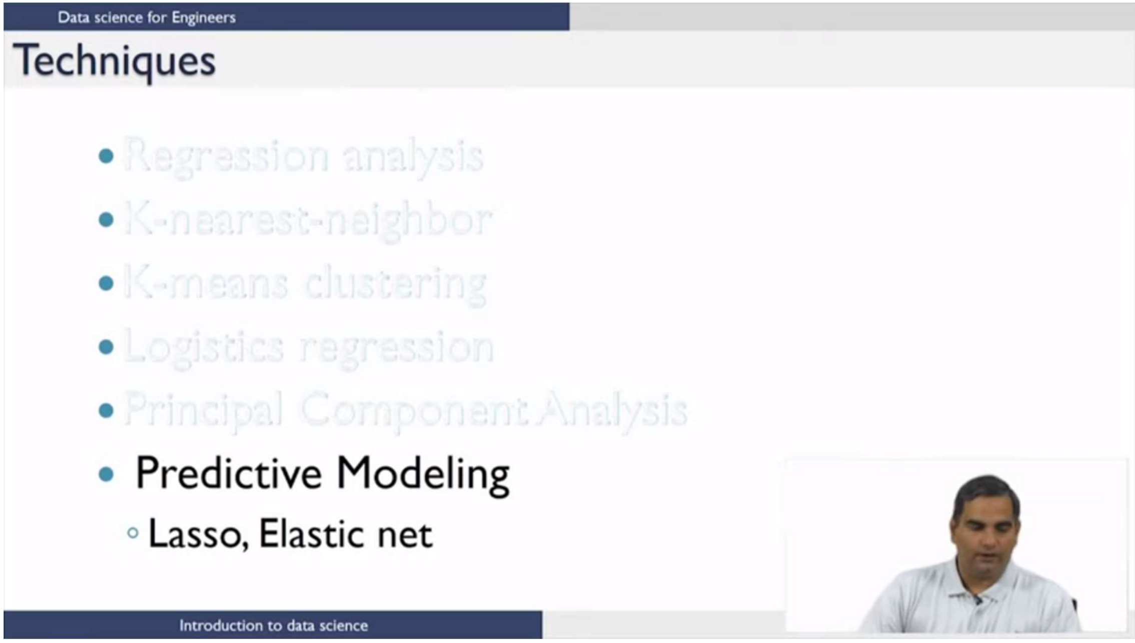
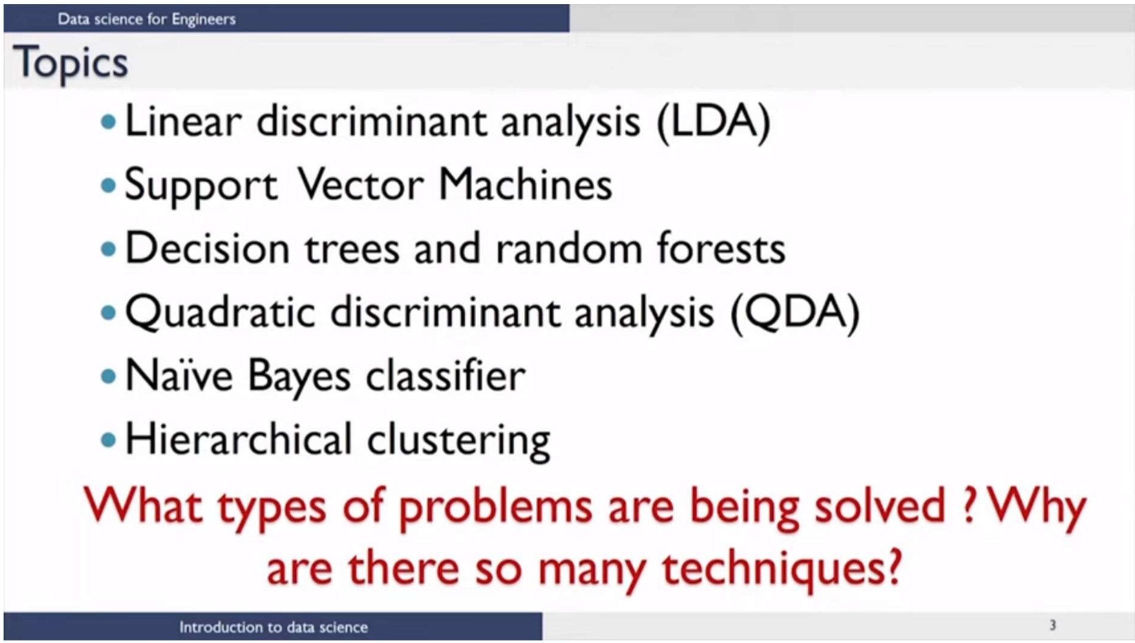
# Introduction to Data Science



