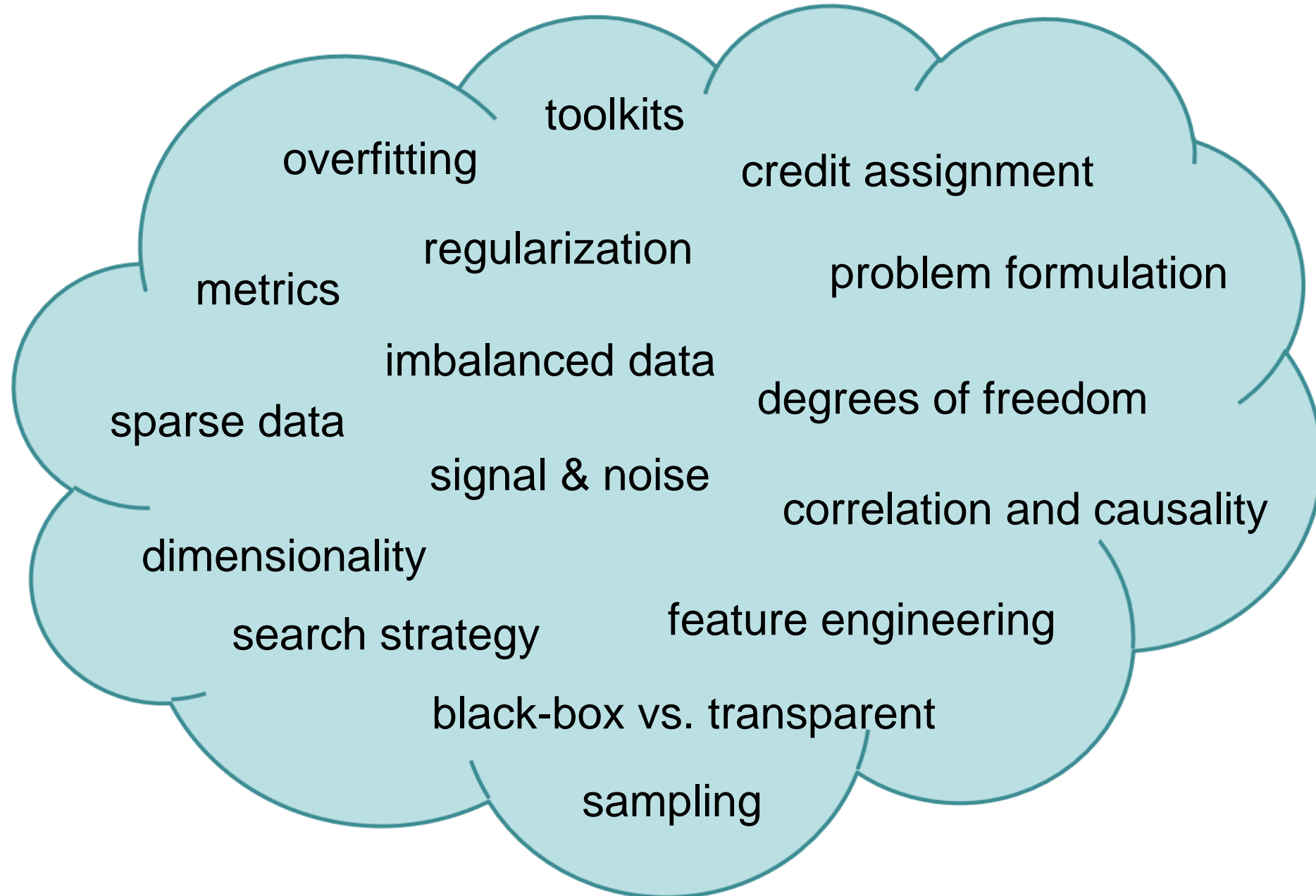


TieCON
AI/ML Bootcamp For Managers

How to Manage an AI Project?
AI Concepts are a Guide!

