Topics in Data Science

Lecture 17

Department of Mathematics and Statistics

PURDUE UNIVERSITY NORTHWEST

Discriminant Analysis

- Discriminant Analysis (DA) is a multivariate classification technique that separates objects into two or more mutually exclusive groups based on measurable features of those objects. The measurable features are sometimes called predictors or independent variables, while the classification group is the response or what is being predicted.
- DA uses data where the classes are known beforehand to create a model that may be used to predict future observations. It's useful as an analytic technique trying to understand the relationship between independent variables and a discrete dependent variable. It differs from regression analysis in that the dependent variable must be discrete. It differs from cluster analysis in that the classes must be known before the model is created.

Discriminate Analysis

Discriminant Function Analysis (DA) undertakes the same task as multiple linear regression by predicting an outcome. However, multiple linear regression is limited to cases where the dependent variable on the Y axis is an interval variable so that the combination of predictors will, through the regression equation, produce estimated mean population numerical Y values for given values of weighted combinations of X values. DA is used when:

- the dependent is categorical with the predictor IV's at interval level such as age, income, attitudes, perceptions, and years of education, although dummy variables can be used as predictors as in multiple regression. Logistic regression IV's can be of any level of measurement.
- There are more than two DV categories, unlike logistic regression, which is limited to a dichotomous dependent variable.

Types of Discriminant Analysis

- **Linear Discriminant Analysis (LDA):** This is the most commonly used form of discriminant analysis. It assumes that the predictor variables are normally distributed and have equal covariance matrices across groups.
- **Quadratic Discriminant Analysis (QDA):** Unlike LDA, QDA relaxes the assumption of equal covariance matrices and allows for different variances and covariances across groups. It is more flexible but requires a larger sample size.
- **Regularized Discriminant Analysis (RDA):** RDA is a modified version of LDA that handles situations where the number of predictor variables is larger than the number of observations. It uses regularization techniques to prevent overfitting.
- **Flexible Discriminant Analysis (FDA):** Uses non-linear combinations of inputs such as splines to handle non-linear separability.

Assumptions of Discriminant Analysis

- **Normality:** Discriminant analysis assumes that the predictor variables are normally distributed within each group. Deviations from normality can affect the accuracy of the classification.
- **Linear Separability:** In LDA, a linear decision boundary should be sufficient to separate the classes. Non-linear relationships may require alternative methods such as non-linear discriminant analysis.
- **Homoscedasticity:** Homoscedasticity assumes that the variances of the predictor variables are equal across groups. Violations of this assumption may lead to biased classifications.

Interpretation of Discriminant Analysis

- **Discriminant Function:** The discriminant function represents the linear combination of predictor variables that maximally separates the groups. It can be used to classify new cases into the appropriate groups.
- **Discriminant Loadings:** These coefficients indicate the relative importance of each predictor variable in discriminating between groups. Larger loadings suggest stronger discriminatory power.
- **Wilks' Lambda:** This statistic measures the overall significance of the discriminant analysis. A smaller value indicates a more significant discrimination between groups.

