



# Recap of Predictive Modeling and Decision Trees

---

Dr. Goutam Chakraborty



# Outline

---

- A quick recap of basics of predictive modeling via machine learning approaches
- Three essentials of predictive modeling
  - How to predict new cases?
  - How to select useful input variables (when you have a large number of them)?
  - How to optimize complexity of models?



# Predictive Models: Statistical vs. Machine Learning Approach

## ■ Statistical Approach

- Explanation more important than prediction
- Smaller data size, limited number of variables
- Statistical significance *via p-value* is the dominant philosophy

## ■ Machine Learning approach

- Prediction more important than explanation
- Large data size, many variables
  - Need methods for variable selection
- Model *performance on unseen data* is the dominant philosophy
  - Handled via data partitioning



# Predictive Modeling Essentials

---

- ▶ **Predict new cases.**
- ▶ **Select useful inputs.**
- ▶ **Optimize complexity.**



# Predictive Modeling Essentials (Contd.)

- ▶ **Predict new cases.**
- ▶ Select useful inputs.
- ▶ Optimize complexity.

**Decision (nominal)  
and  
Estimate (interval).**

Sometimes we are interested in the ranking predictions (e.g., credit score)



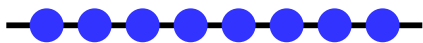
# Predictive Modeling Essentials (Contd.)

---

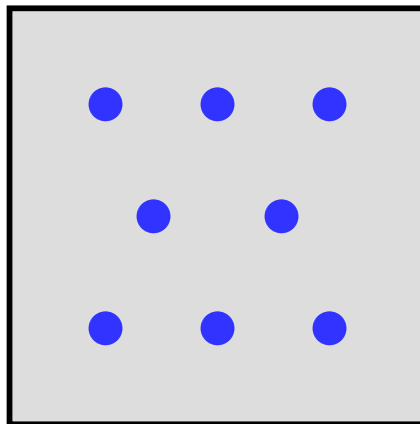
- ▶ Predict new cases.
- ▶ **Select useful inputs.**
- ▶ Optimize complexity.