Eric Saund, Ph.D.
May, 2019

www.saund.org



Analogy Map

AI / ML Concept

Technical Meaning

Analogy in Project Management

Problem Formulation

- grokking data
- requirements
- metrics
- data preparation
- models

-
-
-
-
-

-
-
-
-
-

Dimensionality

- data
- model
- search
- sampling
- overfitting

-
-
-
-
-

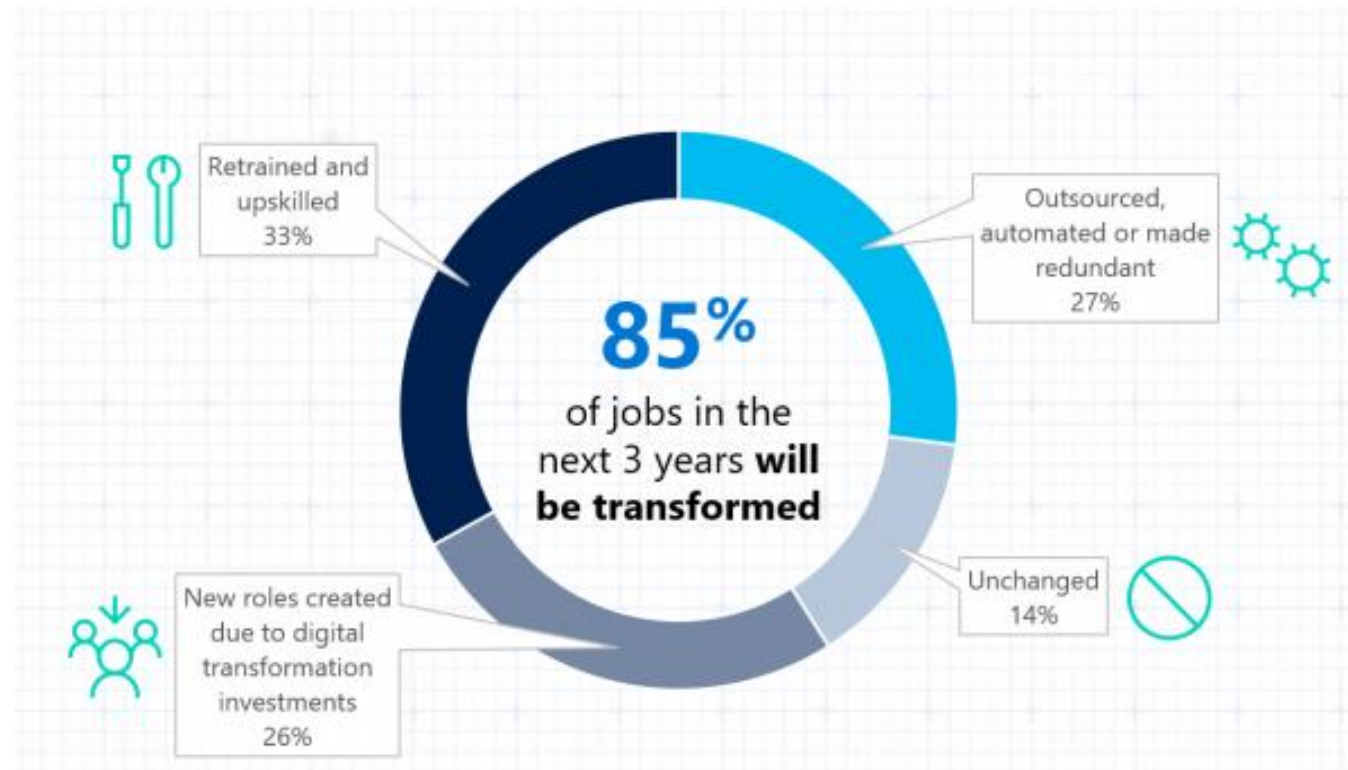
-
-
-
-
-

“AI is Coming for Your Job”

Enterprise
INNOVATION April 25, 2019

Brave new world – how AI will impact work and jobs

By Ralph Haupter, President, Microsoft Asia | 2018-05-01



<https://www.enterpriseinnovation.net/article/brave-new-world-how-ai-will-impact-work-and-jobs-730637861>

This “Robot” Cannot Do Your Job

Home › Entertainment › 10 facts you need to know about hu...

10 facts you need to know about humanoid robot Sophia



<https://www.yovizag.com/10-facts-humanoid-robot-sophia/>

AI Today: Conversational Agents



Why Are Conversational Agents So Dumb?

“Alexa, who won the 1934 world series?”

“The Saint Louis Cardinals beat the Detroit Tigers 4-3 in the 1934 World Series.”

“Who was the starting pitcher for the Tigers?”

“Jordan Zimmerman started for the Tigers in their 6-4 loss to the Twins yesterday.”