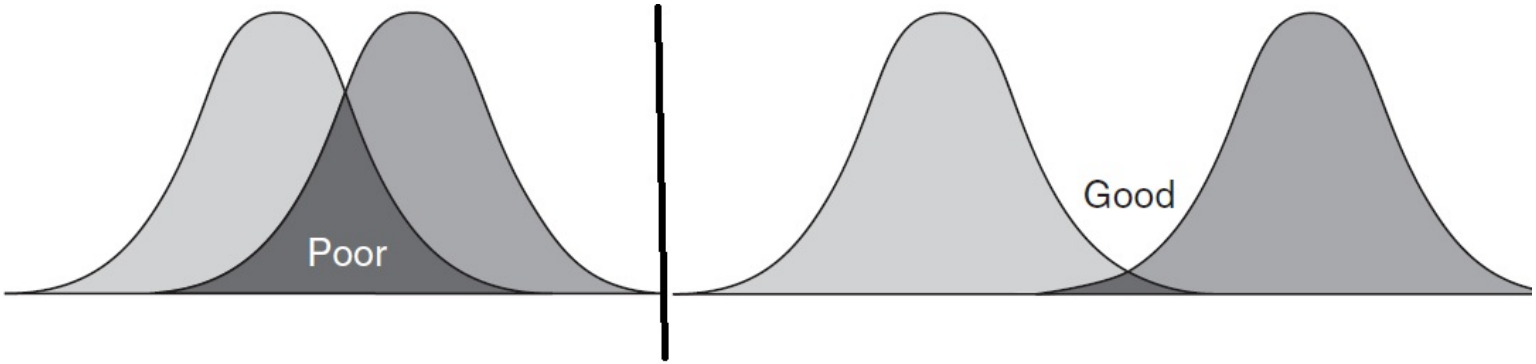
Benefits of Discriminant Analysis

- **Predictive Power:** Discriminant analysis helps predict group membership based on the given predictor variables, allowing for accurate classification of future cases.
- **Variable Importance:** It identifies the most important predictor variables that contribute to group differentiation, aiding in understanding the key factors driving the classification.
- **Data Reduction:** Discriminant analysis can reduce a large number of predictor variables to a smaller set of discriminant functions, simplifying the analysis and interpretation

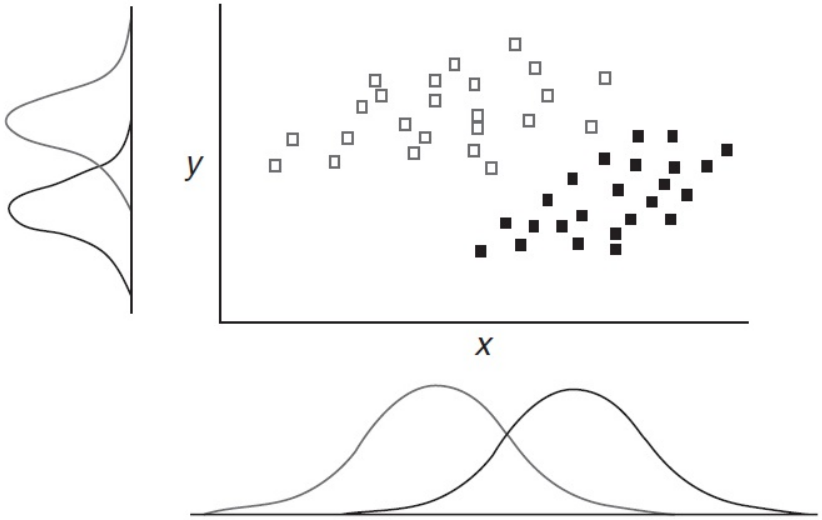
Limitations of Discriminant Analysis

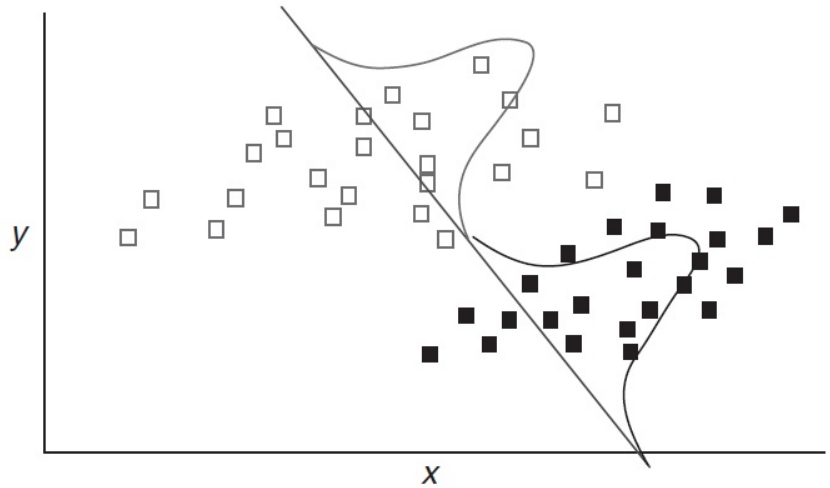
- **Assumption Sensitivity:** Discriminant analysis relies on several assumptions, such as normality and linearity, which may not always hold in real-world datasets.
- **Overfitting:** In situations where the number of predictor variables exceeds the sample size, overfitting can occur, leading to poor generalization and inaccurate predictions.
- **Multicollinearity:** High correlations among predictor variables can affect the stability and interpretability of the discriminant function coefficients.

