Inspire…Educate…Transform.

# Data Science: Big Picture

## Also, introduction to CPEE program

**Dr. K. V Dakshinamurthy**
President, INSOFE

May 30, 2014

# Introduction

- An exciting and interesting career path
  - Visualization, communication skills
  - Programming and hacking skills
  - Mathematical skills

# Program

- Mentors
- Scientists
- Piazza
- R and Hadoop programming environment

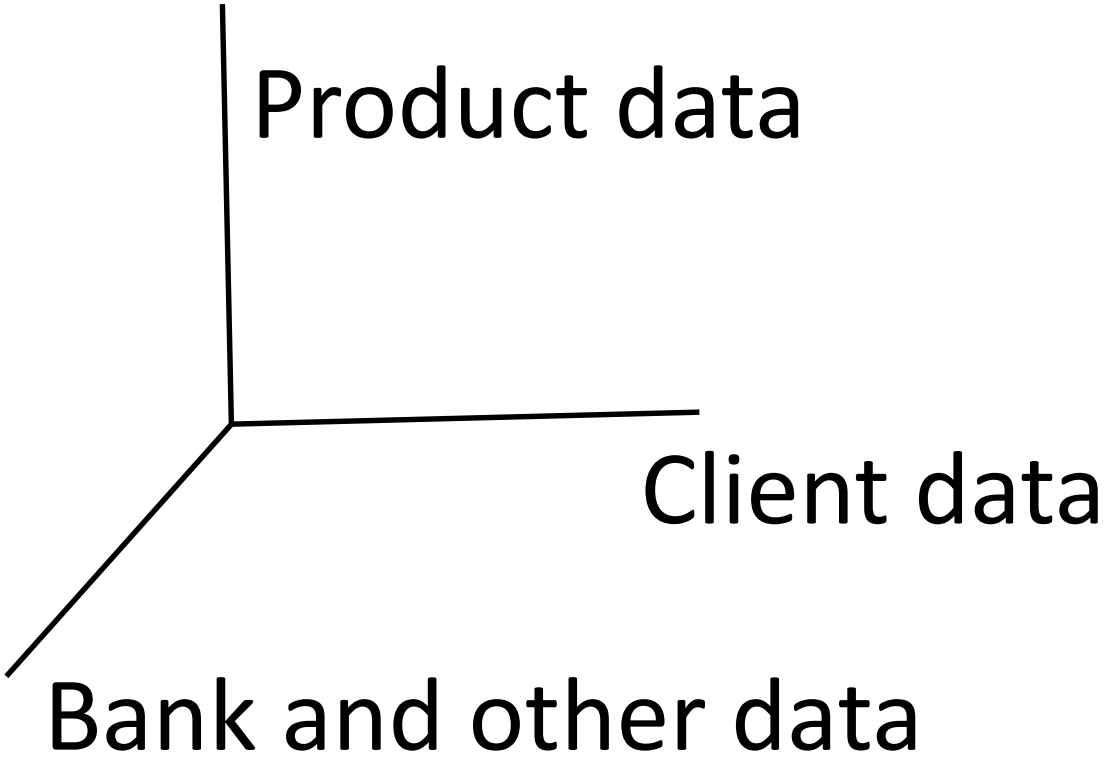
# Structure

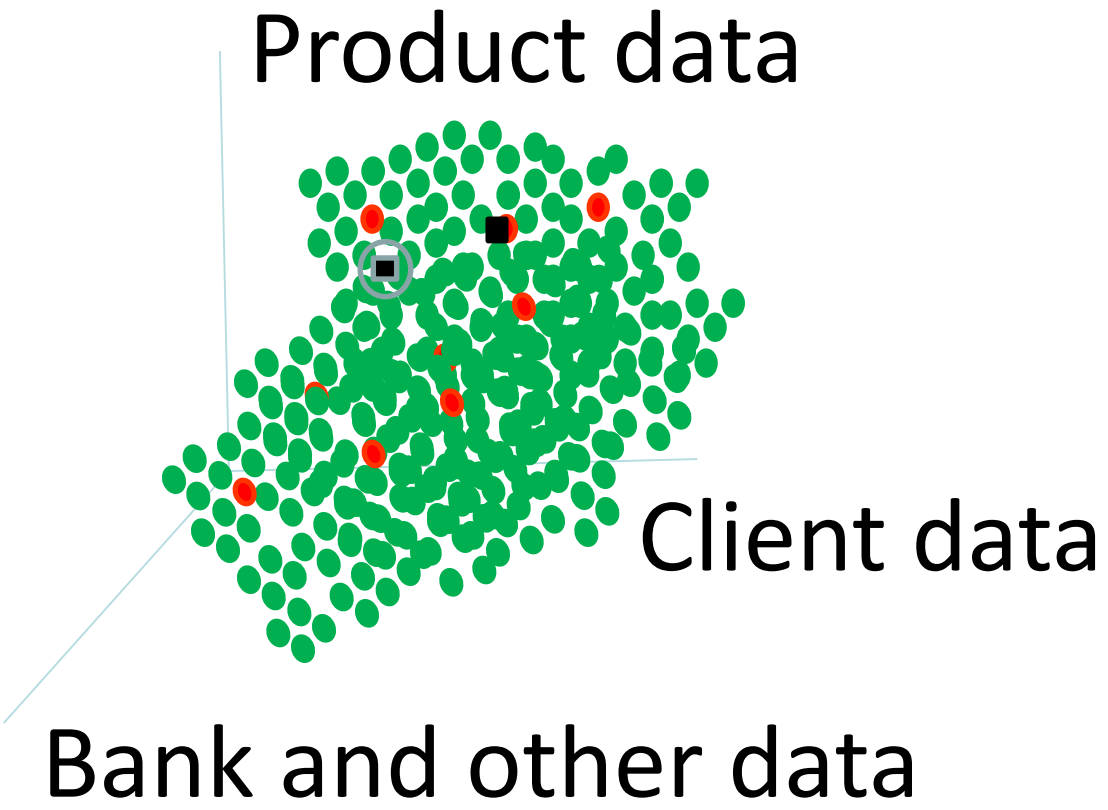
- Lectures:  Morning

- Afternoon:  Quizzes (written), lab and project

# Help banks minimize loss?

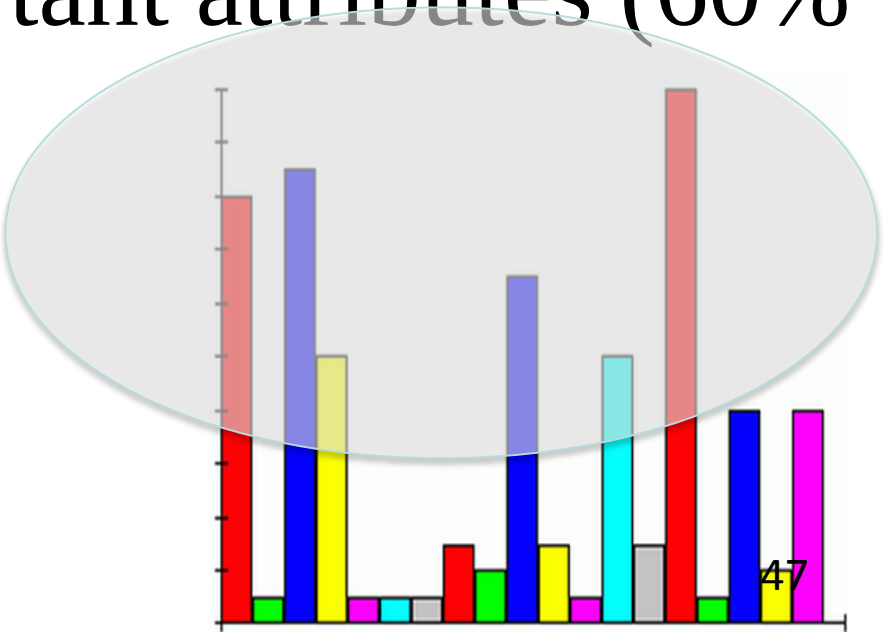
- What is the probability of default?
- Can we predict who is likely to not pay?


Product data
Client data
Bank and other data


Product data
Client data
Bank and other data

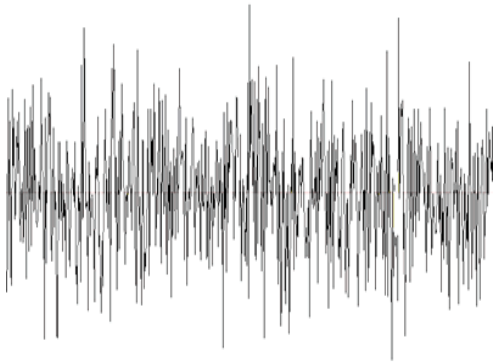
# Lots of experiments

- What attributes to use
  - Use all attributes on 10% data (50% accuracy)
  - Use only the important attributes (60% accuracy)



# Refining the attributes

- Is a 25 year old going to be different from 26?
- Group all below 30 as young
- Accuracy went up to 65%



# Data scientist's way of thinking

# GOAL

# Search for patterns and not theories

A specific metallic component is wearing out with use in an acidic environment.

How do we understand and use this component

# Approach 1…Science

- A chemist studies what happens to the material in acid and understands that the metal is reacting with the hydrogen in the acid to form a vapor. She develops the equations and science behind them that the wearing is a function of acidity and temperature.