The Ingredients of Intelligence

Knowledge

- *data mining*

- ***predictive analytics***

Pattern Matching

- *computer vision*
- *speech recognition*

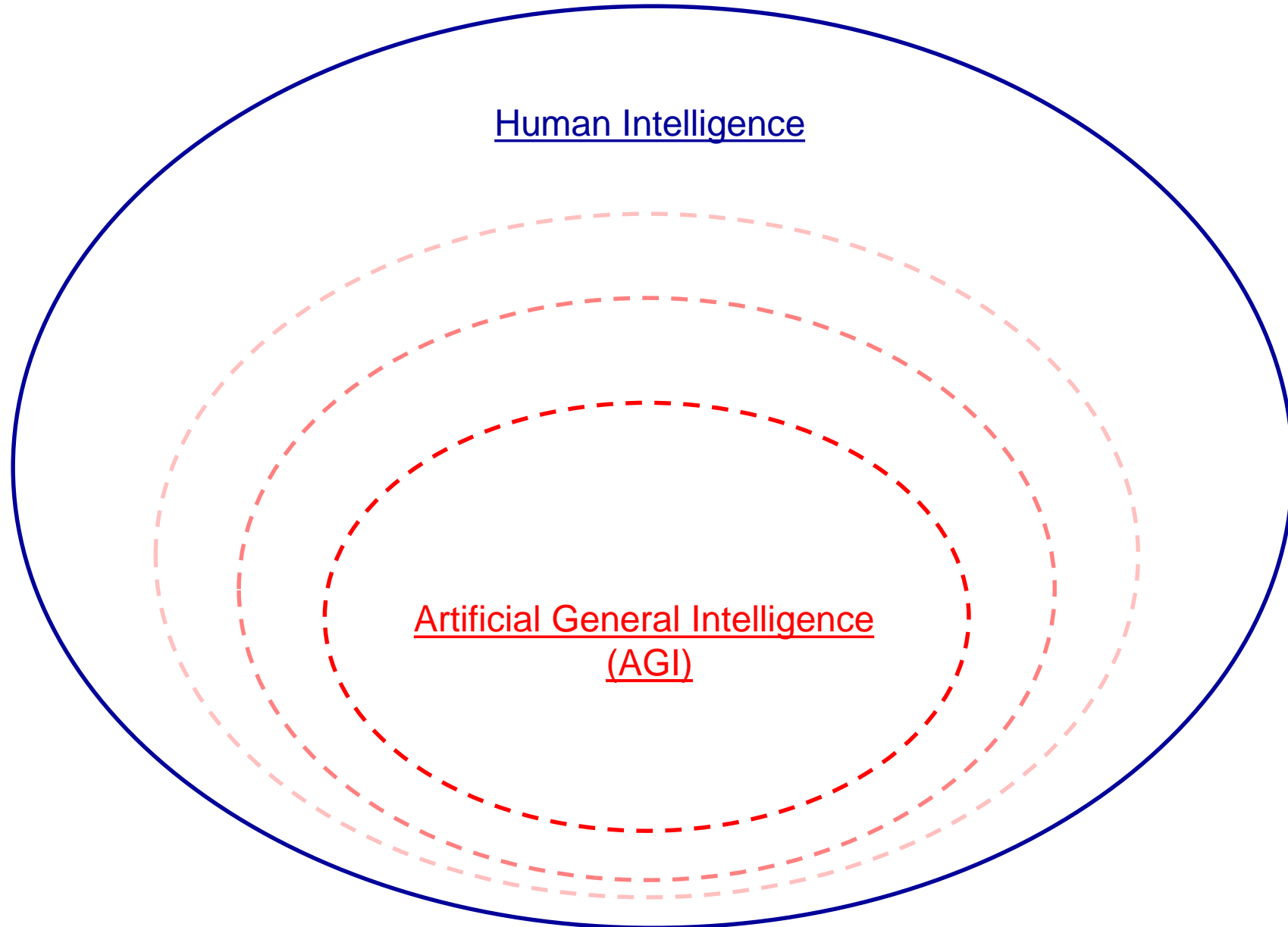
- *Natural Language Understanding*

Modern Machine Learning

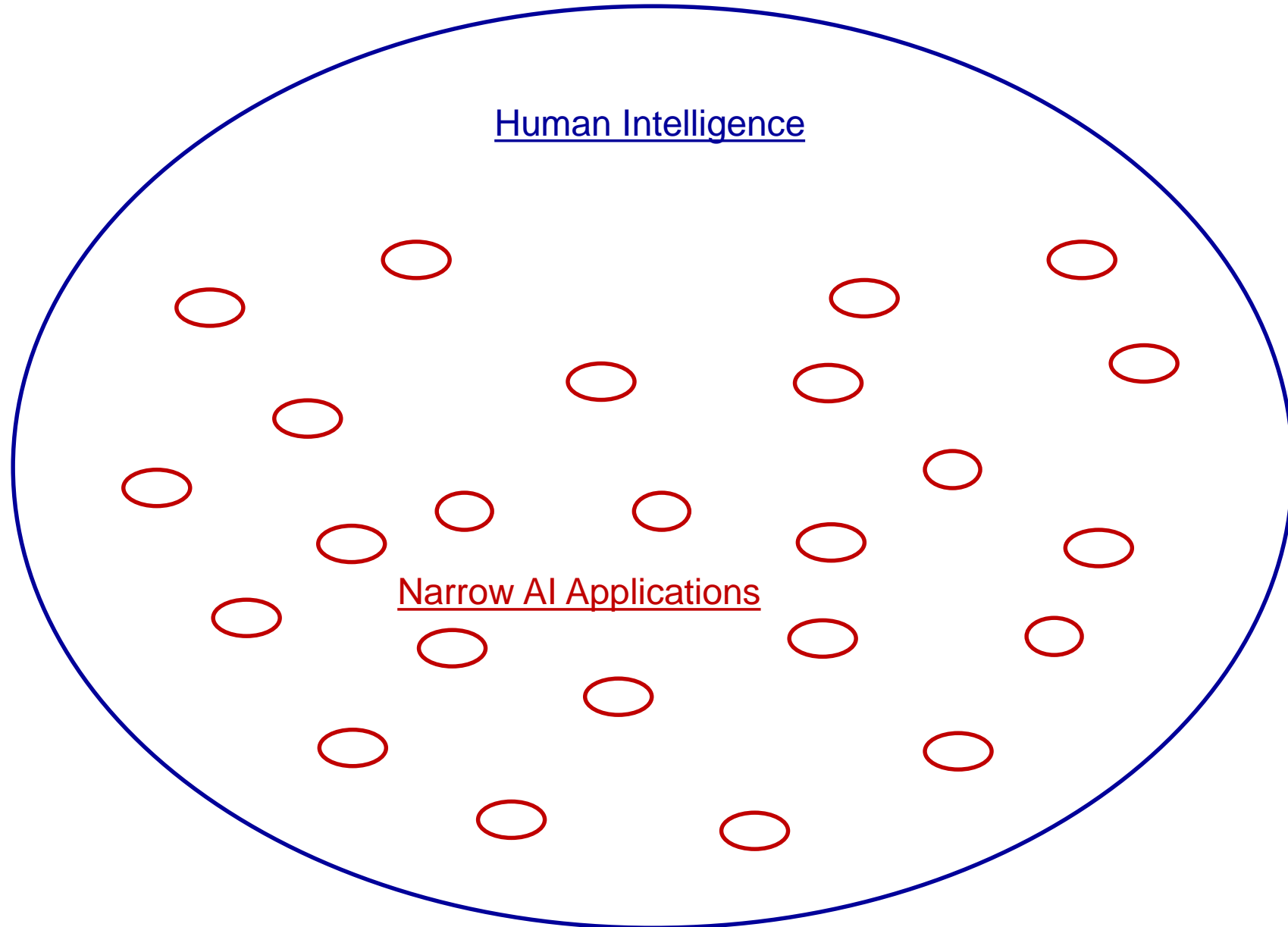
Reasoning

- *constrained optimization*

Human Intelligence and Artificial General Intelligence



Human Intelligence and Narrow Artificial Intelligence



Analogy Map

AI / ML Concept

Technical Meaning

Analogy in Project Management

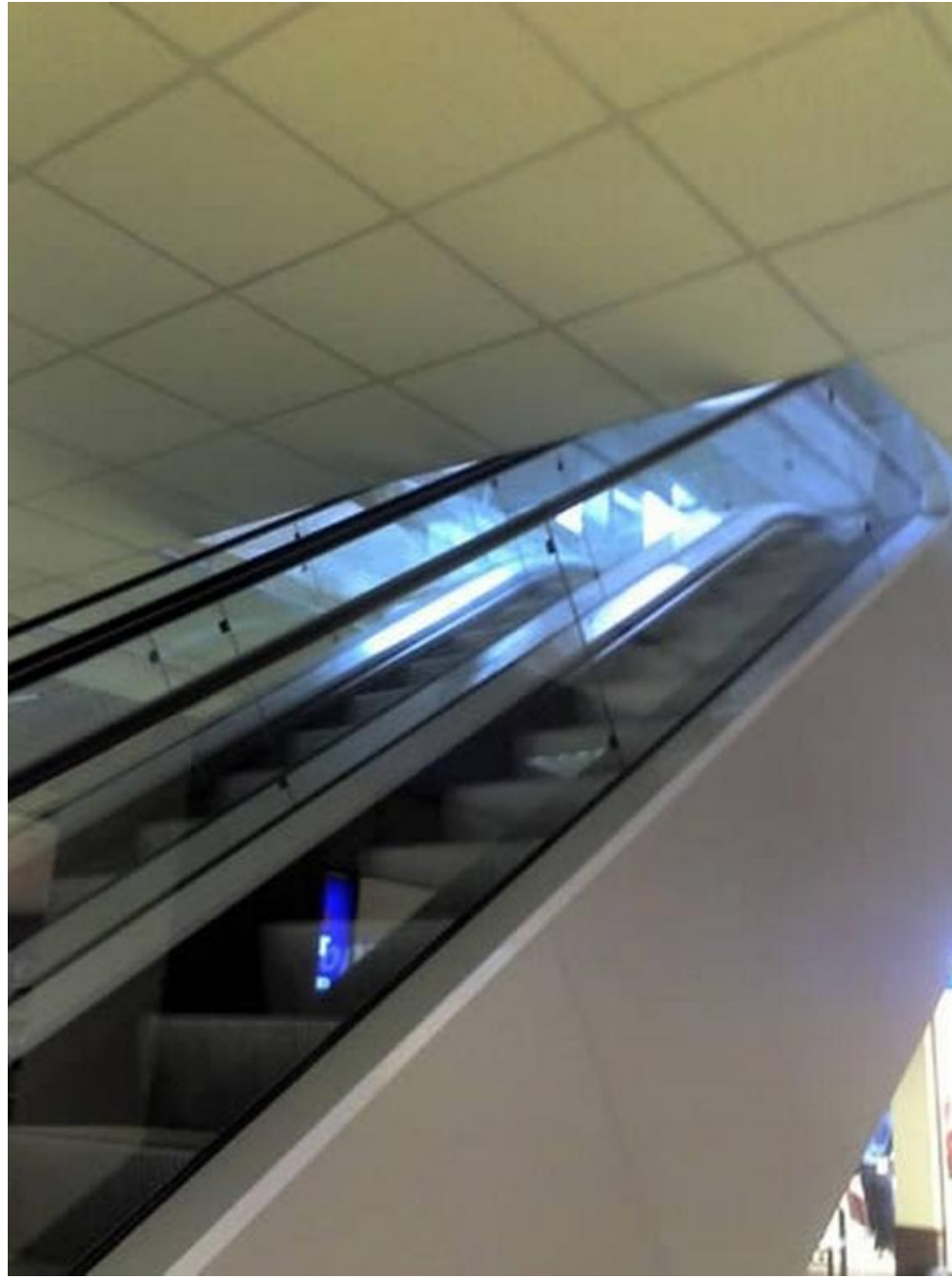
Problem Formulation

- grokking data
- requirements
- metrics
- data preparation