Discriminant Analysis



Not always have a standard shape





How does LDA works?

LDA works by finding directions in the feature space that best separate the classes. It does this by maximizing the difference between the class means while minimizing the spread within each class.
Let's assume that we have a cancer drug

- It works great for some people
- It makes other people feel worse

How do we decide who to give the drug to?
We may look into the gene expression data

Maximizing Class Separability : Role of LDA

LDA is used to identify a linear combination of features that best separates classes within a dataset. For example we have two classes that need to be separated efficiently. Each class may have multiple features and using a single feature to classify them may result in overlapping.

To solve this LDA is used as it uses multiple features to improve classification accuracy.

LDA

When data points belonging to two classes are plotted if they are not linearly separable, LDA will attempt to find a projection that maximizes class separability.

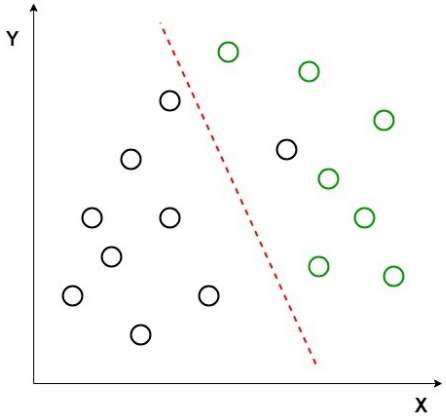


Image shows an example where the classes (black and green circles) are not linearly separable. LDA attempts to separate them using red dashed line. It uses both axes (X and Y) to generate a new axis in such a way that it maximizes the distance between the means of the two classes while minimizing the variation within each class. This transforms the dataset into a space where the classes are better separated

LDA

- After transforming the data points along a new axis LDA maximizes the class separation. This new axis allows for clearer classification by projecting the data along a line that enhances the distance between the means of the two classes.
- Perpendicular distance between the decision boundary and the data points helps us to visualize how LDA works by reducing class variation and increasing separability.
- After generating this new axis using the above-mentioned criteria, all the data points of the classes are plotted on this new axis.

How LDA works?

Let's assume we have two classes with d-dimensional samples such as $x_1, x_2, \cdots, x_n$ where

- n_1 samples belong to c_1
- n_2 samples belong to c_2

Note that if x_i represents a data point, its projection onto the line separated by the unit vector v is $v^T x_i$. Let μ_1 and μ_2 are the means of the classes c_1 and c_2 respectively. After projection, the new mean is $\hat{\mu}_1 = v^T \mu_1$ and $\hat{\mu}_2 = v^T \mu_2$. We want to normalize the difference $|\hat{\mu}_1 - \hat{\mu}_2|$ to maximize the class separation. The scatteredness is calculated as:
$s_1^2 = \sum x_i \in c_1 (x_i - \mu_1)^2$
$s_2^2 = \sum x_i \in c_2 (x_i - \mu_2)^2$
The goal is to maximize the ratio of the between-class scatter to the within-class scatter, which leads us to the following criterion:

