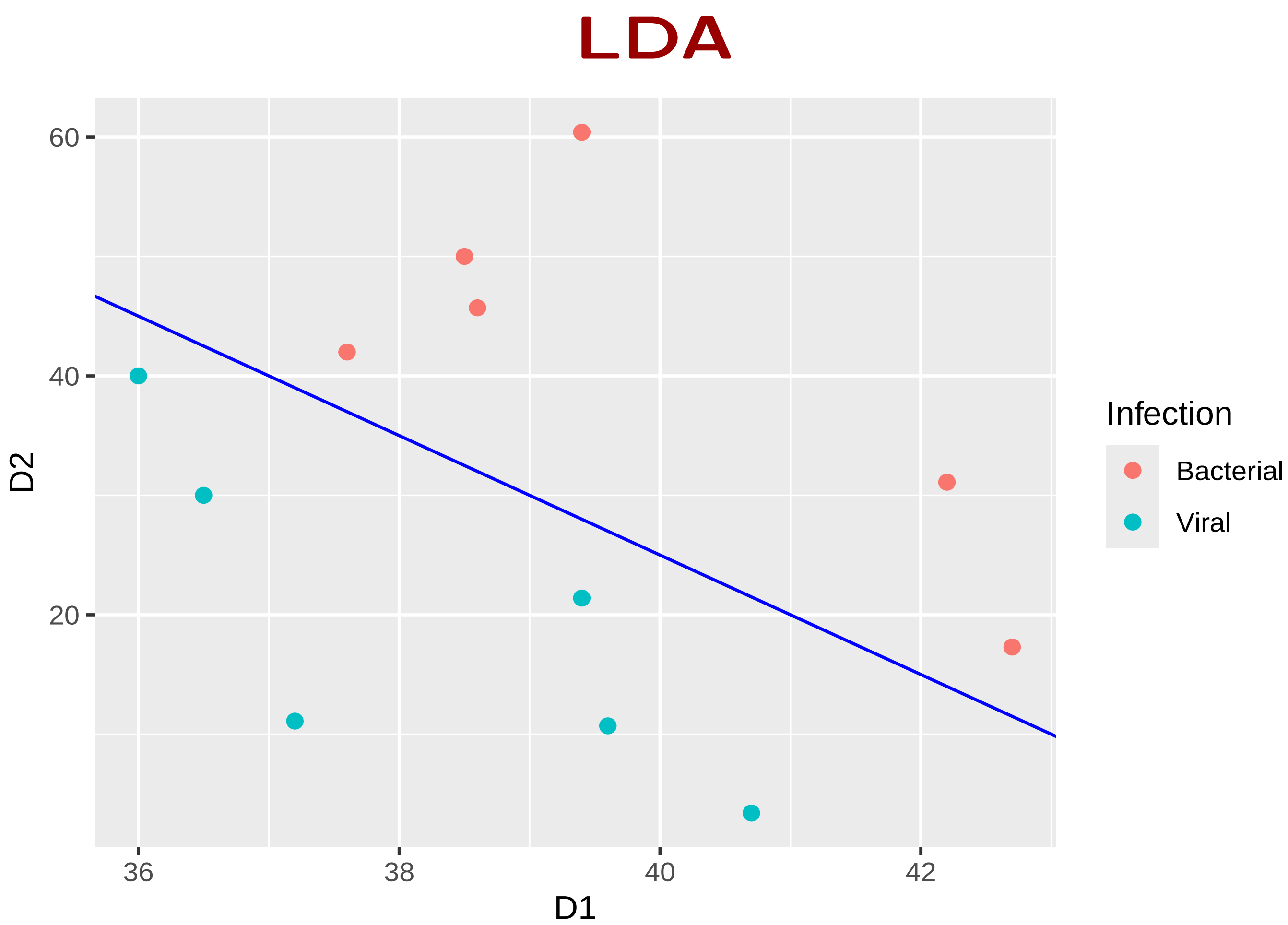
Separation



60
40
20
CRP
Bacterial
Viral
Infection
42
40
38
36
Temp
Bacterial
Viral
Infection