# Approach 1…Engineering

- An engineer tests the degradation in a few concentrations and temperature, plots the degradation, validates the theory and comes up with thumb rules

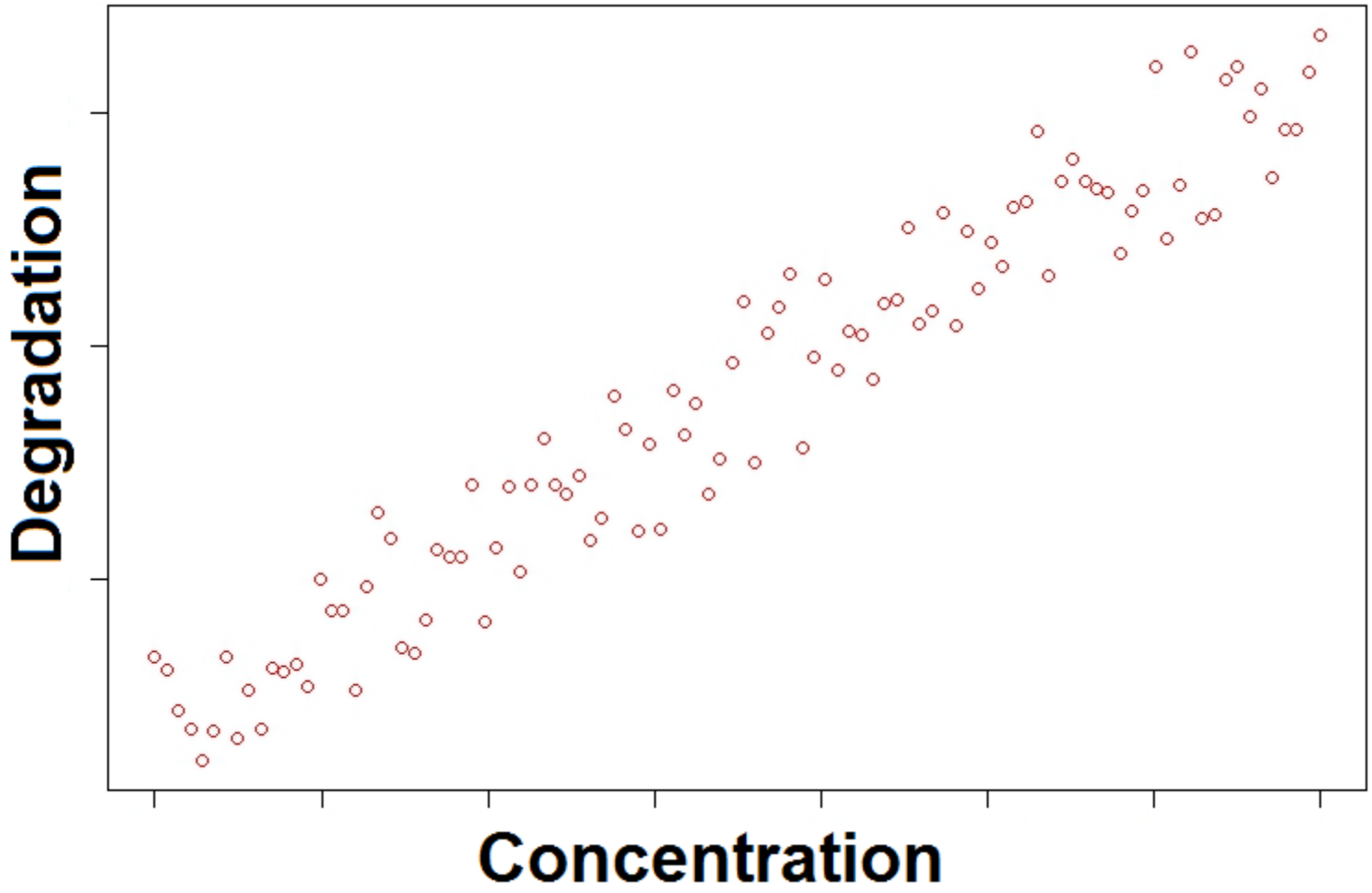
# Deductive learning

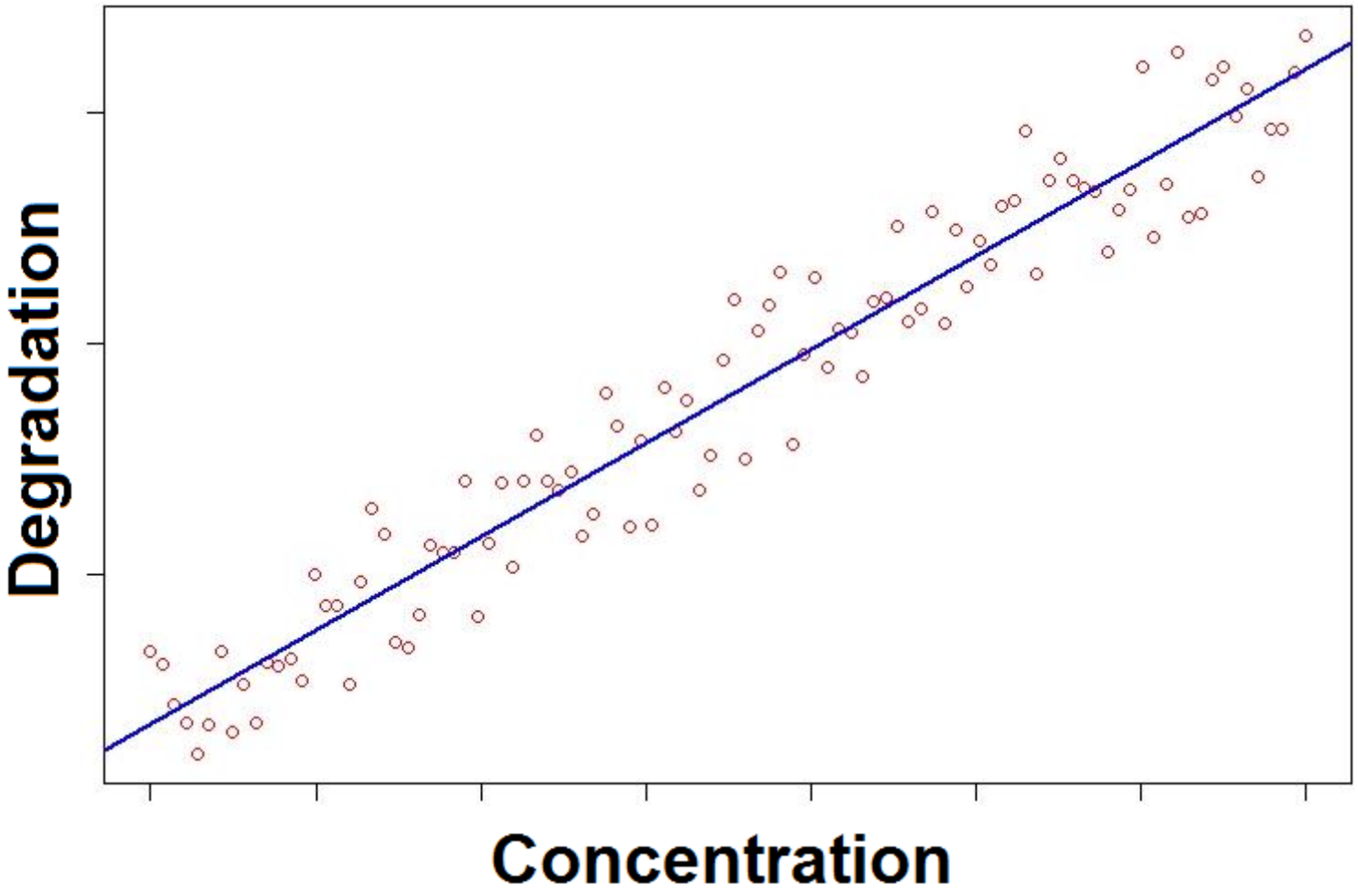
- A few hypothesis that cannot be proved.
- Rest of the science is developed by extending these using mathematics and experimentation.

# WELCOME TO THE INDUCTIVE MODELS

# Approach 2

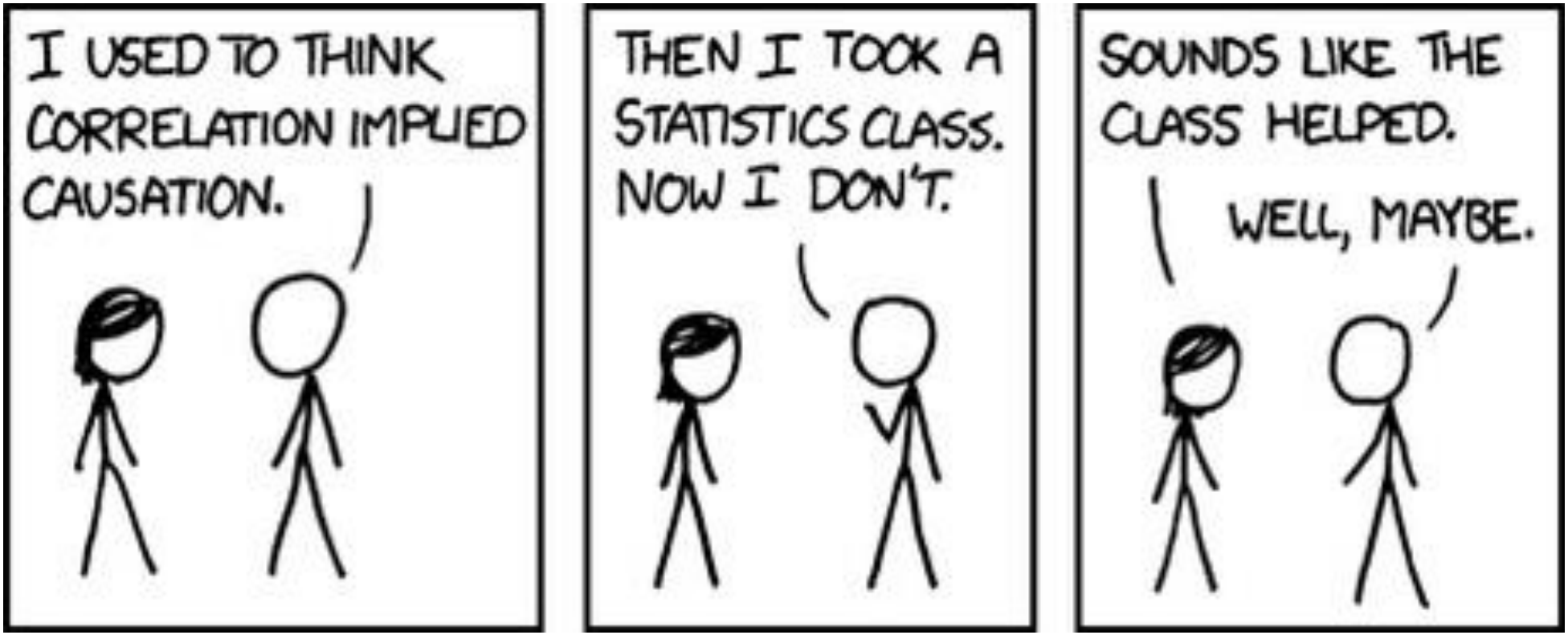
- Measure the degradations in hundreds of conditions (temperatures and concentrations for various times)