$$J(v) = \frac{|\hat{\mu}_1 - \hat{\mu}_2|}{s_1^2 + s_2^2}$$

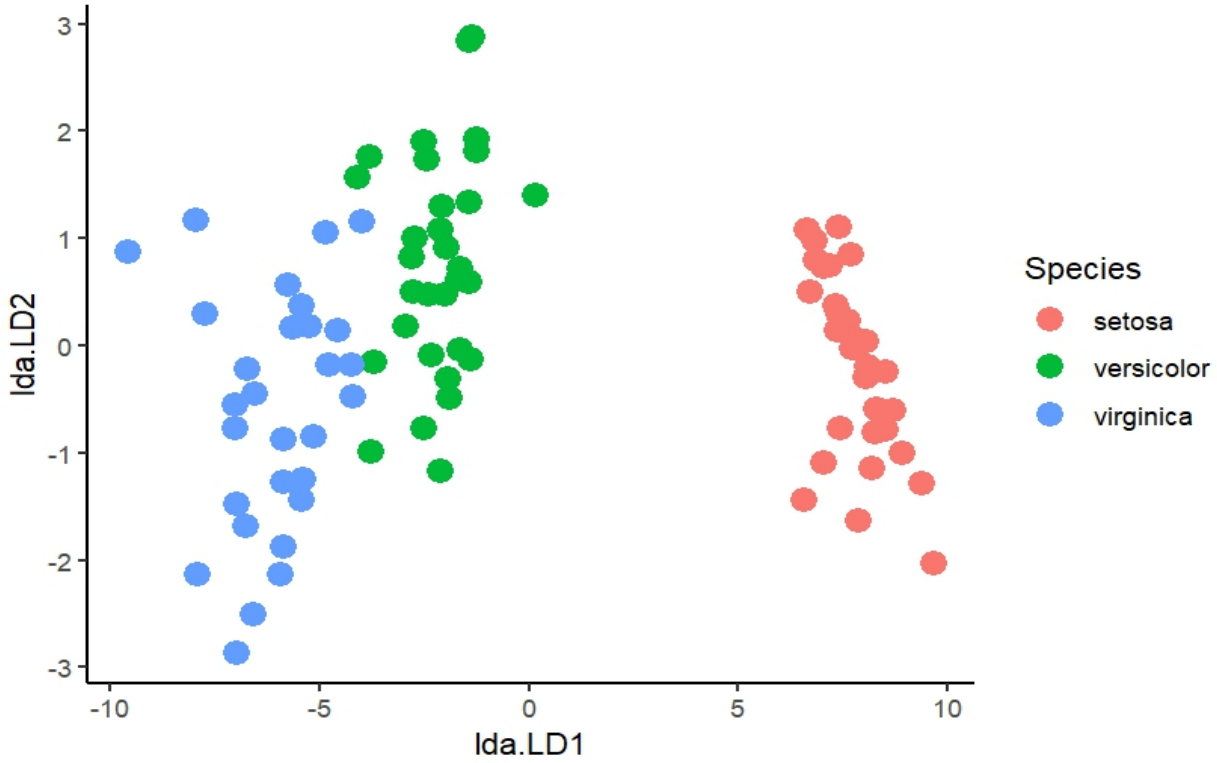
Linear Discriminant Analysis-Example

Consider the iris data set. The length and width measurements are the independent variables that will be used to predict the flower's species, the discrete dependent variable.

```
> library(tidyverse)
> library(MASS)
> library(klaR)
> head(iris)
> training_sample <- sample(c(TRUE, FALSE), nrow(iris),
    replace = T, prob = c(0.6,0.4))
> train <- iris[training_sample, ]
> test <- iris[!training_sample, ]
> lda.iris <- lda(Species ~ ., train)
> lda.iris
> plot(lda.iris)
```

LDA-Example contd.