**Eradicate  
redundancies  
and irrelevancies.**

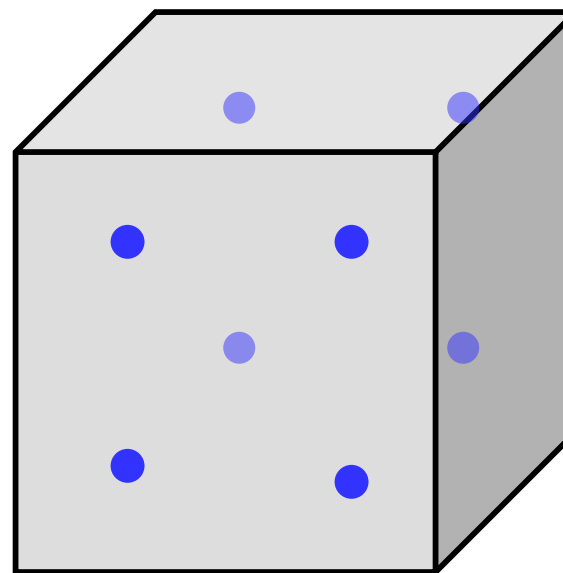
# The Curse of Dimensionality



1-D



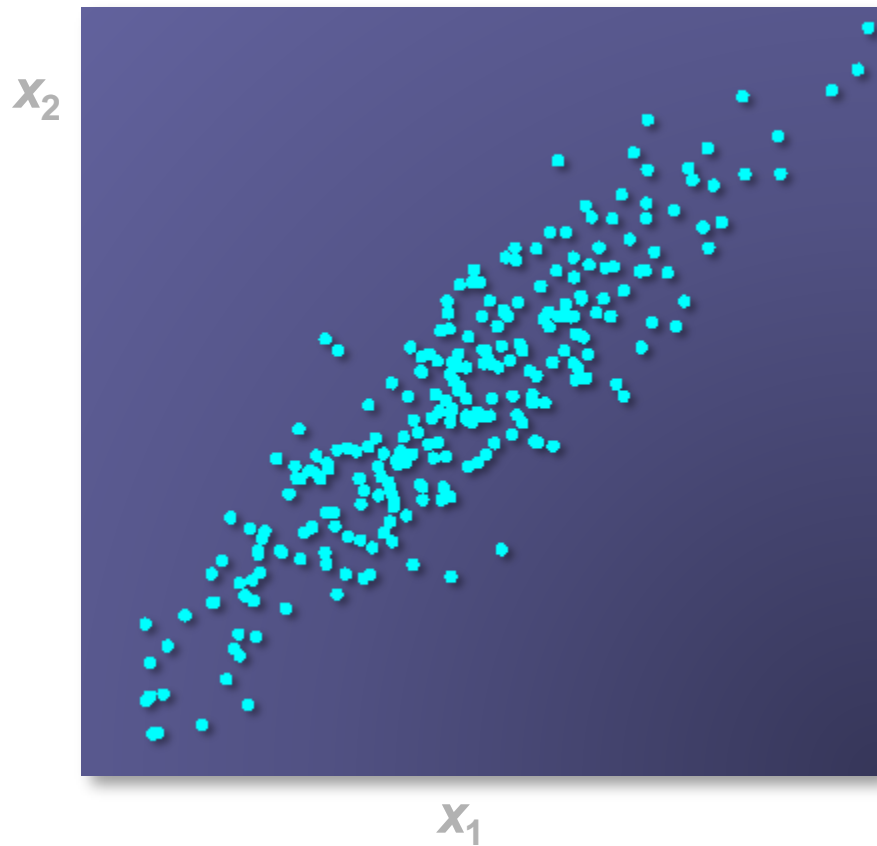
2-D



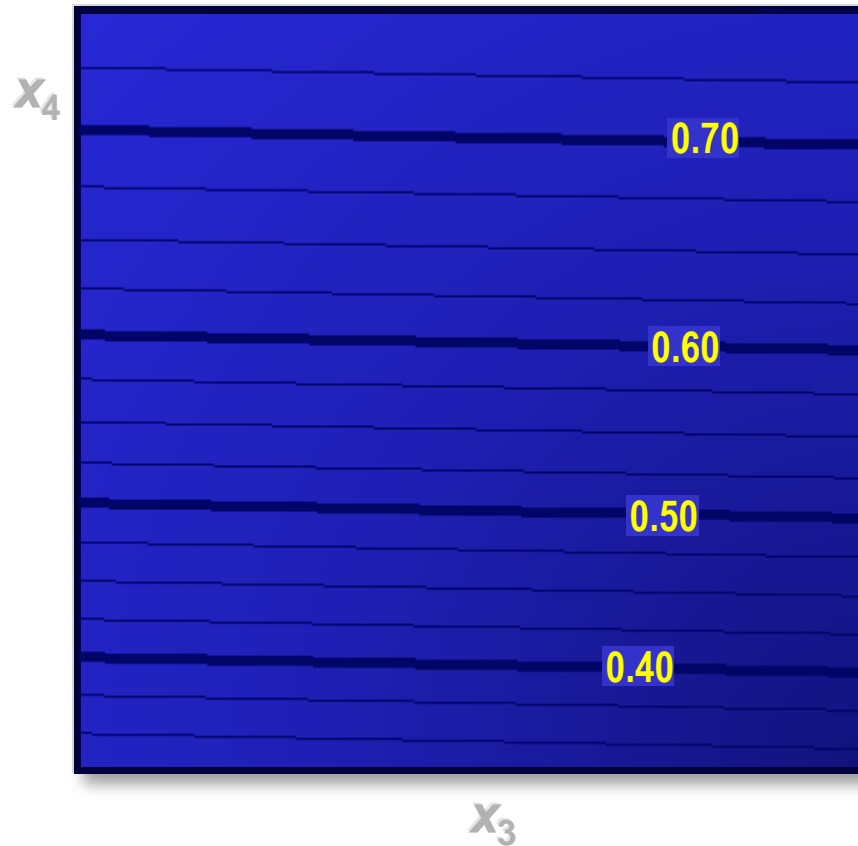
3-D

# Input Reduction – Redundancy

Redundancy



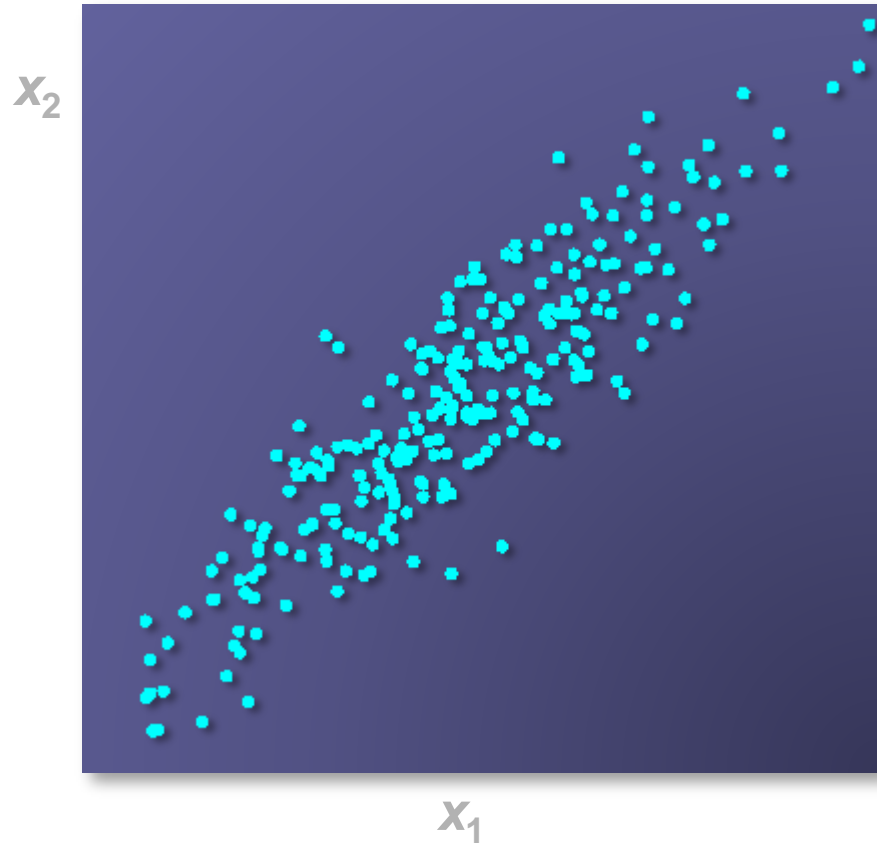