Degradation
Concentration


Degradation
Concentration

# Result is same but

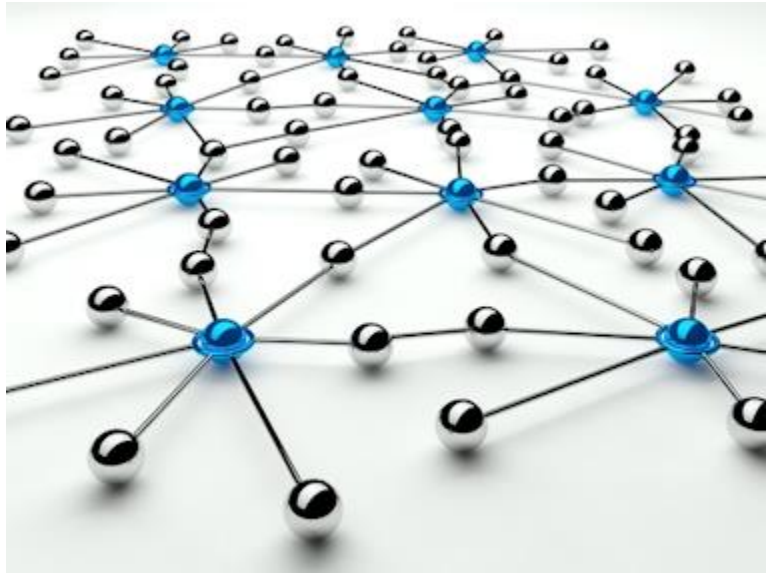
- Do not know the reasons (causations are not known and only correlations are identified)
  - Walmart example (does not tell how to act)


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

# Where it does not work

- Fails when randomness prevails
- Need a lot more data to come to the correct conclusions

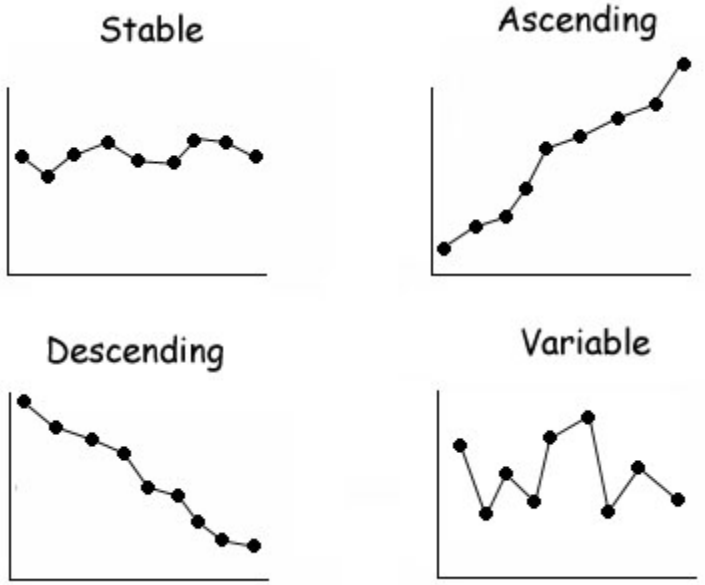
# Where does it work



# Where data is not an issue

# Data science

$$
\begin{aligned}
v_q &= -r_s i_q + \frac{\omega_r}{\omega_b}\Psi_d + \frac{p}{\omega_b}\Psi_q, \\
v_d &= -r_s i_d - \frac{\omega_r}{\omega_b}\Psi_q + \frac{p}{\omega_b}\Psi_d, \\
v_o &= -r_s i_o + \frac{p}{\omega_b}\Psi_o, \\
0 &= r_{aq} i_{aq} + \frac{p}{\omega_b}\Psi_{aq}, \\
v_f &= r_f i_f + \frac{p}{\omega_b}\Psi_f, \\
0 &= r_{ad} i_{ad} + \frac{p}{\omega_b}\Psi_{ad}, \\
T_e &= \frac{3}{2}\frac{P}{2}\frac{1}{\omega_b}(\Psi_d i_q - \Psi_q i_d), \\
p\omega_r &= \frac{P}{2J}(T_a - T_e),
\end{aligned}
\qquad
\begin{aligned}
p\theta_r &= \omega_r, \\
p\theta_e &= \omega_e, \\
\delta &= \theta_r - \theta_e, \\
\omega_m &= \frac{2}{p}\omega_r,
\end{aligned}
$$



# Currently

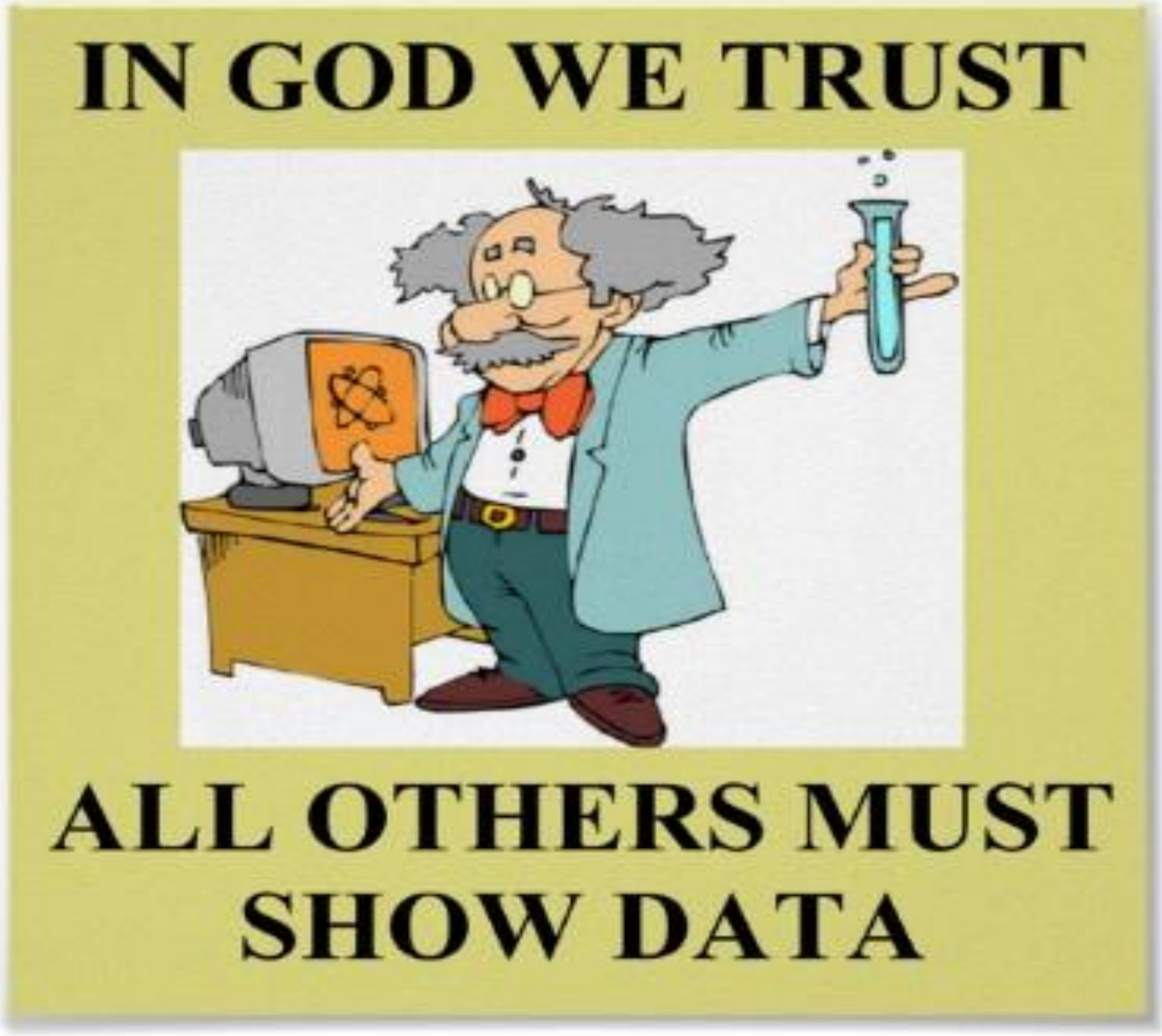
- A skill that employers can’t say no to
- Sexiest job of next decade