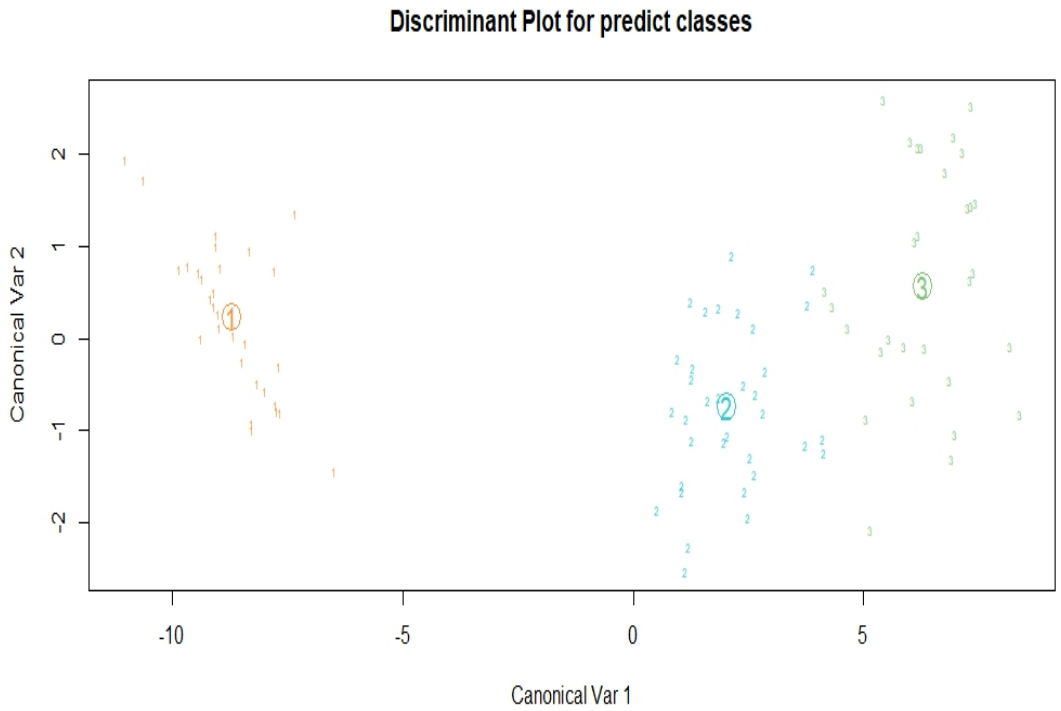
```
library(ggplot2)
lda_df <- data.frame(Species= train[, "Species"],
                     lda = predict(lda.iris)$x)
ggplot(lda_df)+
geom_point(aes(x =lda.LD1,y=lda.LD2, color=Species),size=4) +
  theme_classic()
```



LDA example-contd.