Irrelevancy



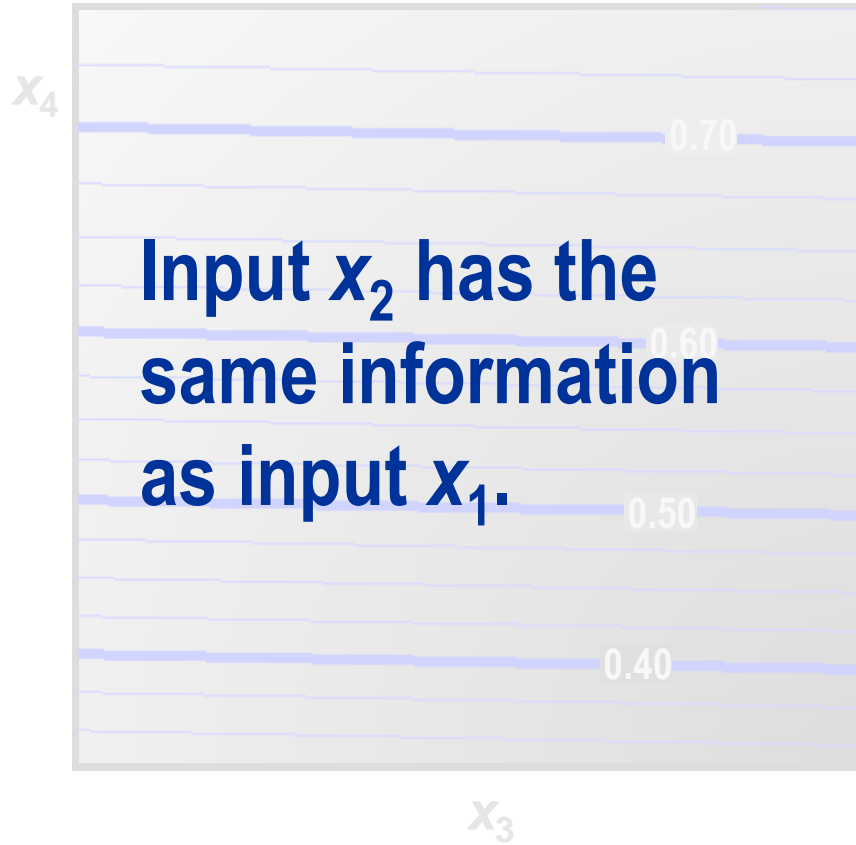


# Input Reduction – Redundancy

Redundancy



Irrelevancy



# Input Reduction – Irrelevancy

Redundancy

Irrelevancy

$x_2$

**Predictions change  
with input  $x_4$  but much  
less with input  $x_3$ .**

$x_1$

$x_4$

0.70

0.60

0.50

0.40

$x_3$

...