# IS INDUCTIVE LEARNING BETTER THAN INTUITION?

# Power of models

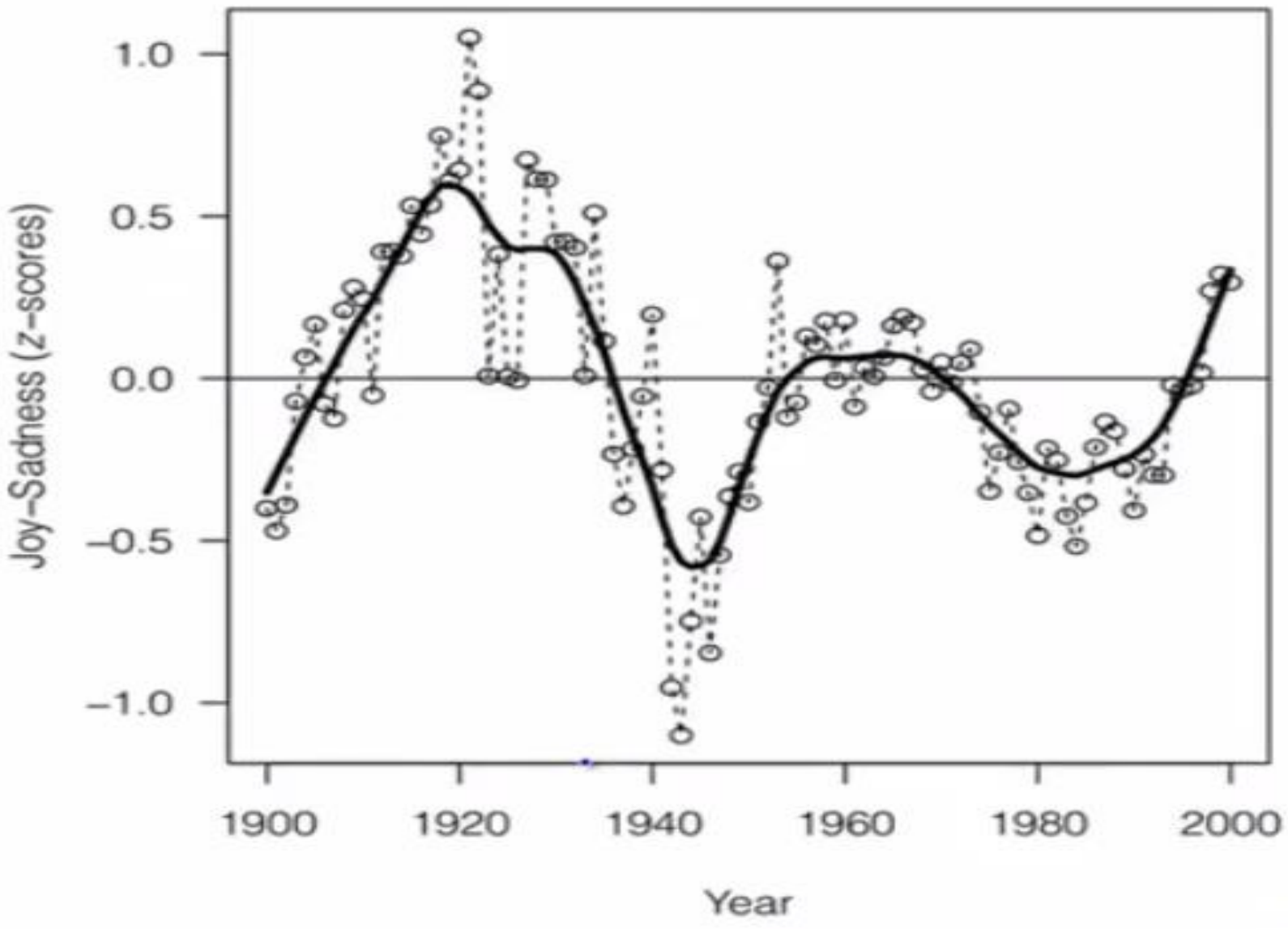
Robyn Dawes 1979: "The Robust Beauty of Improper Linear Models in Decision Making"

