# Discriminant Analysis (DA)

# Supervised vs Unsupervised Learning*

- **Supervised Learning**
  - **Predict some feature of interest** (e.g., disease type, cell type, disease prognosis, species of an organism) **by using one or more other variables** (the predictors)
  - Typically we construct prediction rule from training data
  - We want to  predict new cases and have some criteria to assess the quality of prediction rule

- **Unsupervised learning**
  - Here all variables are on an equal footing.
  - Techniques include
    - Principal Components Analysis, PCA
    - Multidimensional Scaling, MDS, nMDS
    - Hierarchical and non-Hierarchical Clustering (todo)
  - We want to
    - Discover patterns, ordinations
    - Reduce dimensions produce low- dimensional maps to display data and data patterns

*terms from machine learning

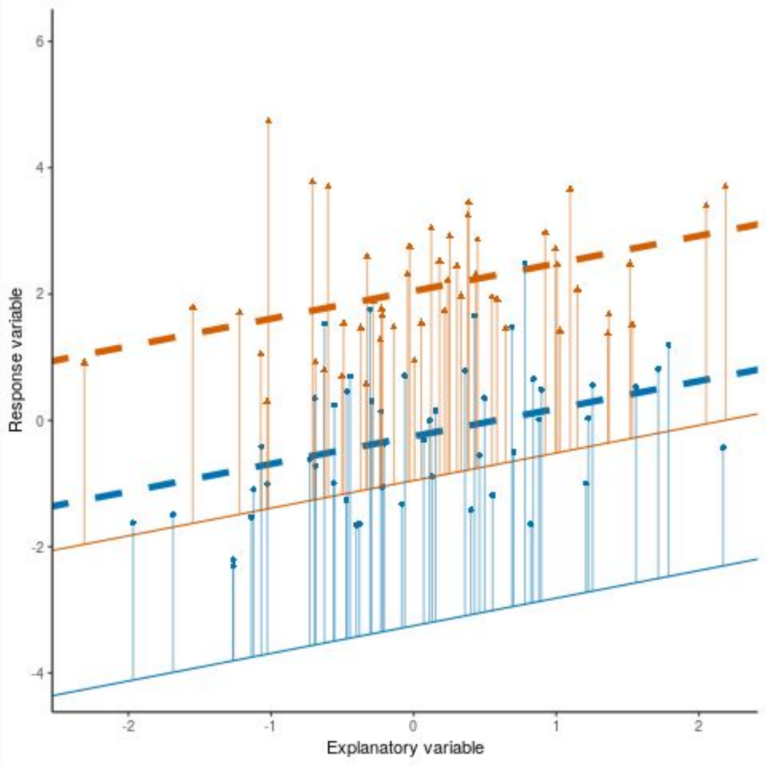
# Supervised vs Unsupervised Learning*

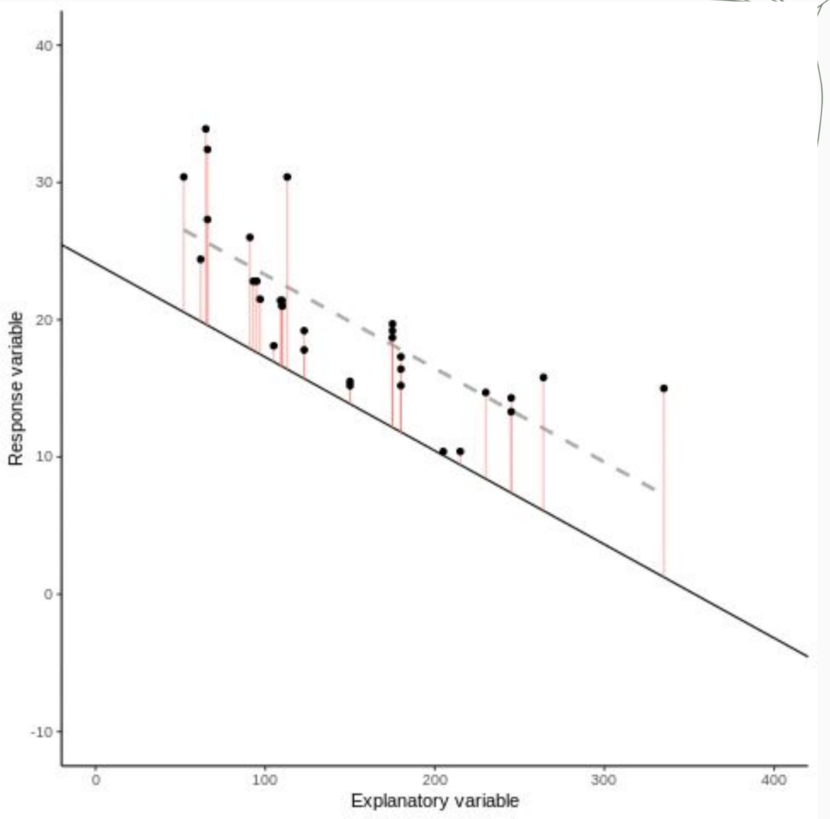
- **Supervised Learning**
  - Find a prediction rule to predict value of one feature, objective or response from the other features
  - Response can be continuous or categorical
    - Continuous response: regression
    - Categorical response: classification