- models

Dimensionality

- data
- model
- search
- sampling
- overfitting

Problem Formulation

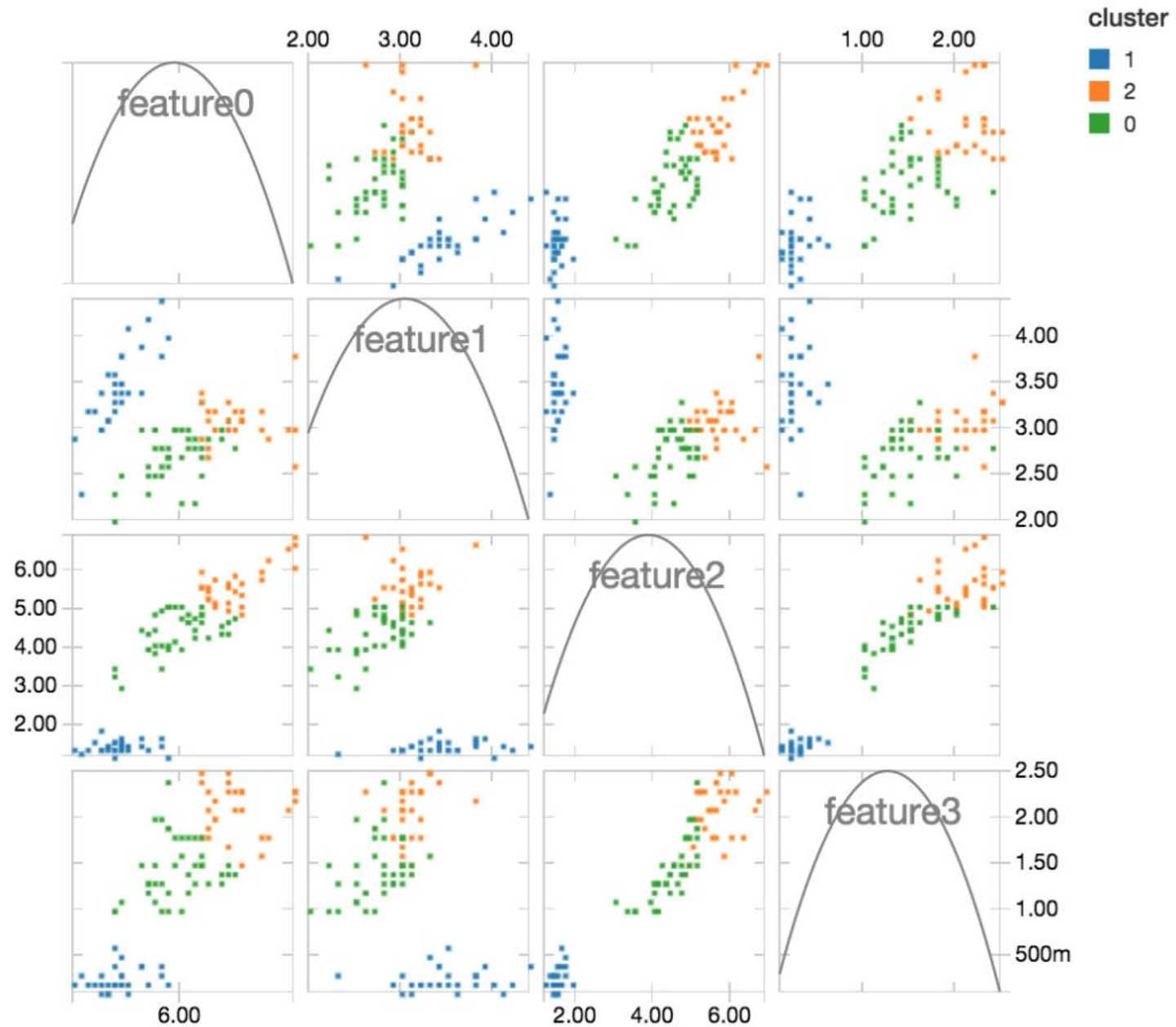


Problem Formulation



<https://widspune.blogspot.com/2019/02/data-science-problem-formulation.html>

Problem Formulation: 1. Understand the Data



Problem Formulation: 2. Requirements



How the customer explained it



How the project leader understood it



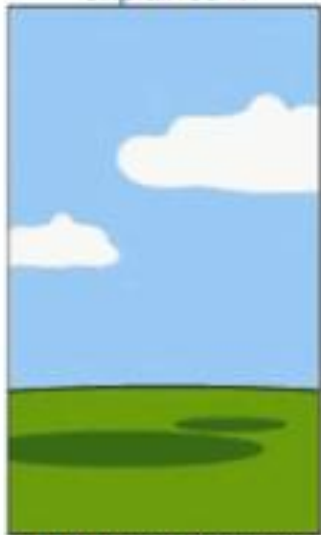
How the engineer designed it



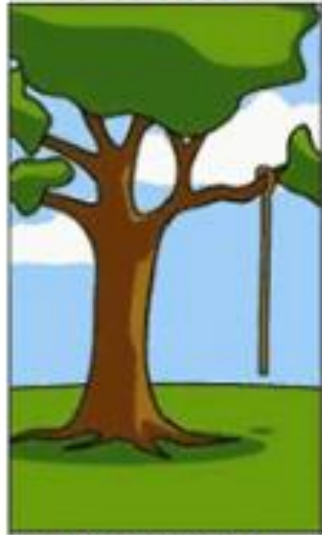
How the programmer wrote it



How the sales executive described it



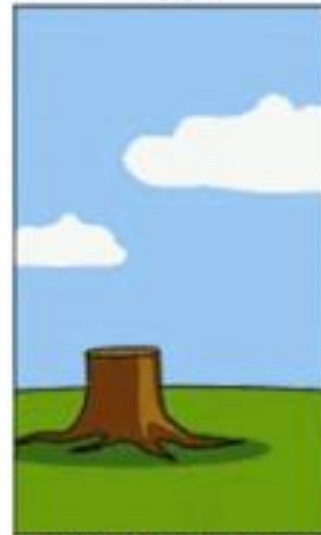