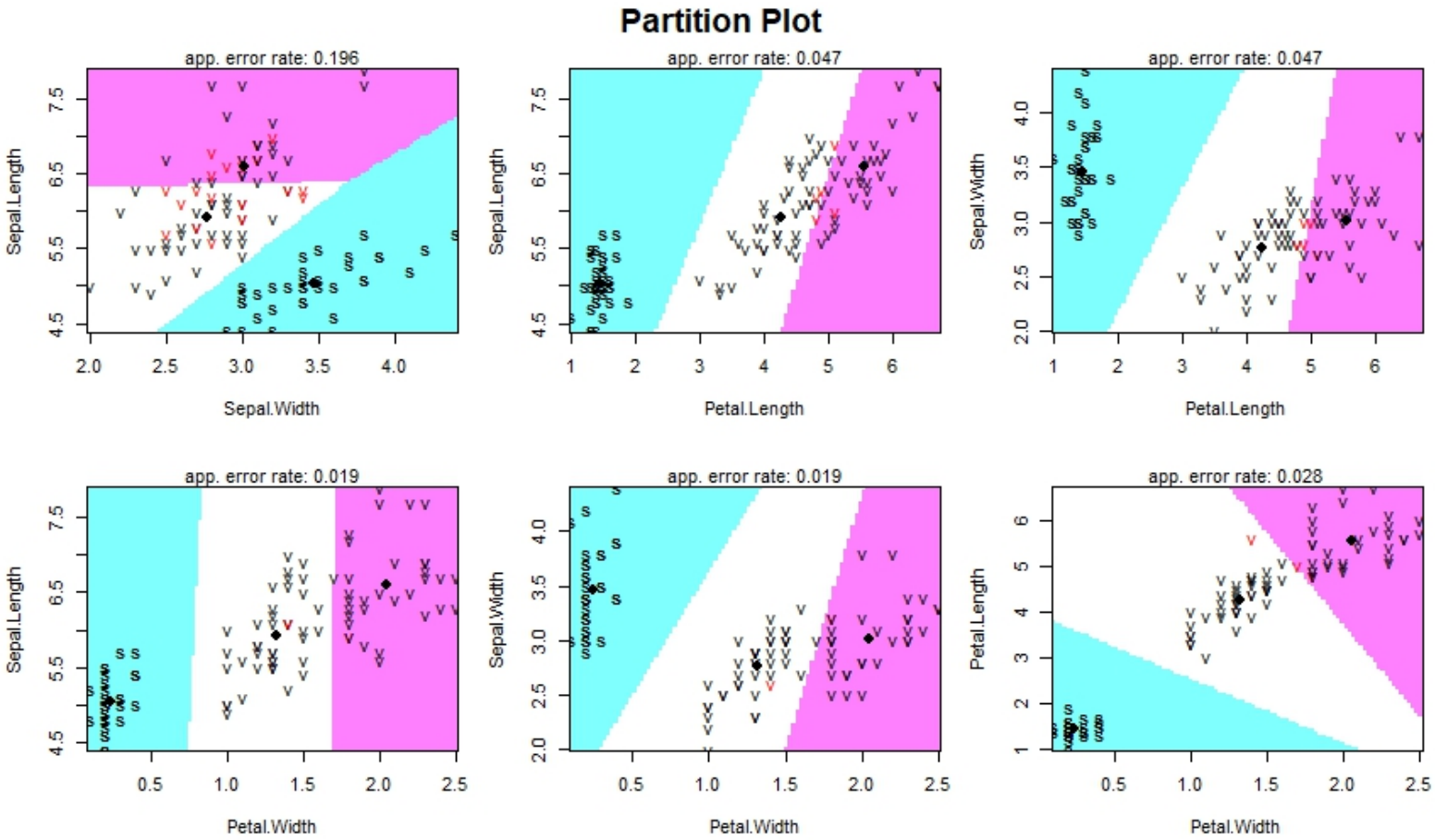
For flexible discriminant analysis

```
> library(mda)
> m2=fda(Species~., data=train)
> plot(m2)
> table(train$Species, predict(m2))
> T=table(train$Species, predict(m2))
> sum(diag(T))/sum(T)
```



LDA Partition Plots

```
>library(klaR)
> partimat(Species ~ Sepal.Length + Sepal.Width +
Petal.Length + Petal.Width, data=train, method="lda")
```



Assumptions underlying LDA

- Independent subjects.
- Normality: the variance-covariance matrix of the predictors is the same in all groups.
- If the latter assumption is violated: use quadratic discriminant analysis in the same manner as linear discriminant analysis.

Quadratic discriminant analysis

- Quadratic discriminant analysis is a method you can use when you have a set of predictor variables and you'd like to classify a response variable into two or more classes. It is considered to be the non-linear equivalent to linear discriminant analysis.
- Quadratic Discriminant Analysis (QDA), an extended version of Linear Discriminant Analysis (LDA), is a method used for classifying data points.
- While LDA operates under the premise that all classes have a shared covariance matrix, QDA diverges by allowing individual covariance matrices for each class. This distinction endows QDA with added flexibility, although at the cost of estimating additional parameters.

```
> library(MASS)
> library(ggplot2)
> attach(iris)
> sample <- sample(c(TRUE, FALSE), nrow(iris),
                   replace=TRUE, prob=c(0.7,0.3))
> train <- iris[sample, ]
> test <- iris[!sample, ]
> model <- qda(Species~., data=train)
> model
> predicted <- predict(model, test)
> head(predicted$class)
> head(predicted$posterior)
> mean(predicted$class==test$Species)
```

QDA Partition Plots

```
> library(klaR)
> partimat(Species ~ Sepal.Length + Sepal.Width +
Petal.Length + Petal.Width, data=train, method="qda")
```