## Techniques

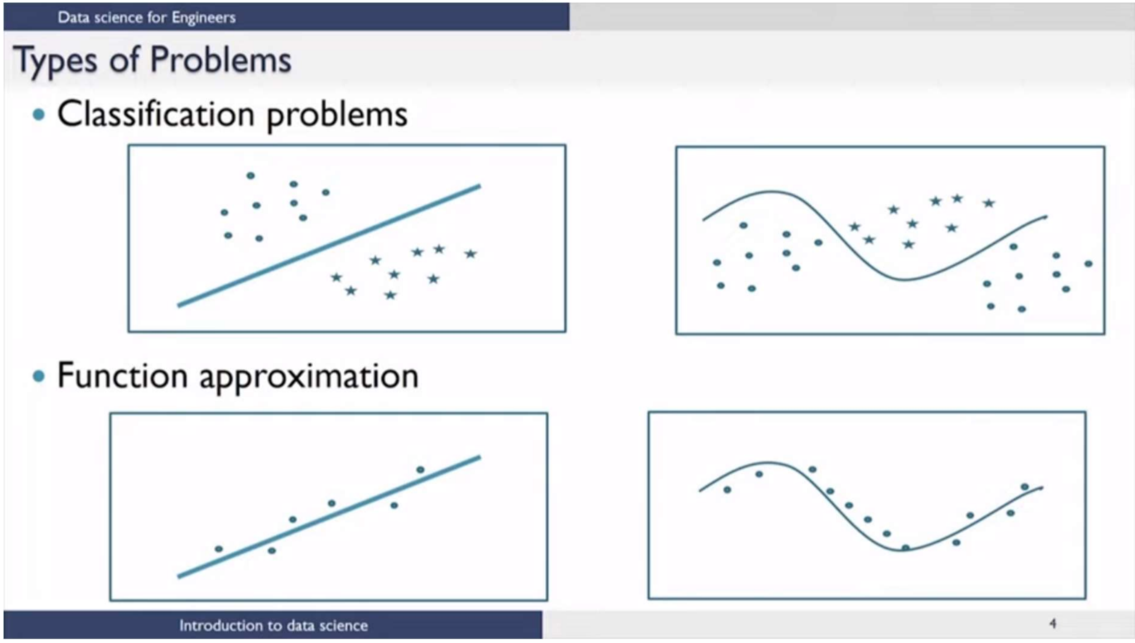
- Regression analysis
- K-nearest-neighbor
- K-means clustering
- Logistics regression
- Principal Component Analysis
- Predictive Modeling
  - Lasso, Elastic net



## Topics

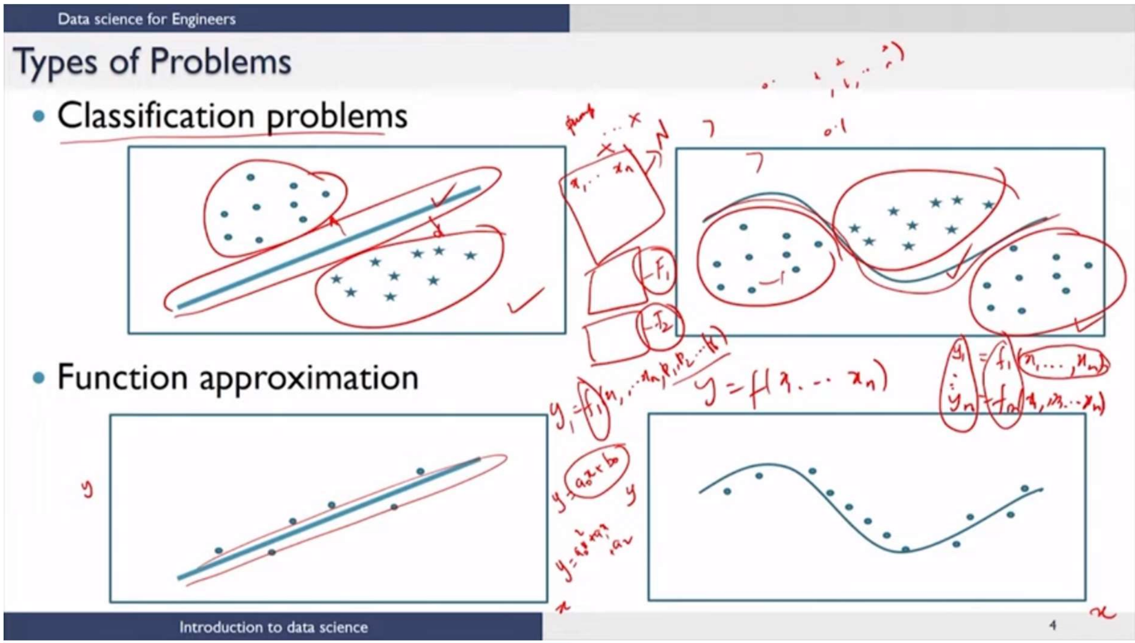
- Linear discriminant analysis (LDA)
- Support Vector Machines
- Decision trees and random forests
- Quadratic discriminant analysis (QDA)
- Naïve Bayes classifier
- Hierarchical clustering

What types of problems are being solved ? Why are there so many techniques?