How the project was documented



What operations installed



How the customer was billed



How the helpdesk supported it



What the customer really needed

Nature of the task:

- prediction,
- forecasting
- recommendation
- exploration & discovery
- question answering

Requirements for a solution:

- speed
- accuracy
- consistency
- reliability
- throughput
- trustworthiness,
- comprehensibility

Constraints:

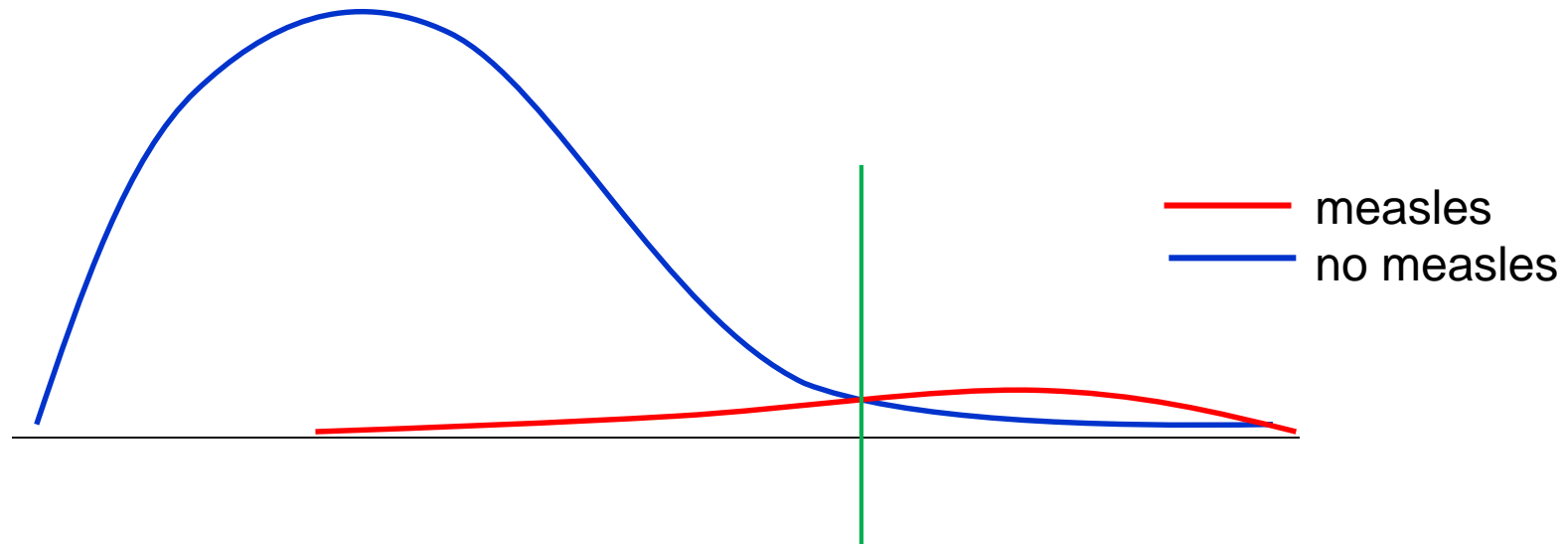
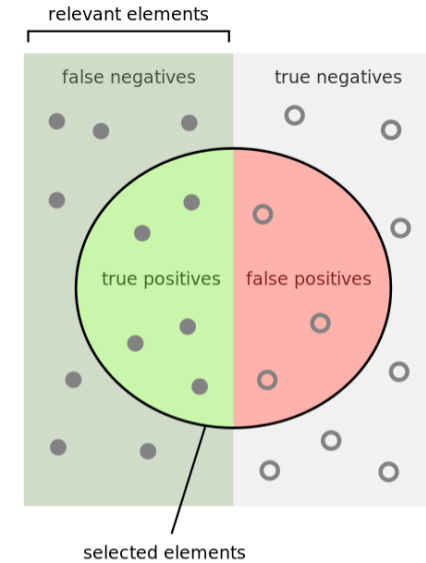
- data availability
- data collection
- groundtruth
- cloud or on-prem
- privacy & legal restrictions
- compute power
- development resources
- timeframe