- **Unsupervised learning**
  - Find similarities or distances between cases based on their feature values, e.g. clustering, ordination

*terms from machine learning

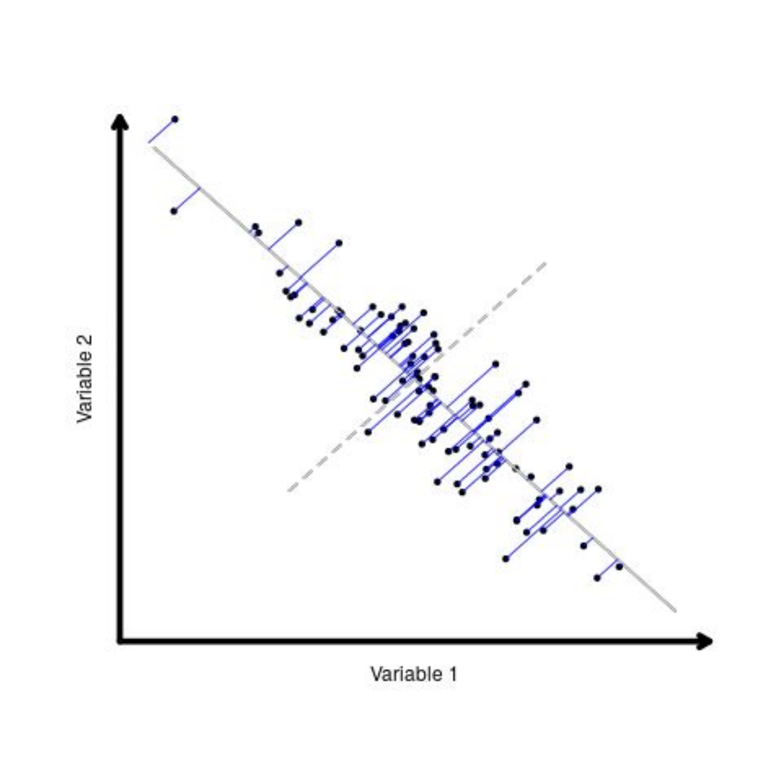
# Linear discriminant analysis all about y = mx + c …