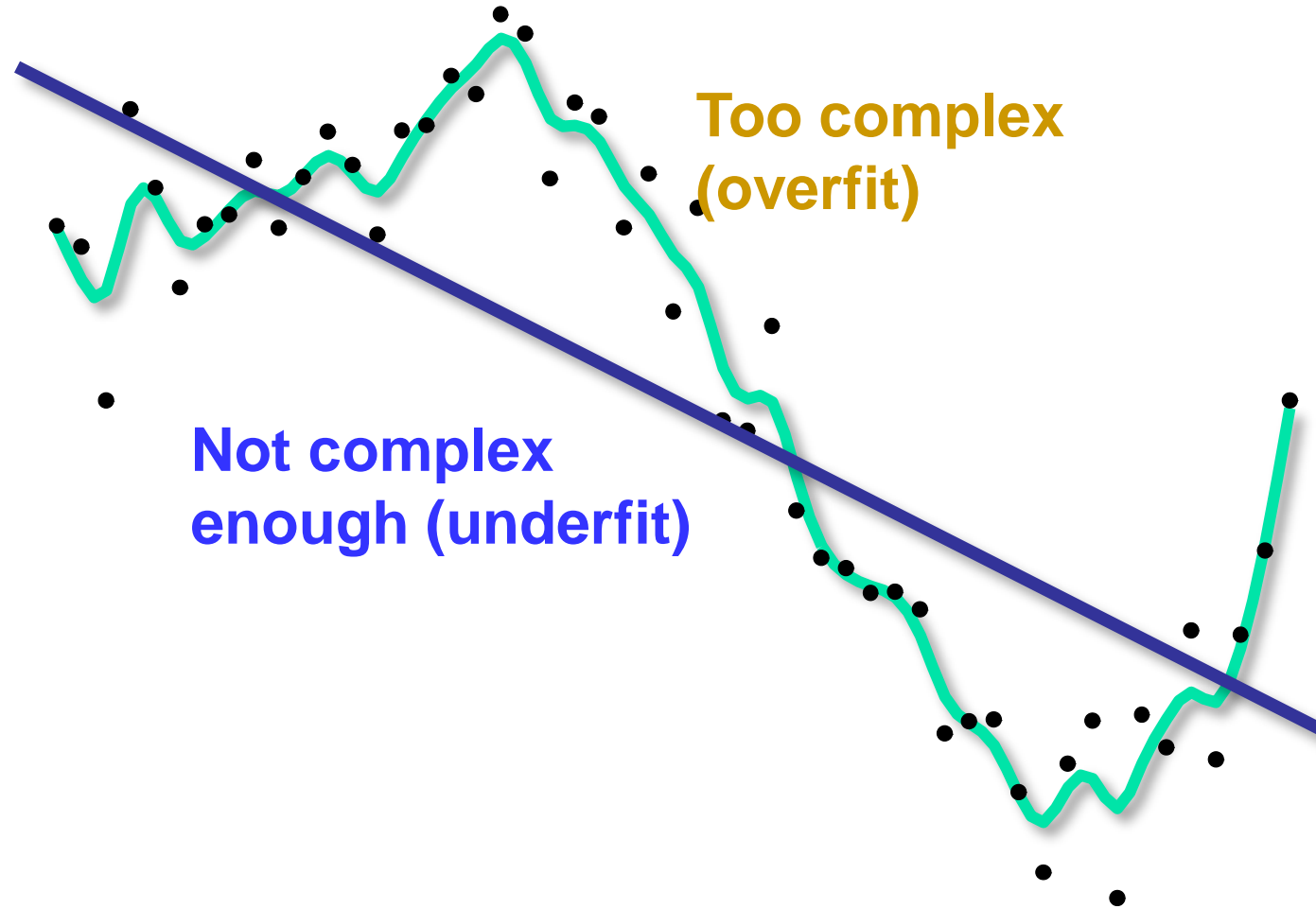


# Predictive Modeling Essentials (Contd.)

---

- ▶ Predict new cases.
- ▶ Select useful inputs.
- ▶ **Optimize complexity.**

# Model Complexity





# Modeling Essentials – Optimize Review

---

▶ Predict new cases.

Decide, rank,  
and estimate.

▶ Select useful inputs.

Eradicate  
redundancies  
and irrelevancies.

▶ **Optimize complexity.**

**Tune models with  
validation data.**