**Simple models do better than experts**

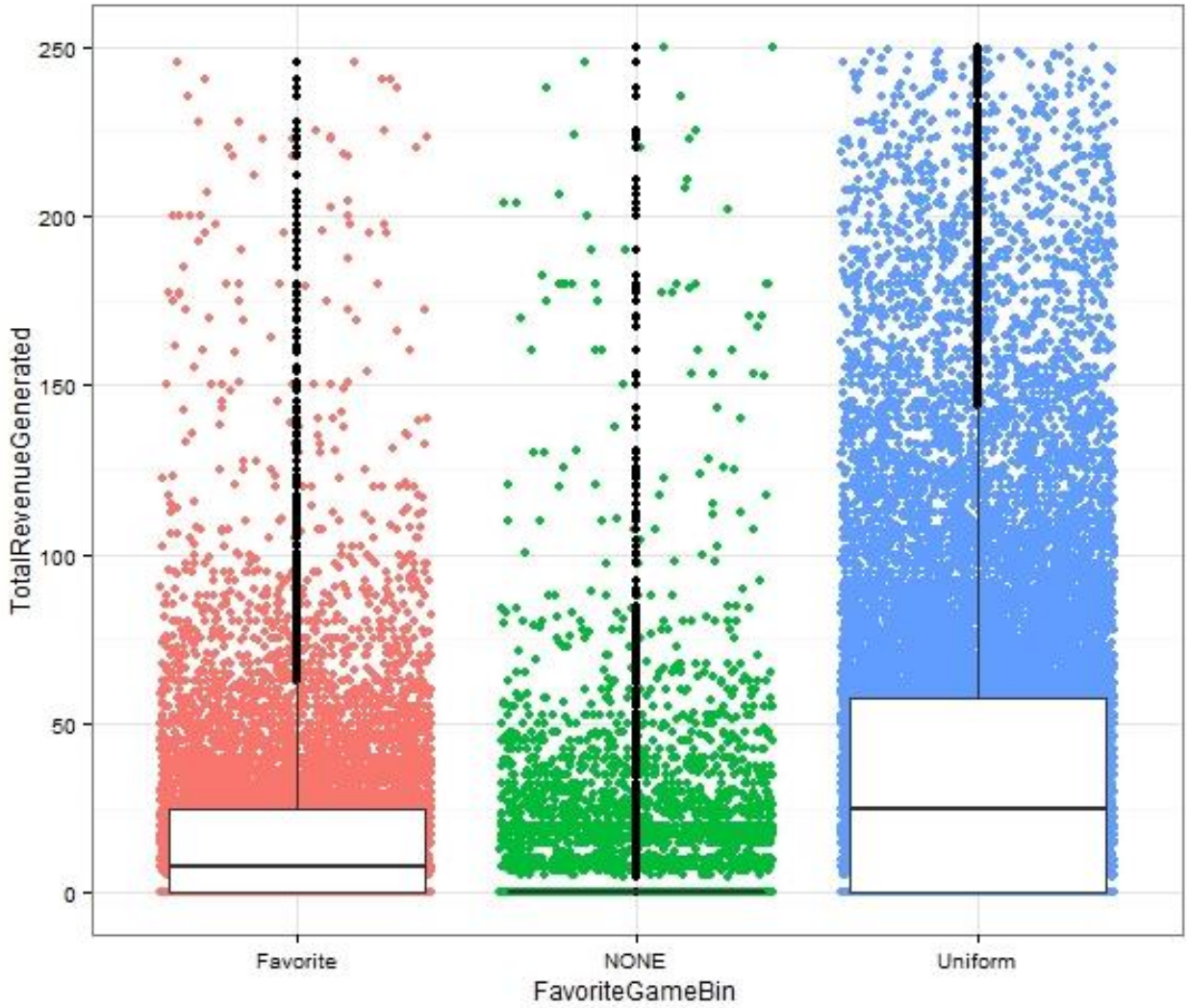

IN GOD WE TRUST
ALL OTHERS MUST
SHOW DATA

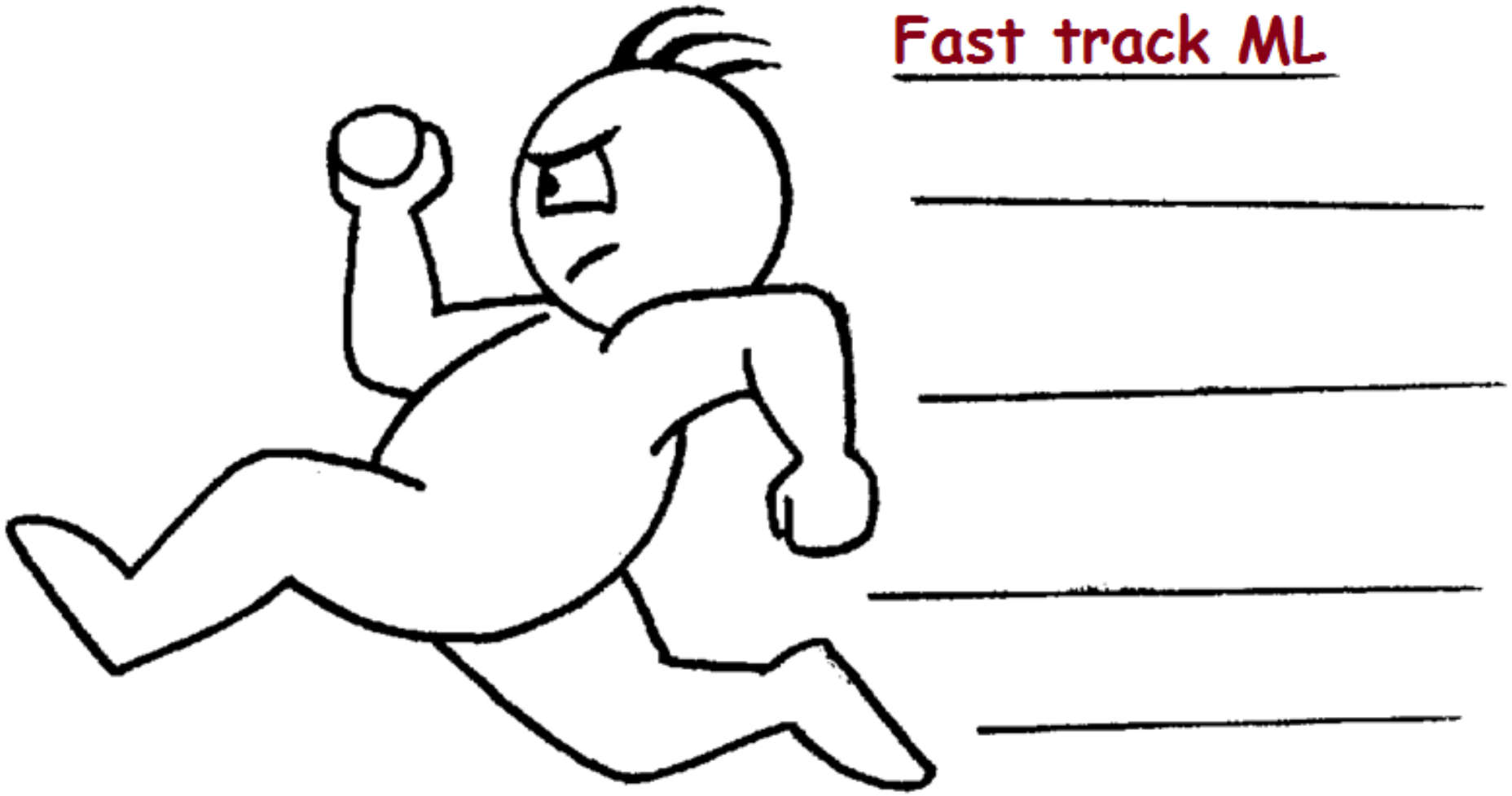
EXAMPLE

# Are Americans happy or unhappy

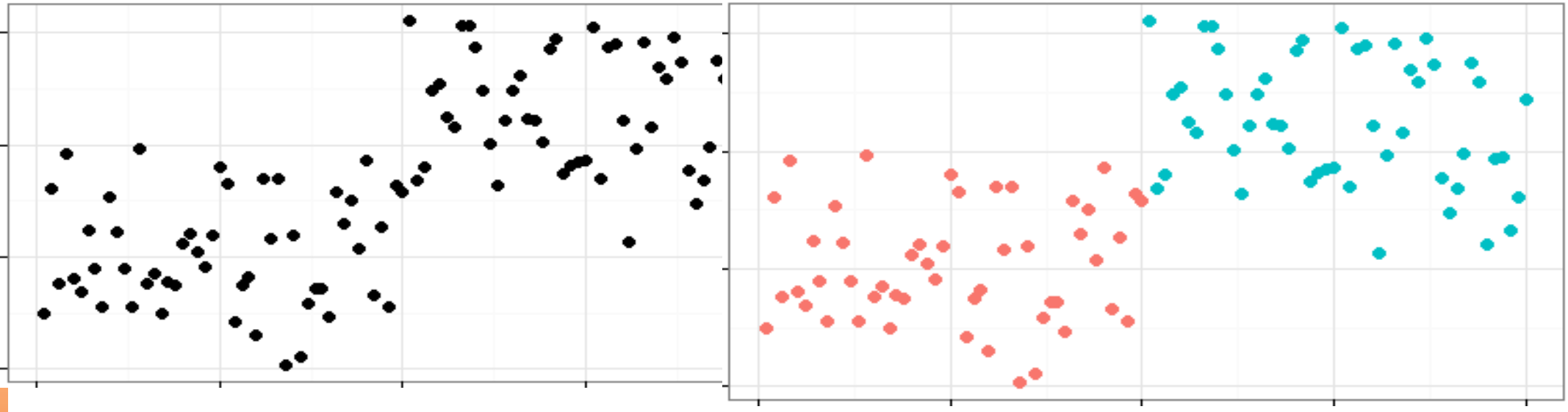