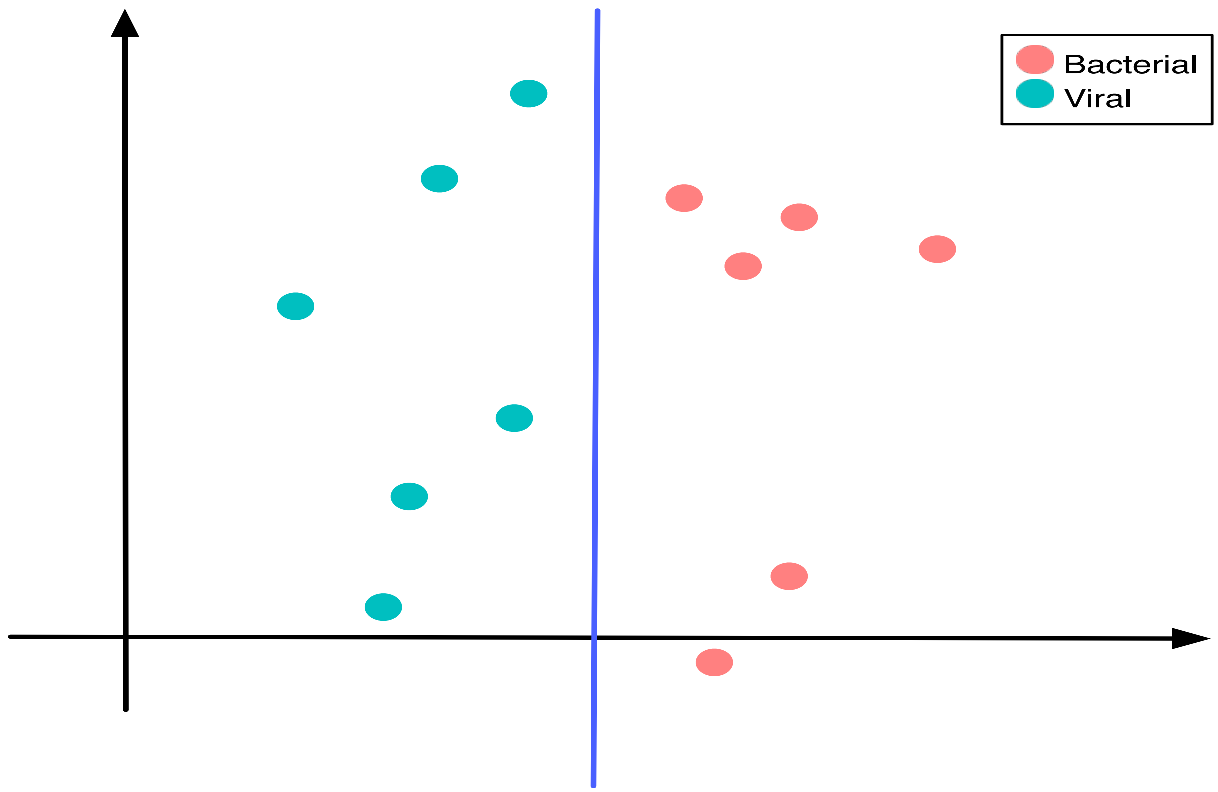
Separation




LDA
60
40
20
D2
36
38
40
42
D1
Infection
Bacterial
Viral