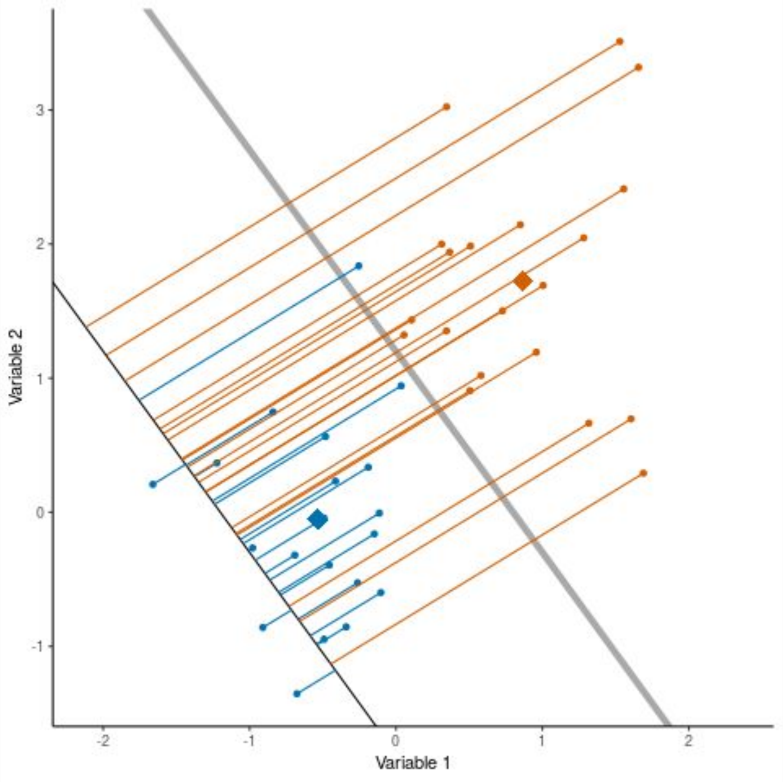



# PCA still has the straight lines

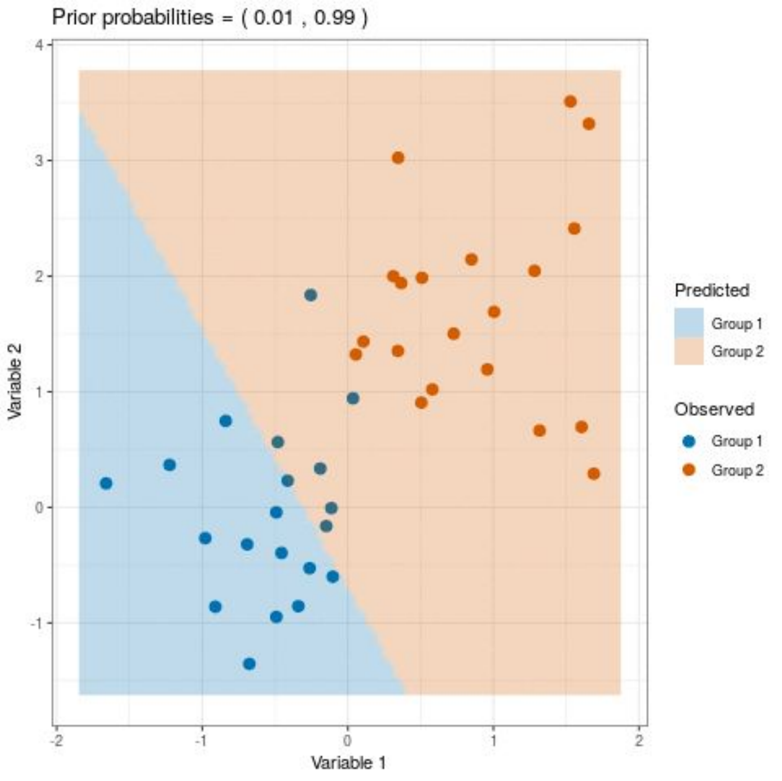
**Note the slight difference in how we calculate distance (point to line, rather than point to point)**