# Which customers are buying

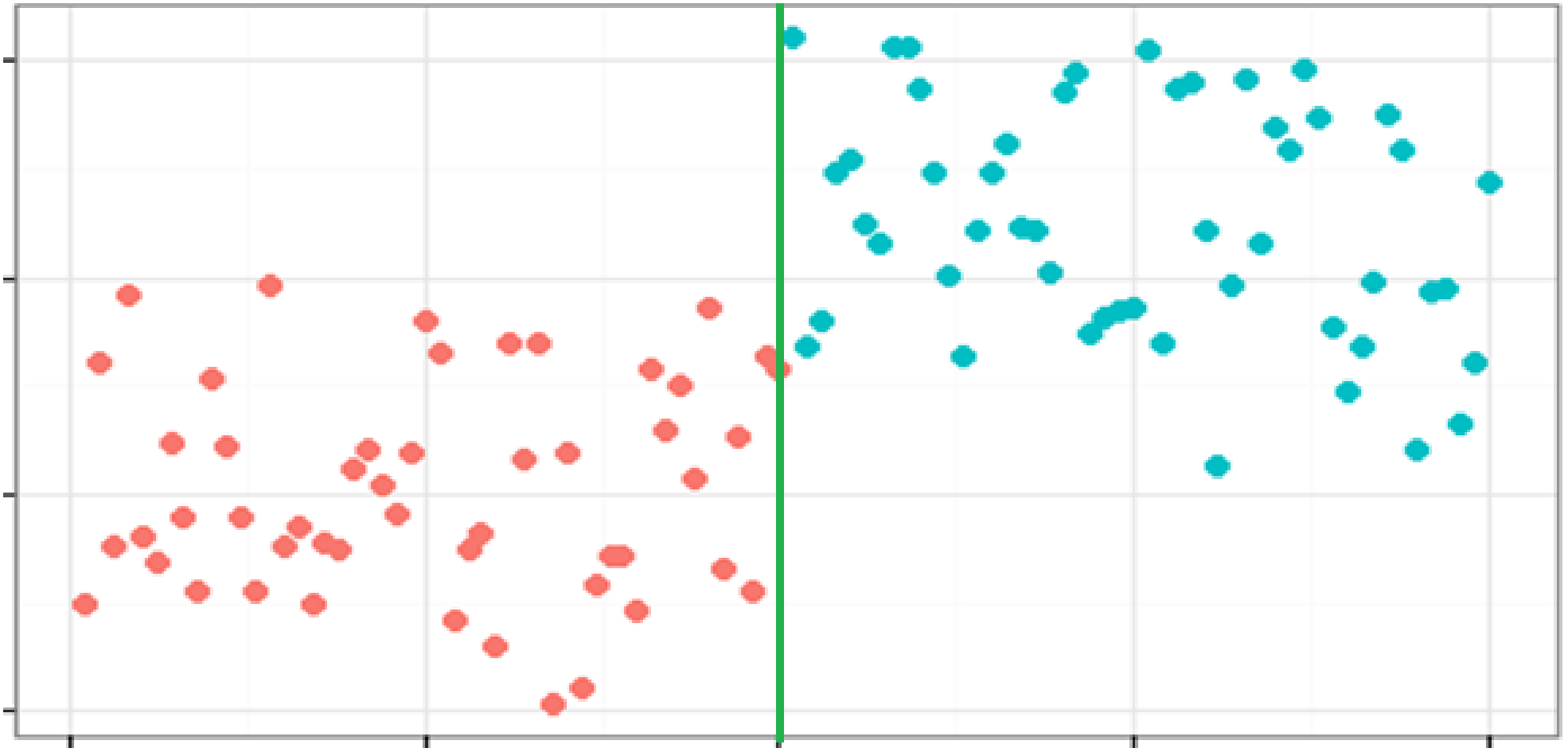



Fast track ML

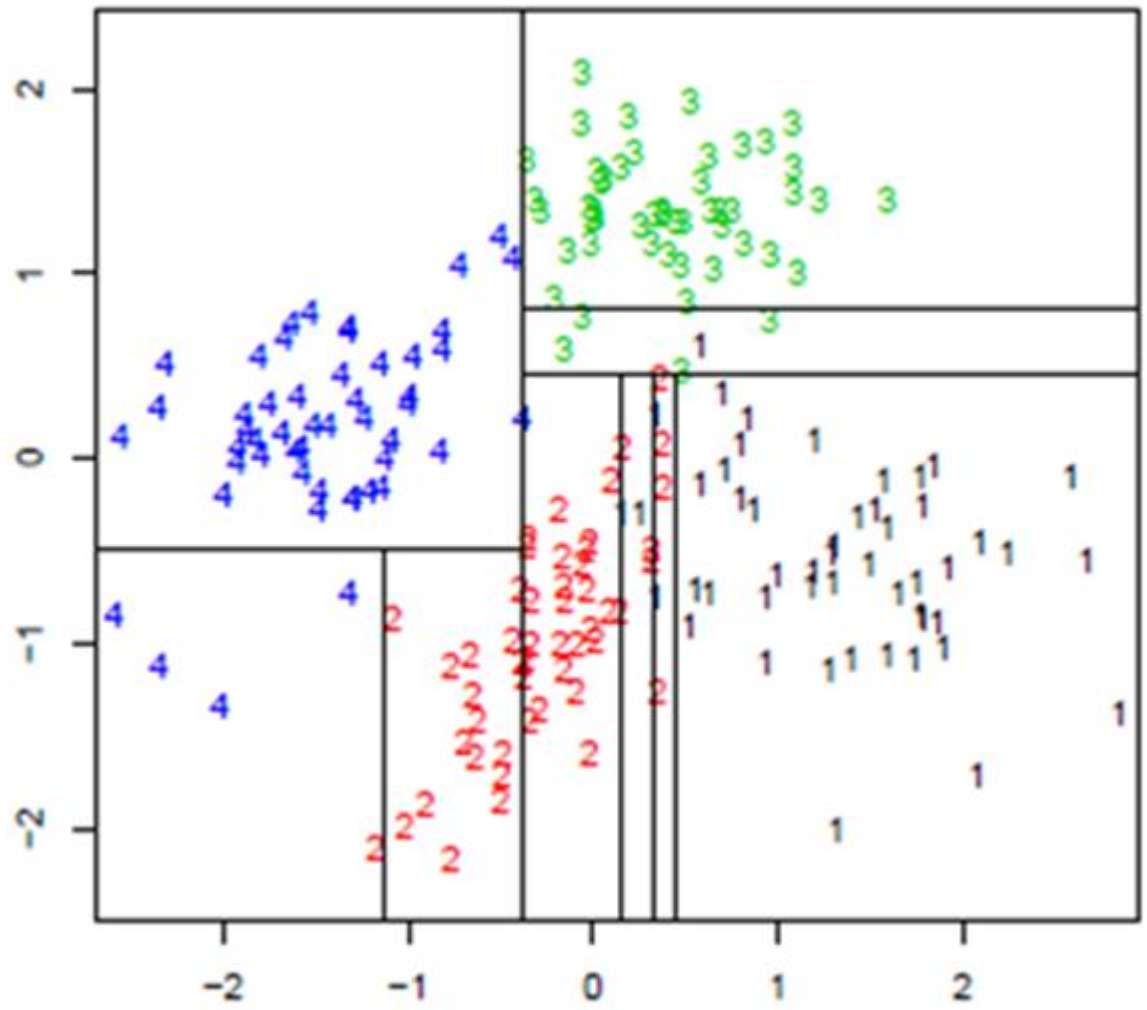
# Classification; Supervised and Unsupervised



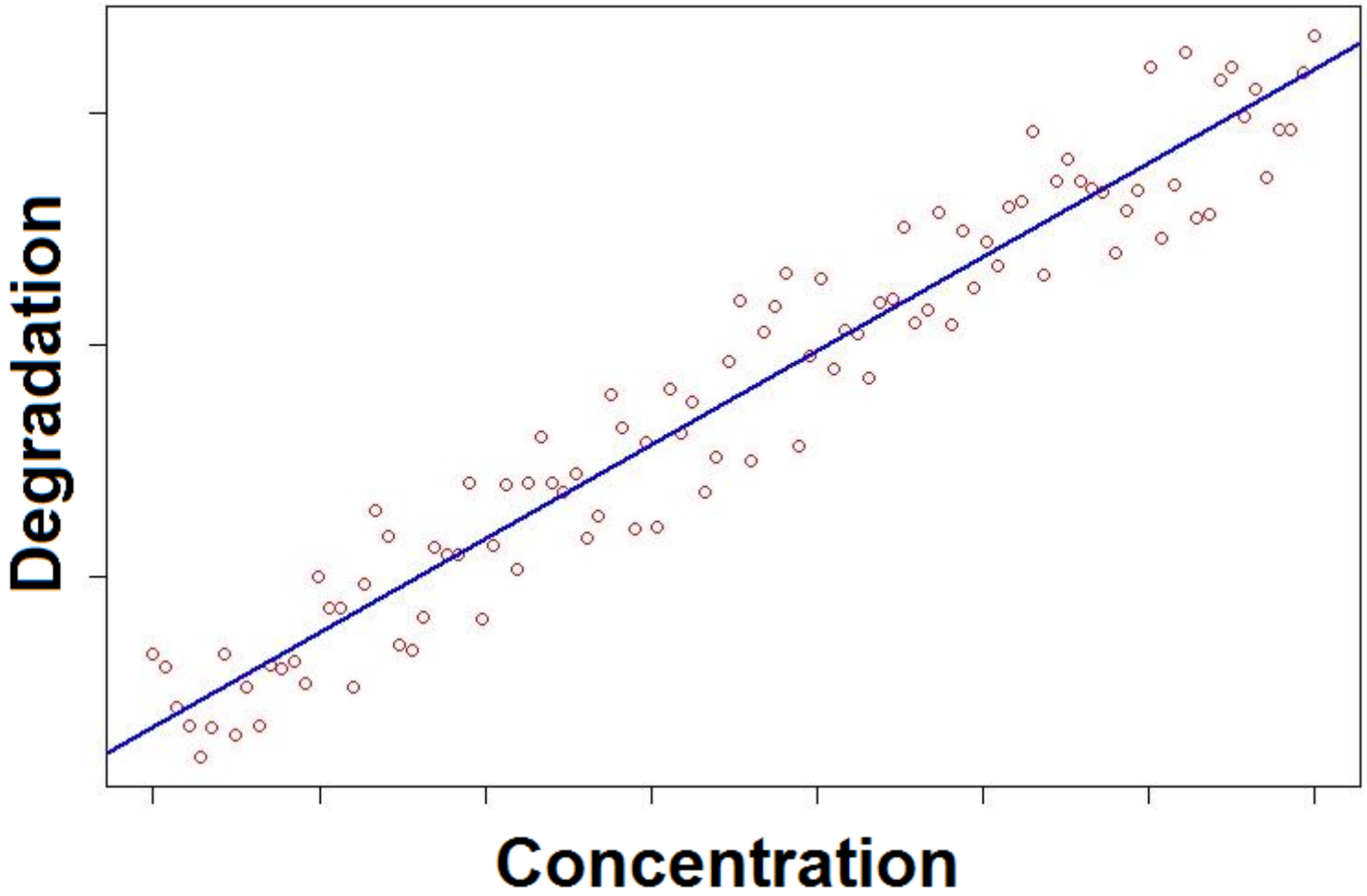
# Classification is partitioning the space