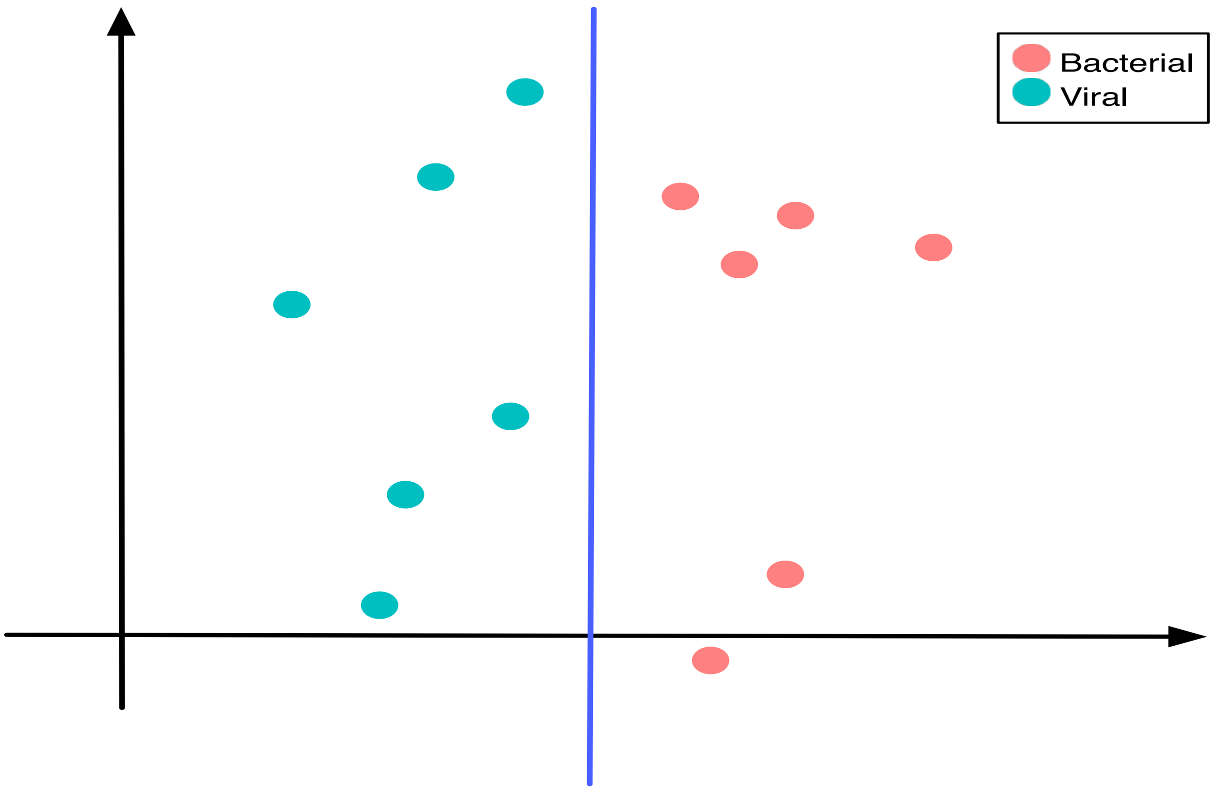
LDA

$$LD = v_1X_1 + v_2X_2$$



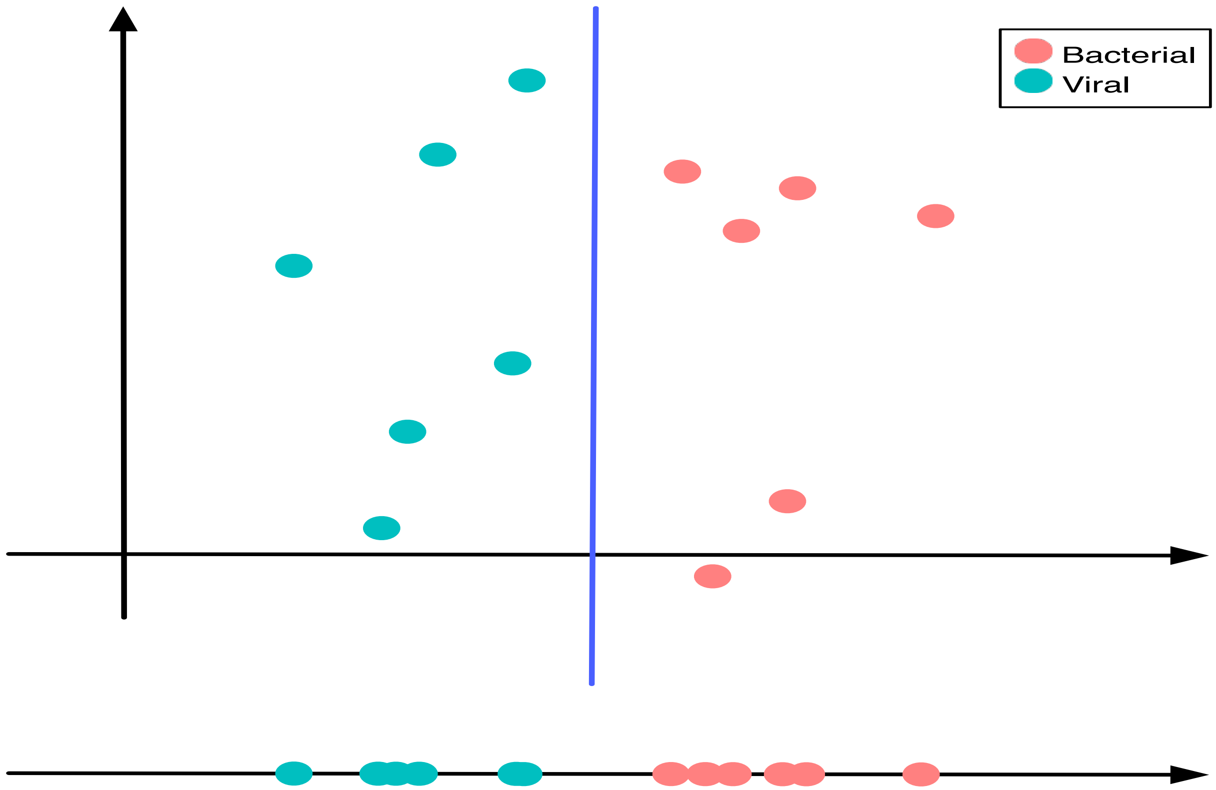
LDA

$$LD = 0.11 \cdot \text{CRP} + 