Problem Formulation: 3. Metrics

- TP, FP, TN, FN,
Precision, Recall, F1, AUC

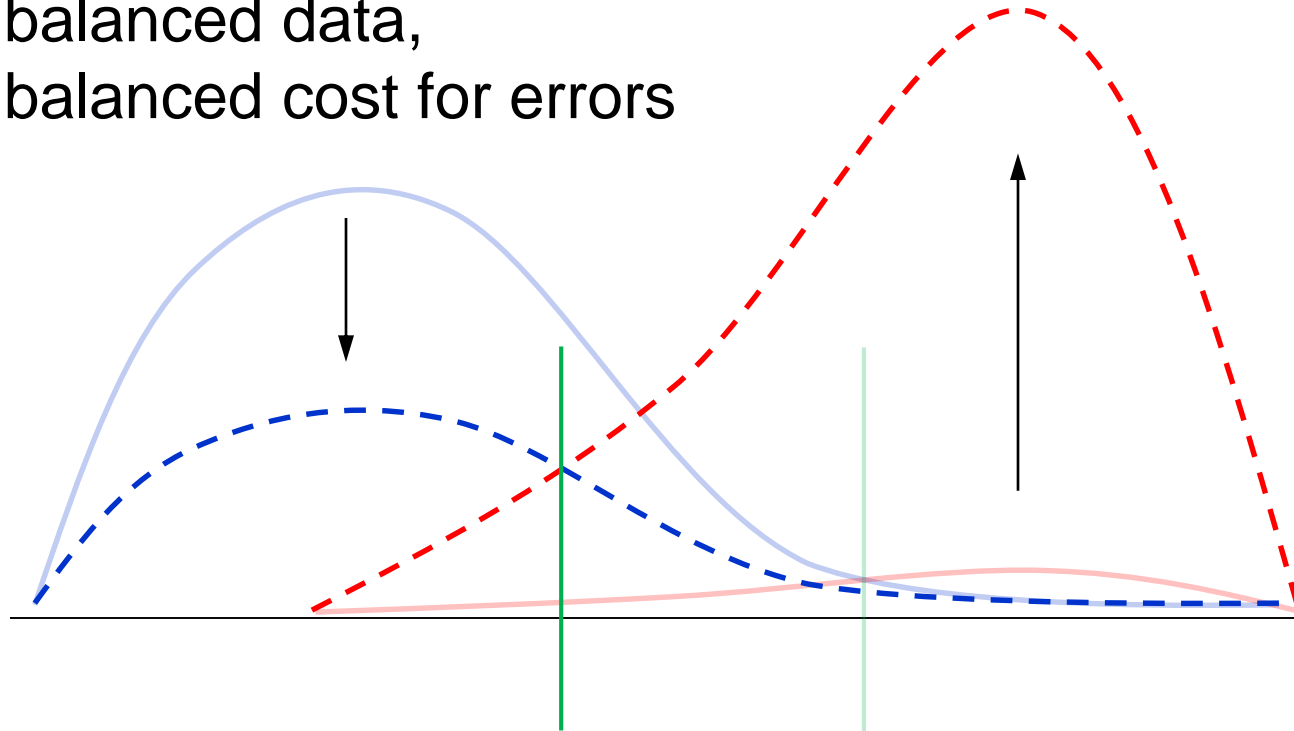
https://en.wikipedia.org/wiki/Precision_and_recall

- Imbalanced data,
- Imbalanced cost for errors



Problem Formulation: 3. Metrics

- Imbalanced data,
- Imbalanced cost for errors



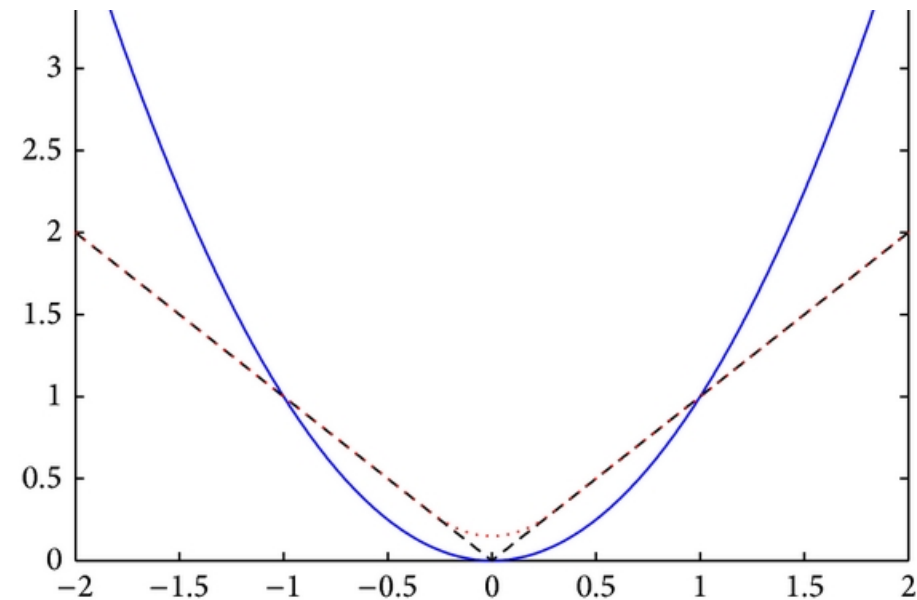
- Individual metrics on multiple criteria