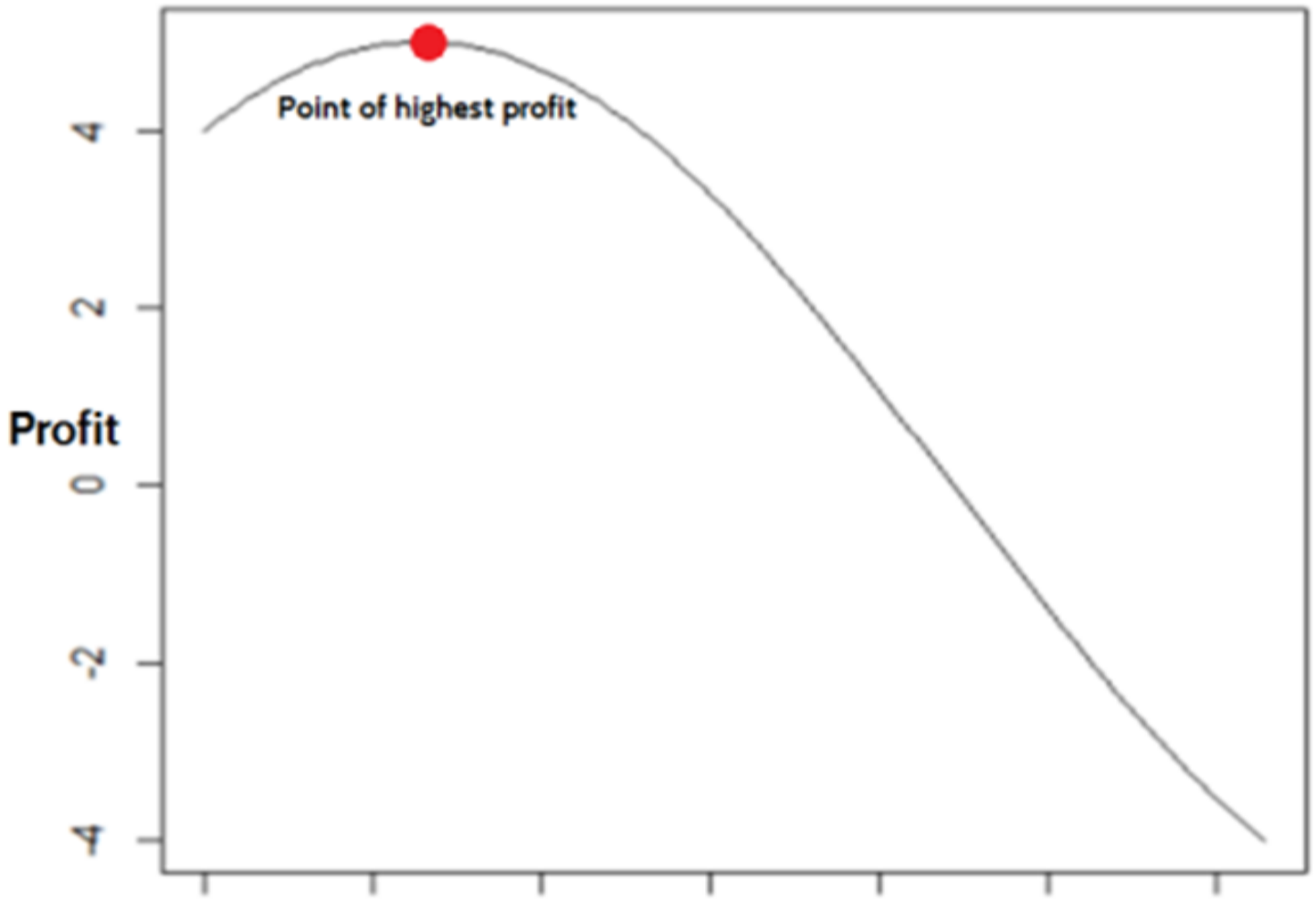
# Classification is partitioning…



# Forecasting is finding the line/plane closest to all points



# Optimization

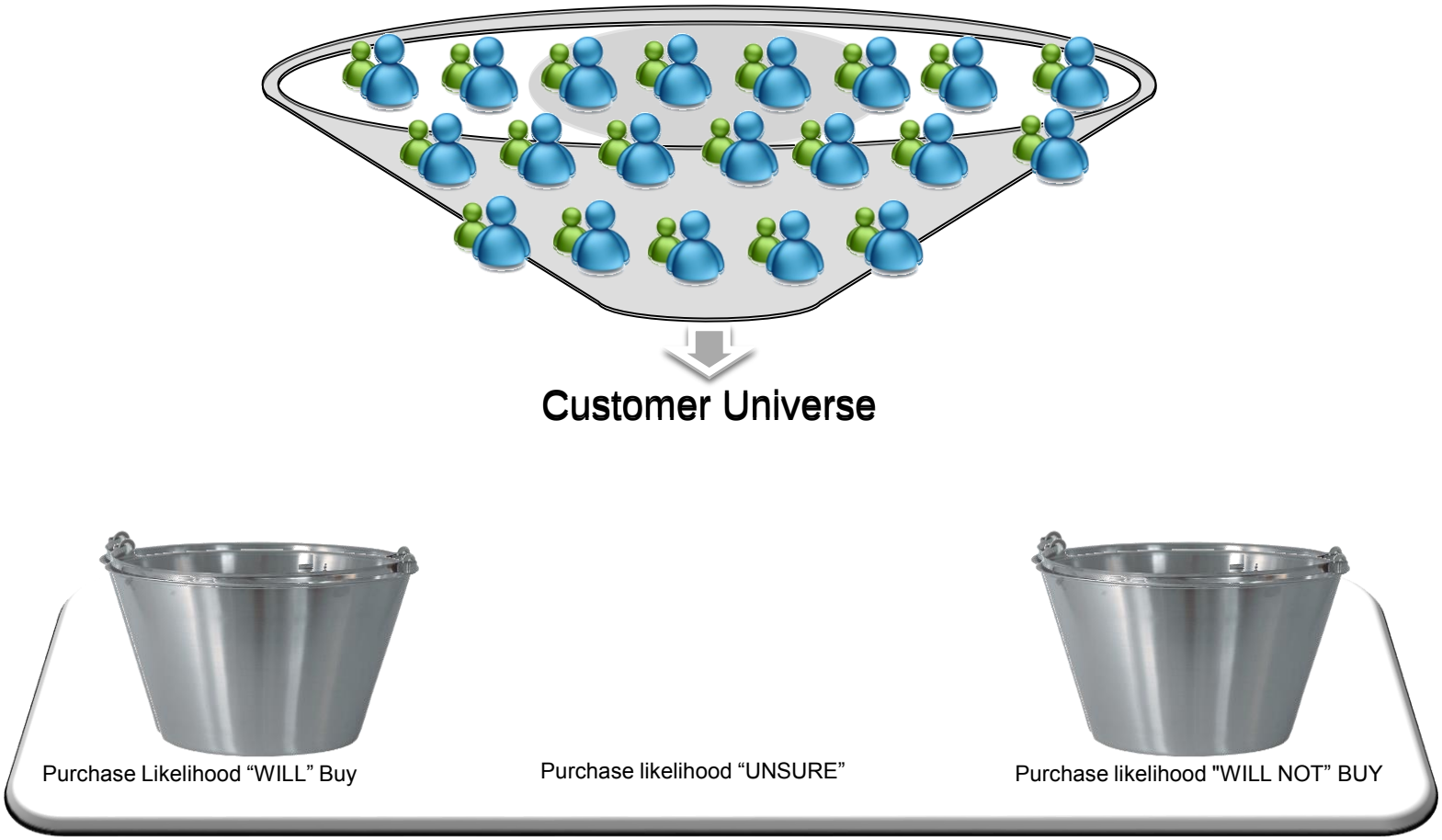


# Simulation

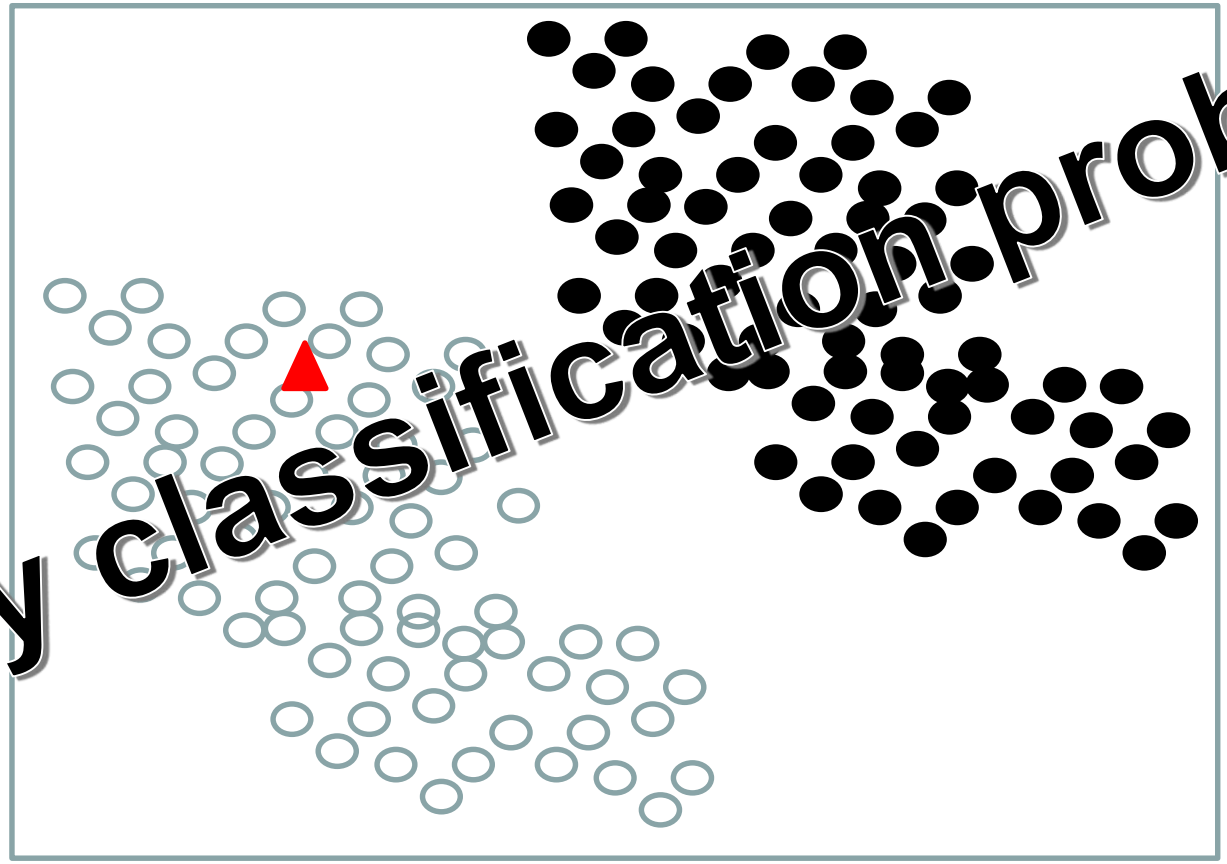
- Individuals are easy but groups and interactions are complex