0.70 \cdot \text{Temp}$$



LDA

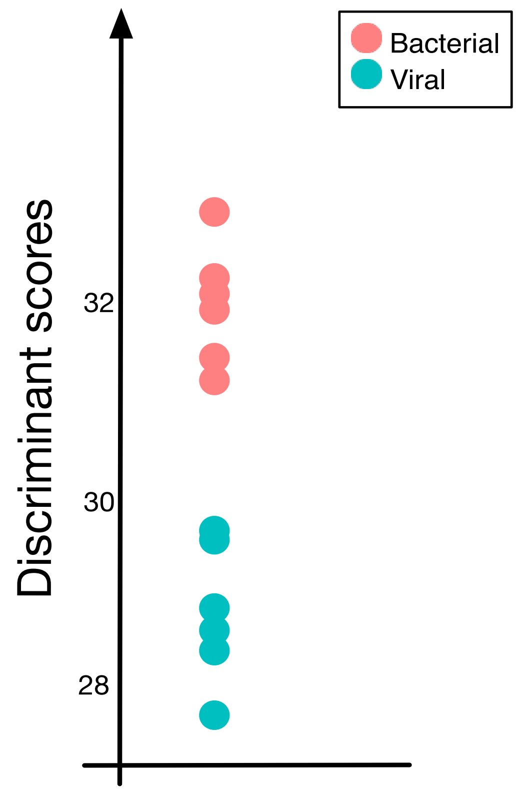
$$LD = 0.11 \cdot \text{CRP} + 0.70 \cdot \text{Temp}$$