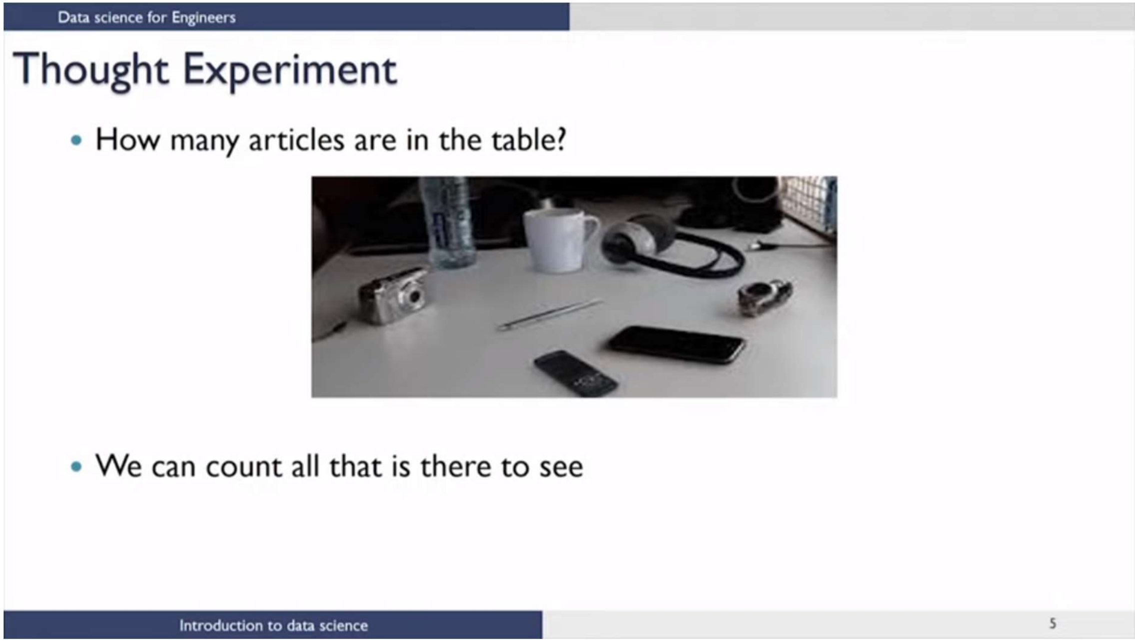
## Types of Problems

- Classification problems
- Function approximation



## Types of Problems

- Classification problems
- Function approximation



## Thought Experiment

- How many articles are in the table?
- We can count all that is there to see

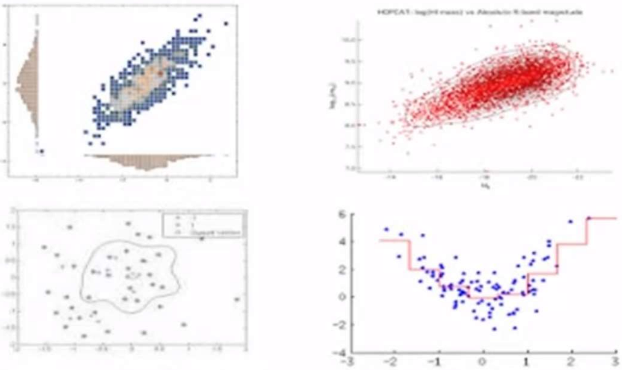
# Thought Experiment (Metaphorical)

- What about things that we cannot see?