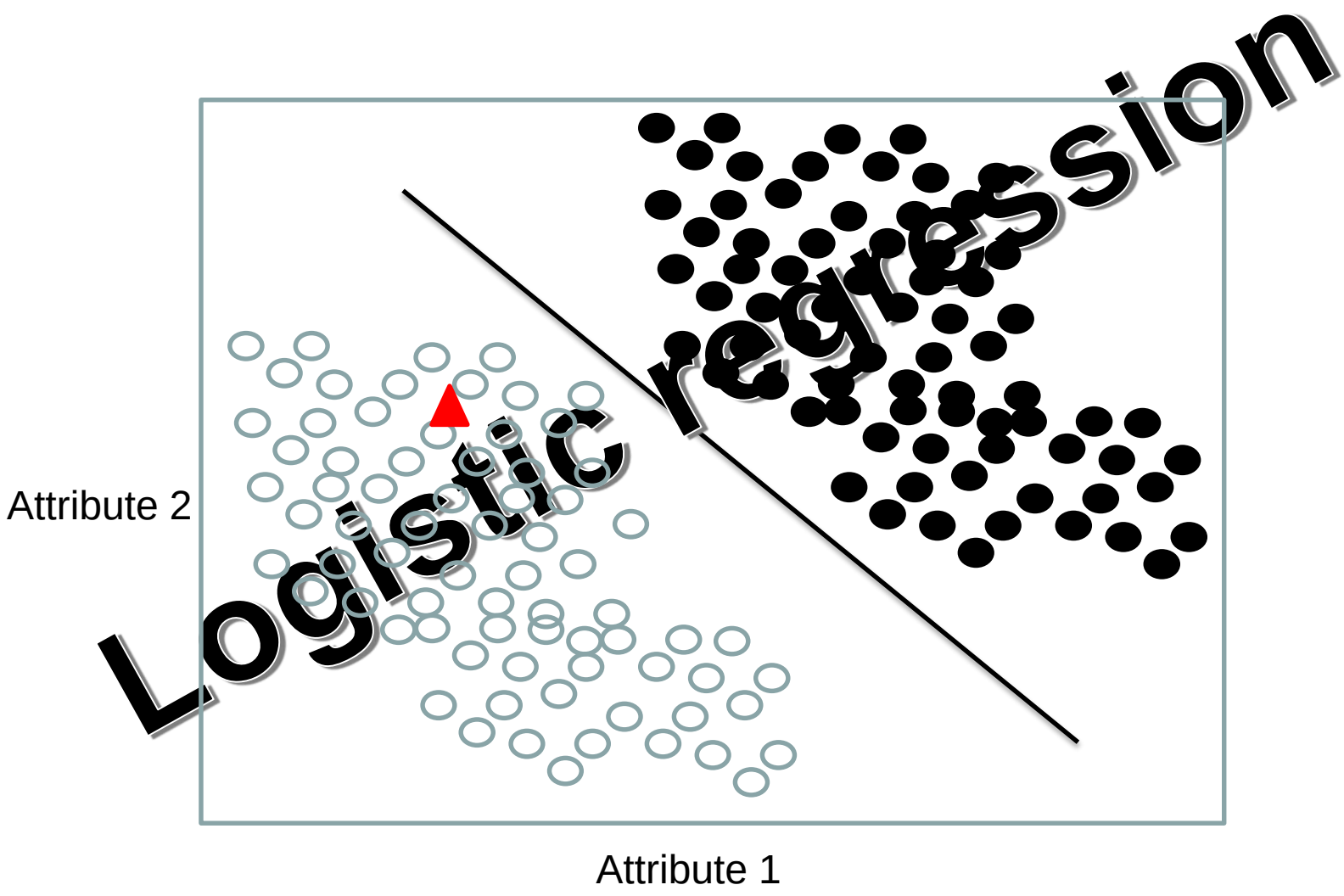
# FAST TRACK MODELS

# Data Science Models

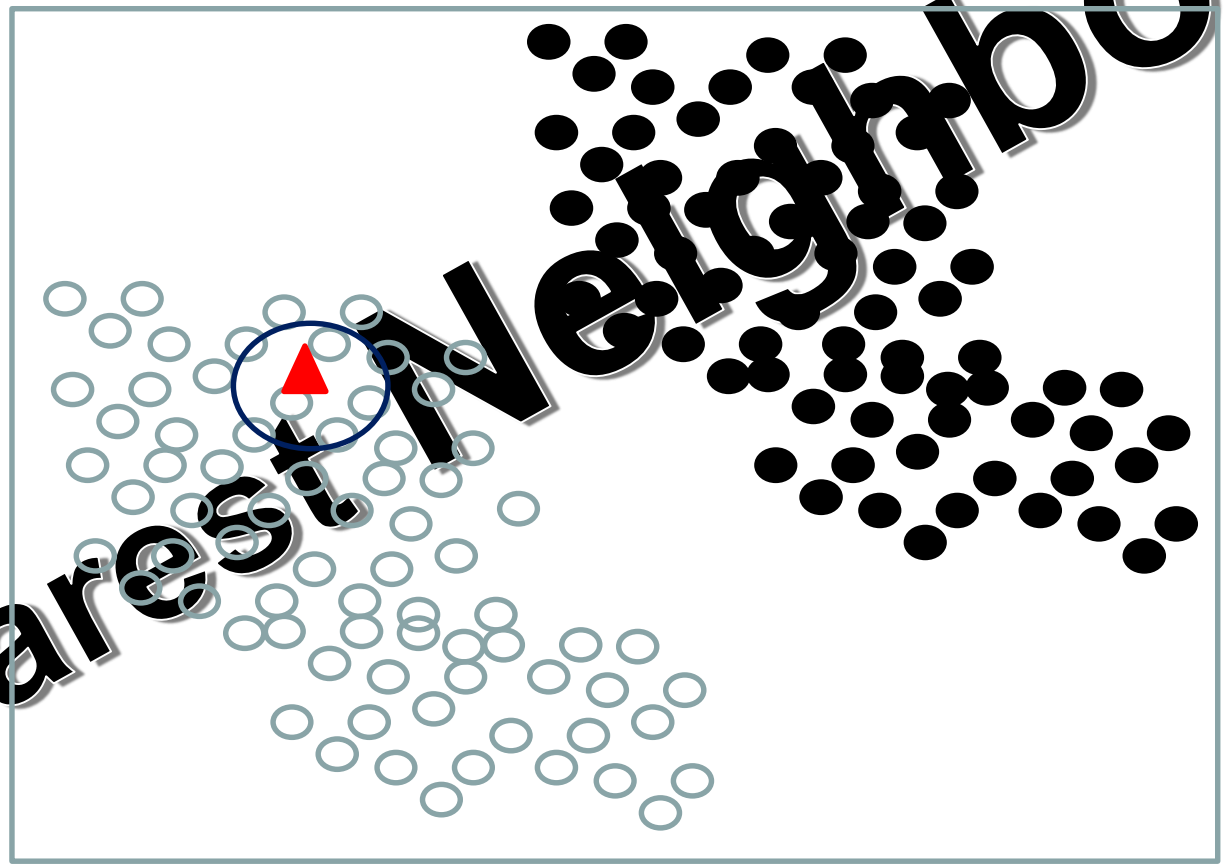


# Binary classification problem

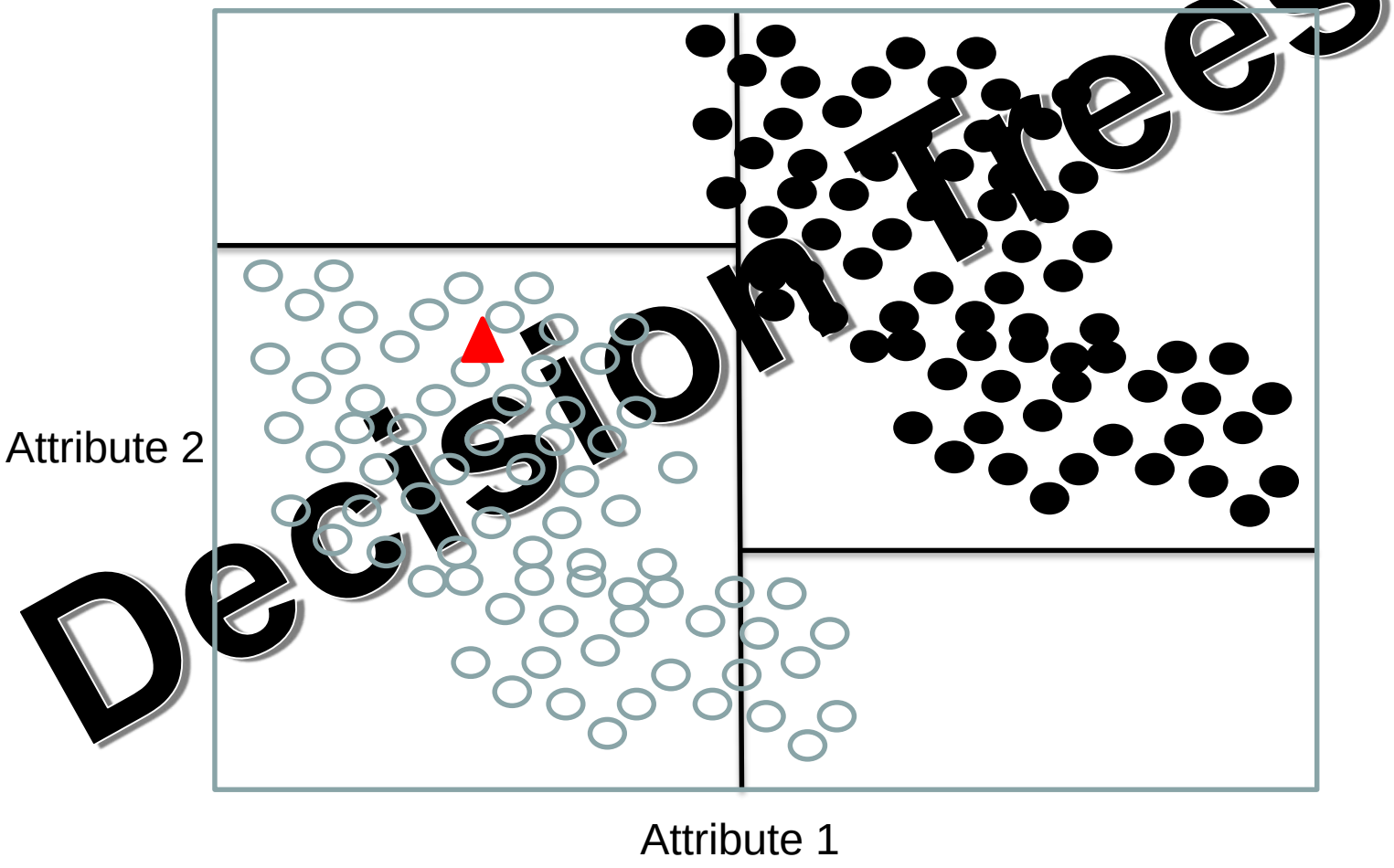



Logistic regression
Attribute 2
Attribute 1

# K-Nearest Neighbors



# Decision Trees



Non-linear solutions
Attribute 2
Attribute 1

Inspire...Educate...Transform.

# International School of Engineering


Plot 63/A, 1st Floor, Road # 13, Film Nagar, Jubilee Hills, Hyderabad - 500 033

For Individuals: +91-9502334561/63 or 040-65743991

For Corporates: +91-9618483483

Web: http://www.insofe.edu.in

Facebook: https://www.facebook.com/insofe

Twitter: https://twitter.com/Insofeedu

YouTube: http://www.youtube.com/InsofeVideos

SlideShare: http://www.slideshare.net/INSOFE

LinkedIn: http://www.linkedin.com/company/international-school-of-engineering



*This presentation may contain references to findings of various reports available in the public domain. INSOFE makes no representation as to their accuracy or that the organization subscribes to those findings.*