LDA

$$LD = 0.11 \cdot \text{CRP} + 0.70 \cdot \text{Temp}$$

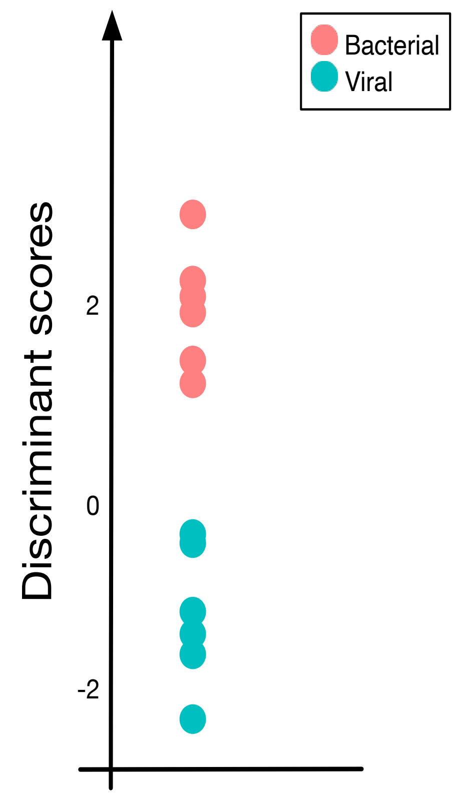
	Infection	CRP	Temp	scores
1	Viral	40.0	36.0	29.600
2	Viral	11.1	37.2	27.261
3	Viral	30.0	36.5	28.850
4	Viral	21.4	39.4	29.934
5	Viral	10.7	39.6	28.897
6	Viral	3.4	40.7	28.864
7	Bacterial	42.0	37.6	30.940
8	Bacterial	31.1	42.2	32.961
9	Bacterial	50.0	38.5	32.450
10	Bacterial	60.4	39.4	34.224
11	Bacterial	45.7	38.6	32.047
12	Bacterial	17.3	42.7	31.793



LDA

$$LD = 0.11 \cdot (\text{CRP} - \overline{\text{CRP}}) + 0.70 \cdot (\text{Temp} - \overline{\text{Temp}})$$

	Infection	CRP	Temp	scores	centered scores
1	Viral	40.0	36.0	29.600	−1.05175
2	Viral	11.1	37.2	27.261	−3.39075
3	Viral	30.0	36.5	28.850	−1.80175
4	Viral	21.4	39.4	29.934	−0.71775
5	Viral	10.7	39.6	28.897	−1.75475
6	Viral	3.4	40.7	28.864	−1.78775
7	Bacterial	42.0	37.6	30.940	0.28825
8	Bacterial	31.1	42.2	32.961	2.30925
9	Bacterial	50.0	38.5	32.450	1.79825
10	Bacterial	60.4	39.4	34.224	3.57225
11	Bacterial	45.7	38.6	32.047	1.39525
12	Bacterial	17.3	42.7	31.793	1.14125



Similarities between PCA and LDA

- Both rank the new axes in order of importance
  - PC1 (the first new axis that PCA creates) accounts for the most variation in the data.
    * PC2 (the second new axis) does the second best job ...
  - LD1 (the first new axis that LDA creates) accounts for the most variation between categories.
    * LD2 (the second new axis) does the second best job ...
- Both can tell you which variables are driving the new axes.

LDA in R

- Key function: lda function in the MASS package: e.g.,

```
lda(df$Infection ~ df$CRP + df$Temp)
```

```
Call:
lda(df$Infection ~ df$CRP + df$Temp)

Prior probabilities of groups:
Bacterial     Viral
      0.5       0.5

Group means:
            df$CRP  df$Temp
Bacterial 41.08333 39.83333
Viral     19.43333 38.23333

Coefficients of linear discriminants:
               LD1
df$CRP  -0.1060934
df$Temp -0.7011204
```

LDA

LDA has the following assumptions:

- The data are assumed to follow a Gaussian distribution.
- The covariance matrices of different classes/groups are equal.
- The data are linearly separable.

Setup with Two Populations

- Let $\{\mathbf{x}_1^1, \ldots, \mathbf{x}_{n_1}^1\}$ be n_1 observations from the group C_1.
- Let $\{\mathbf{x}_1^2, \ldots, \mathbf{x}_{n_2}^2\}$ be n_2 observations from the group C_2.
- Let $\mathbf{v}$ be a unit vector. Then the projection of $\mathbf{x} \in C_1 \cup C_2$ on the line represented by $\mathbf{v}$ is $\mathbf{v}^\top \mathbf{x}$.
- Let $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ denote the group means in C_1 and C_2, respectively, before the projection.
- Then the projected group mean $\tilde{\mu}_i$ is given by

$$\tilde{\mu}_i := \frac{1}{n_i} \sum_{\mathbf{x} \in C_i} \mathbf{v}^\top \mathbf{x} = \mathbf{v}^\top \boldsymbol{\mu}_i, i = 1, 2.$$