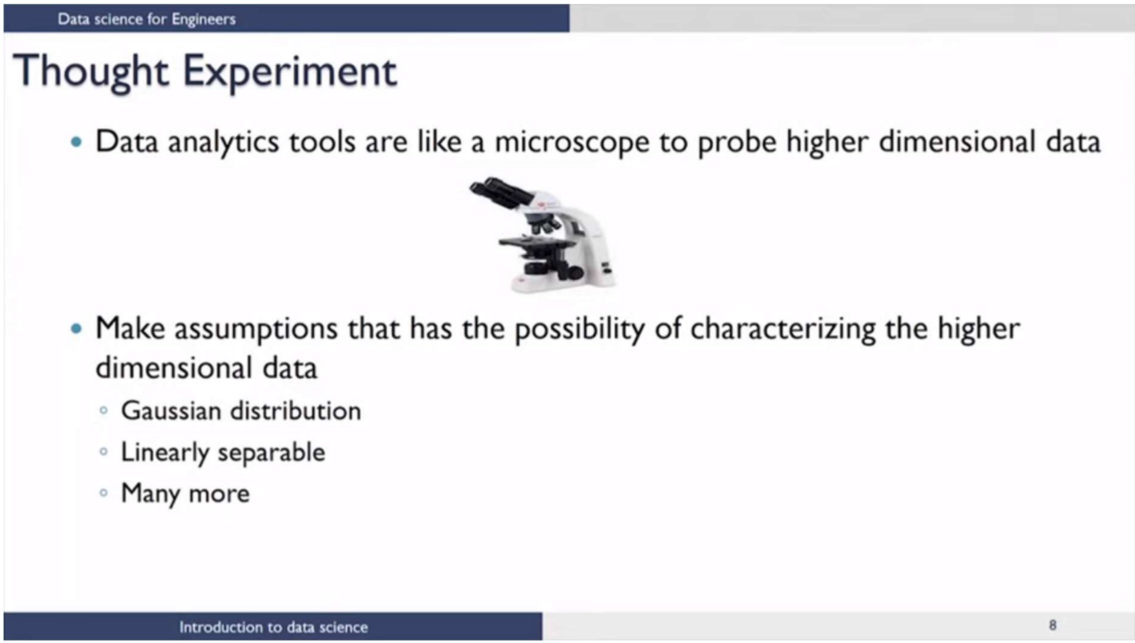
- How do we understand things that we cannot see – appropriate fluorescence chemical?

# Thought Experiment

- If world were 2D?

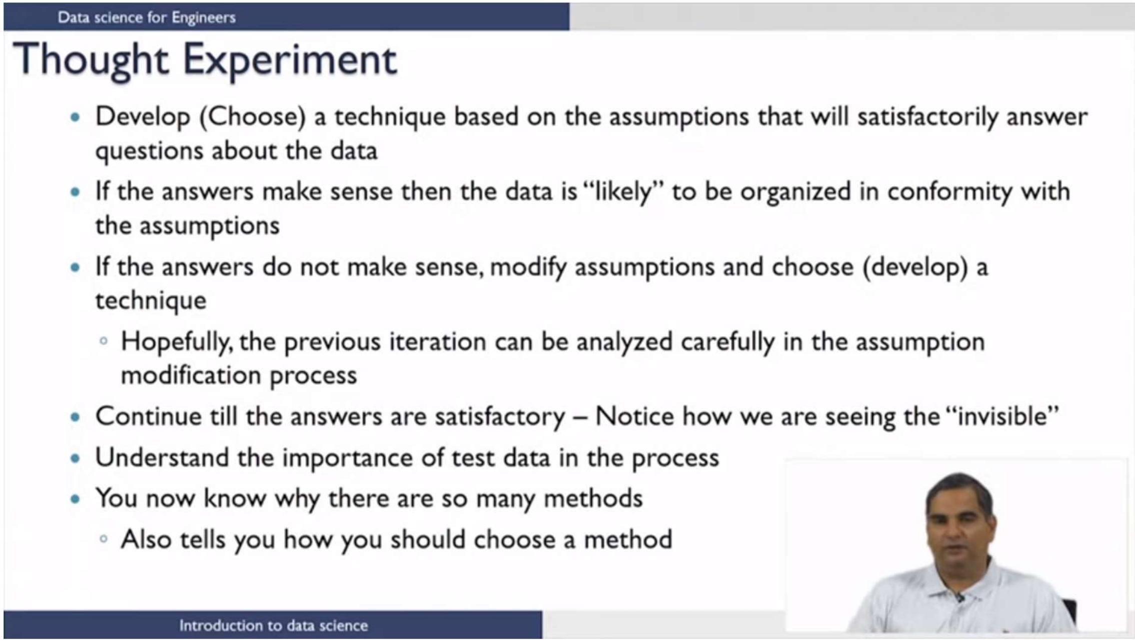


- Data analytics not as critical



## Thought Experiment

- Data analytics tools are like a microscope to probe higher dimensional data
- Make assumptions that has the possibility of characterizing the higher dimensional data
  - Gaussian distribution
  - Linearly separable
  - Many more



## Thought Experiment

- Develop (Choose) a technique based on the assumptions that will satisfactorily answer questions about the data
- If the answers make sense then the data is "likely" to be organized in conformity with the assumptions
- If the answers do not make sense, modify assumptions and choose (develop) a technique
  - Hopefully, the previous iteration can be analyzed carefully in the assumption modification process
- Continue till the answers are satisfactory – Notice how we are seeing the "invisible"
- Understand the importance of test data in the process
- You now know why there are so many methods
  - Also tells you how you should choose a method