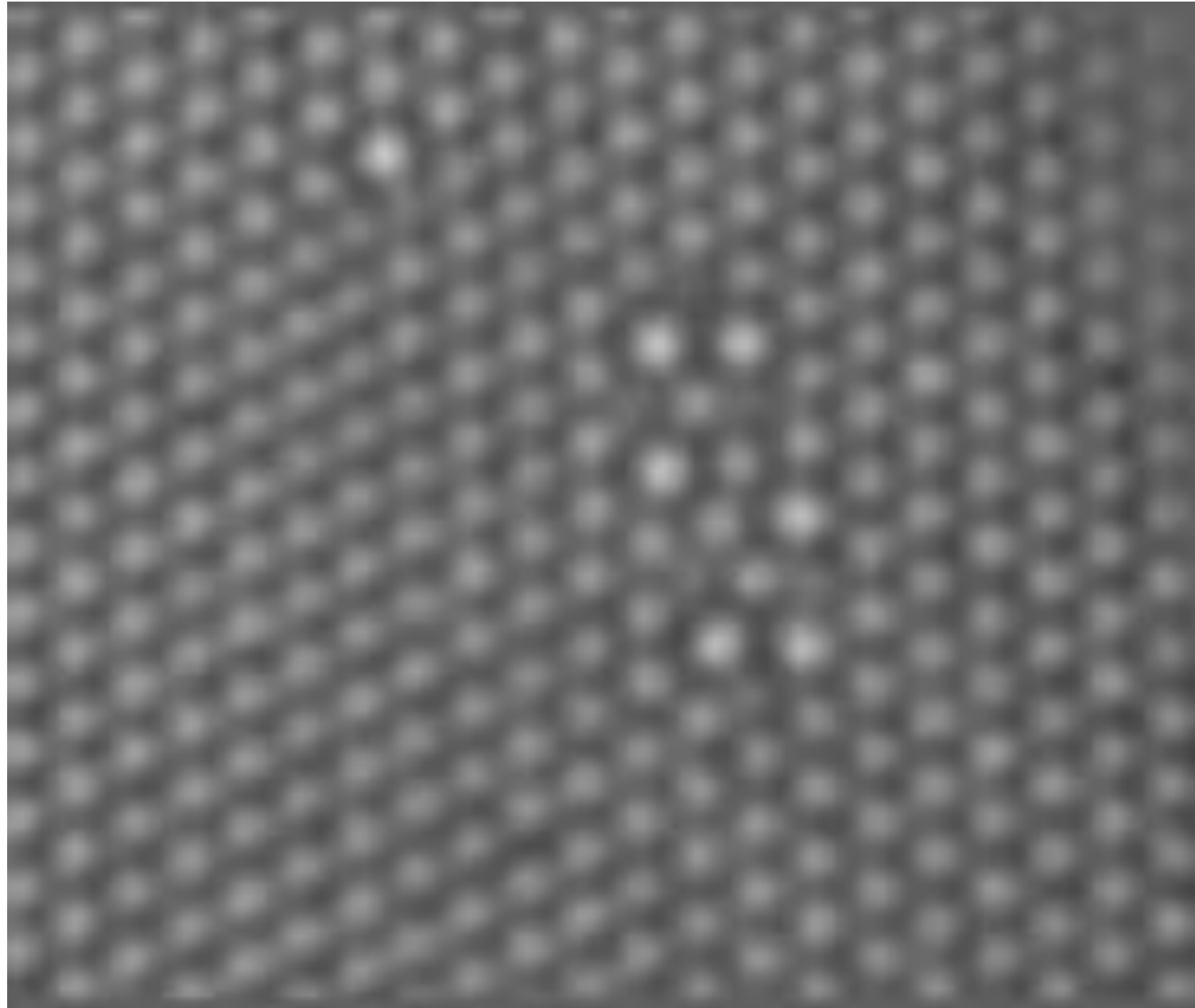
- articulate tradeoffs



- Aggregate metrics

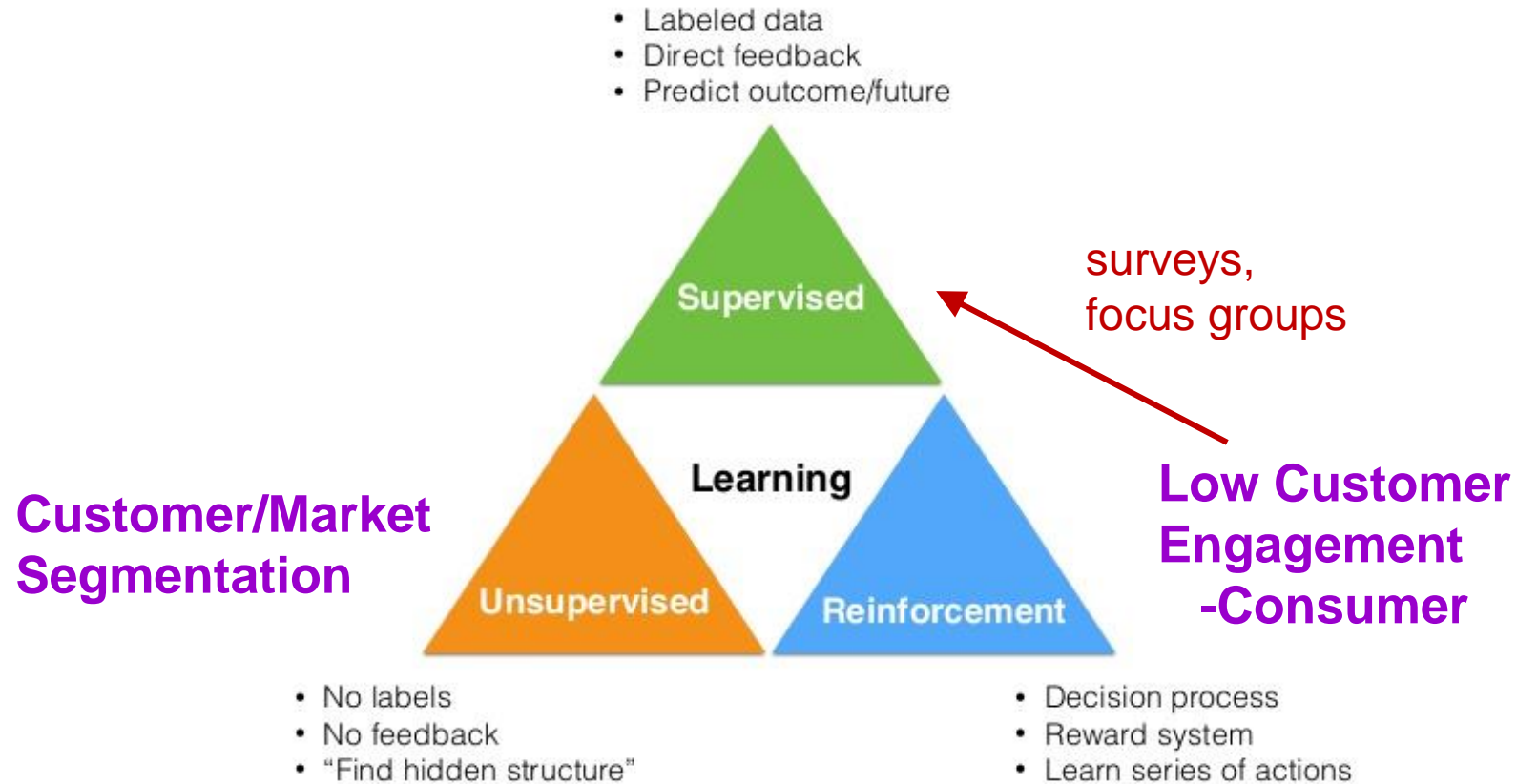
- weight individual metrics
- "loss function"
- nonlinear error