Mathematical Formulation

- **Scatter** matrix: sample variance $\times$ # of samples.
- Fisher's LDA is to maximize $J(v)$ with respect to $\mathbf{v}$ where

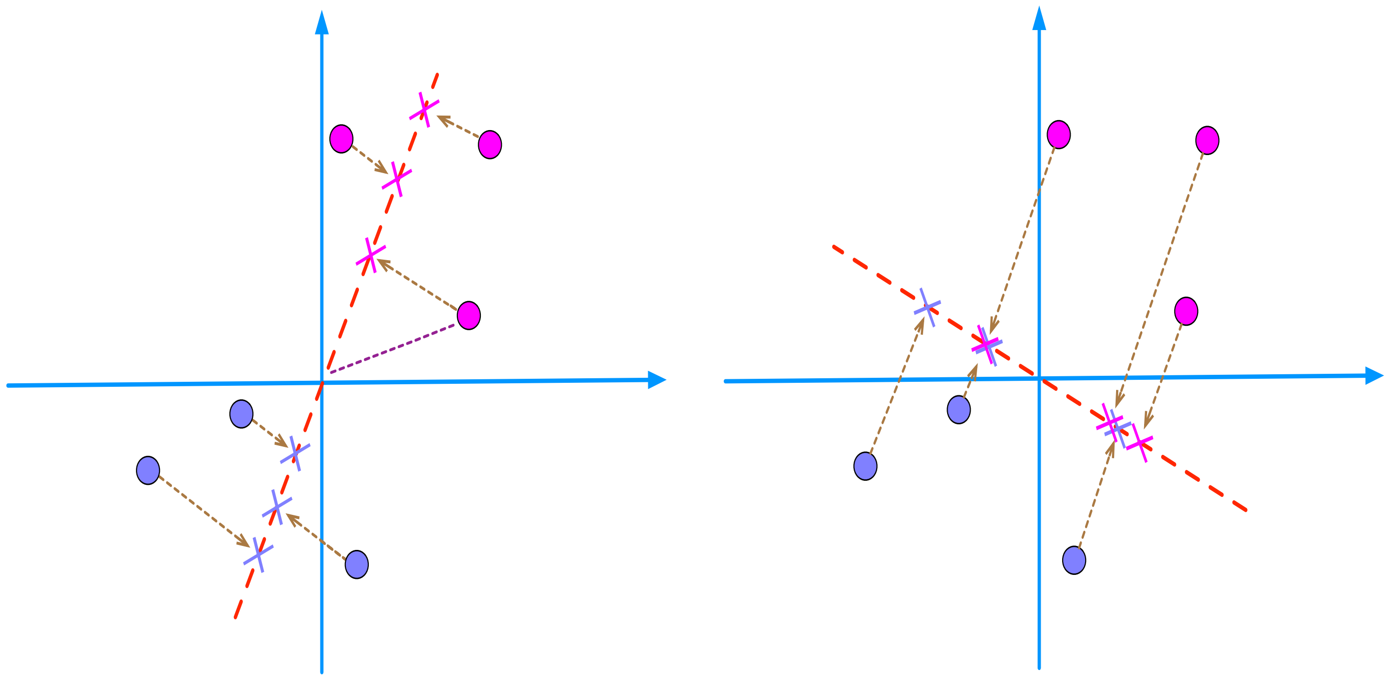
$$J(\mathbf{v}) = \frac{(\mathbf{v}^\top \boldsymbol{\mu}_1 - \mathbf{v}^\top \boldsymbol{\mu}_2)^2}{S_1^2 + S_2^2}$$

  in which $S_i^2 = \sum_{\mathbf{x} \in C_i} (\mathbf{v}^\top \mathbf{x} - \tilde{\mu}_i)^2$ is the scatter of C_i after the projection.
- LDA maximizes the ratio of between-class variance to within-class variance.