# So, what if we want to separate/discriminate groups?



# So, what if we want to separate/discriminate groups?

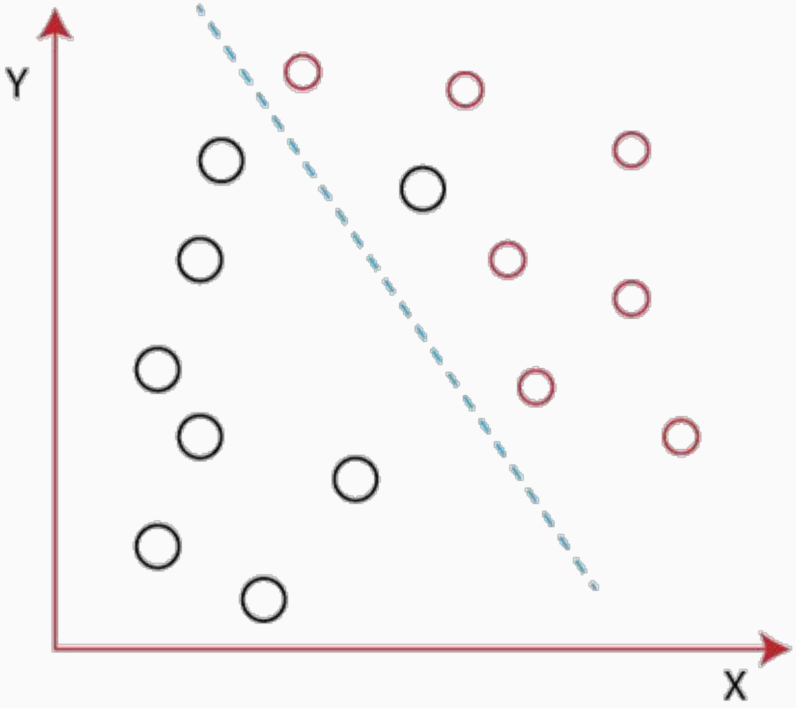


# LDA

- Goal is to predict some feature of interest using one or more variables
- Like regression but predicting groups: think ANOVA meets PCA
- Finding a linear combination of features that classifies observations into groups

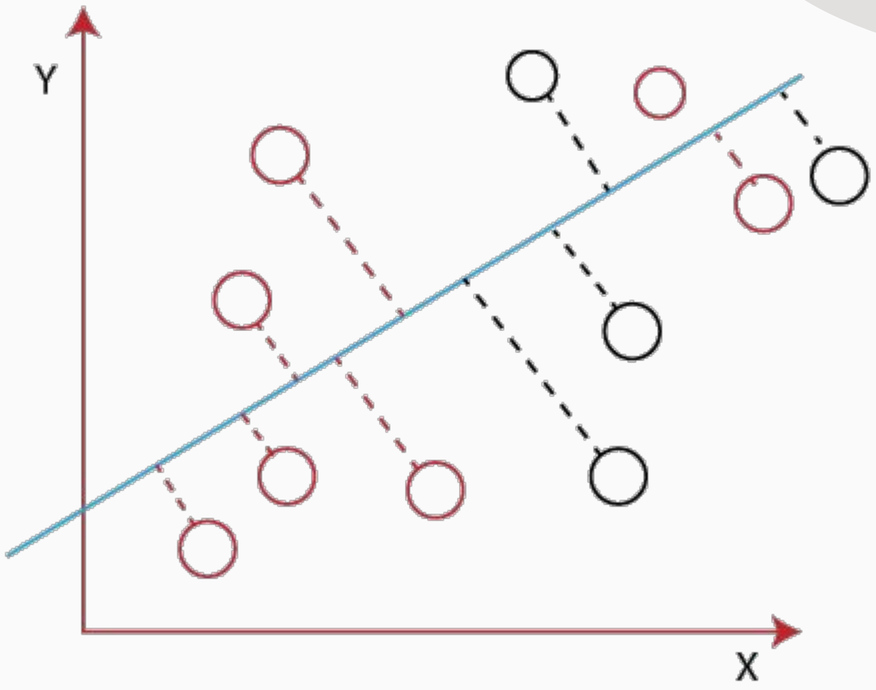
# LDA

- It is impossible to draw a straight line in a 2-d plane that can separate these data points efficiently
- However, using linear discriminant analysis; we can dimensionally reduce the 2-D plane into the 1-D plane.
- Using this technique, we can also maximize the separability between multiple classes.



# LDA…

- Maximizes the distance between means of two classes, whilst
- Minimizing the variance within the individual class.



# What we should be aware of

- The structure of our training data
  - Remember talking about algorithmic bias in Module 1
  - How many observations are in each group (“prior probabilities”) will affect inference
  - It’s your choice to assess if the defaults are correct or not.

# To look out for

- The **proportion of trace**
    - Gives the **percentage separation** achieved by each discriminant function (think of eigenvalues /stress measures from PCA and MDS)
- The **discriminant functions** themselves
    - Give the linear combination of variables giving best separation between groups (think of reconstructing our principal components in PCA)
- The **confusion** (misclassification) **matrix**
    - **The misclassification rate tells us the proportion of cases we miss classified :)**
- **Prediction of new cases!** Using inbuilt functions we see
    - The predicted **class** (group) of the observations, and
    - The **posterior probabilities** for each observation belonging to each group

Glass Identification

**[code along, activity 3]**

PUBLISH ABOUT BROWSE

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

## Uncomfortably high: Testing reveals inflated THC potency on retail *Cannabis* labels

Anna L. Schwabe, Vanessa Johnson, Joshua Harrelson, Mitchell E. McGlaughlin

Published: April 12, 2023 • https://doi.org/10.1371/journal.pone.0282396