Geometric Intuition

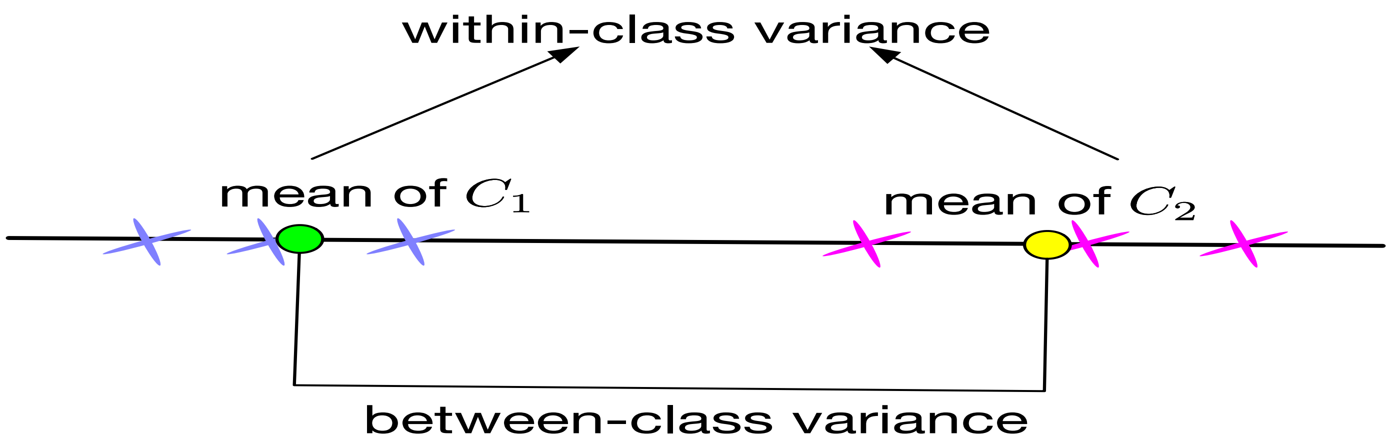
Two criteria are used by LDA to create a new axis defined by $\mathbf{v}$:

- Maximize the distance between the means of the two classes
- Minimize the variation within each class



Mathematical Formulation

- Within-class scatter S_w: measures the spread around means of each class.
  - $S_w := s_1 + s_2$ is the within-class scatter matrix with $s_i := \sum_{\mathbf{x} \in C_i} (\mathbf{x} - \boldsymbol{\mu}_i)(\mathbf{x} - \boldsymbol{\mu}_i)^\top$.
- Between-class scatter S_b: measures the distance between class means.
  - $S_b := (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top$ is the between-class scatter matrix.



Mathematical Formulation

- $J(\mathbf{v})$ can be equivalently written as

$$J(\mathbf{v}) = \frac{(\mathbf{v}^\top \boldsymbol{\mu}_1 - \mathbf{v}^\top \boldsymbol{\mu}_2)^2}{S_1^2 + S_2^2} = \frac{\mathbf{v}^\top S_b \mathbf{v}}{\mathbf{v} S_W \mathbf{v}}.$$

- This optimization problem can be shown to be equivalent to solve the following eigen equation

$$M\mathbf{v} = \lambda \mathbf{v}$$

with $\lambda := \frac{\mathbf{v}^\top S_b \mathbf{v}}{\mathbf{v}^\top S_w \mathbf{v}}$ and $W := S_w^{-1} S_b$.

- The maximum separation occurs when $\mathbf{v} \propto S_w^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$.
- $\mathbf{v}$ is the normal to the discriminant hyperplane.
- No general rule is available to separate the two groups, but a good choice is $\mathbf{v}^\top \mathbf{x} > c$ where $c = \mathbf{v}^\top \cdot \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$.

Comments

- LDA maximizes between-class scatter while minimizing within-class scatter.
- LDA assumes Gaussian distribution and identical covariance matrices for groups.
- LDA can be extended to multi-class problems and address some limitations of logistic regression.
- The next lecture will focus on the general formulation of LDA.