<u>Phase</u>	<u>Scale</u>	<u>Method</u>
Data Grokking	10^2	manual
Model Development	$10^3 - 10^5$	outsource
Production	$10^6 +$	automated

High Customer Engagement - B2B



Analogy Map

AI / ML Concept

Technical Meaning

Analogy in Project Management

Problem Formulation

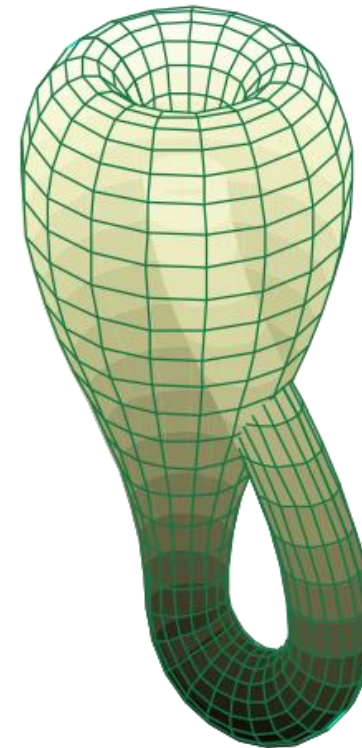
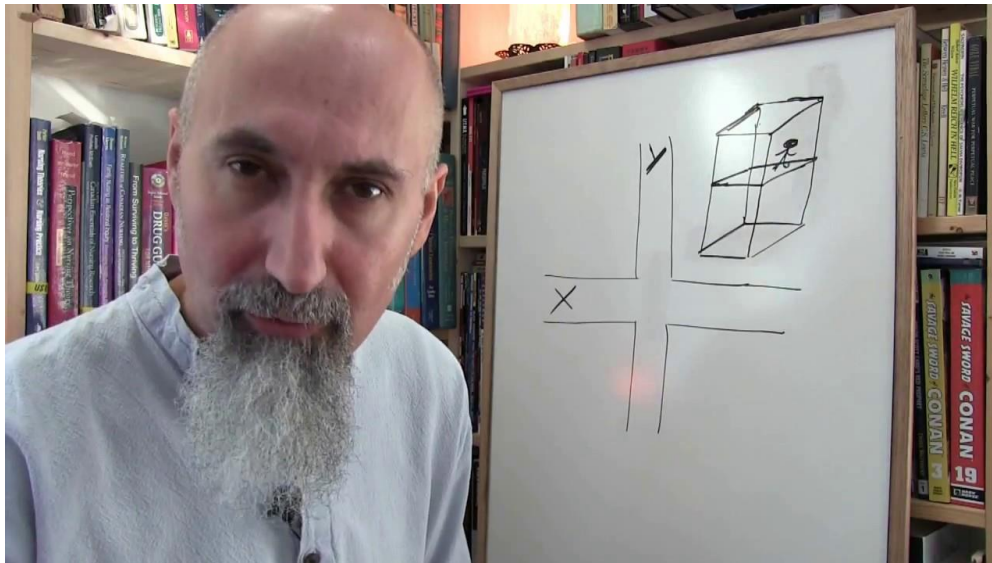
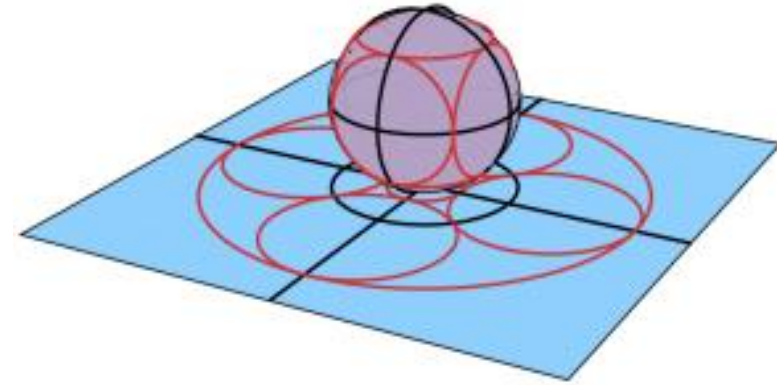
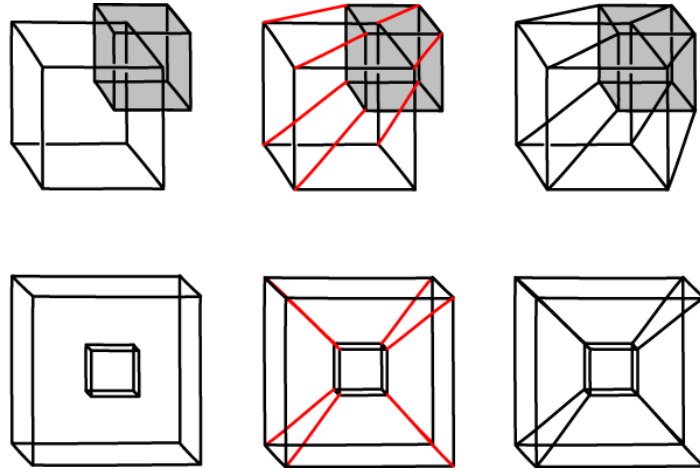
- grokking data
- requirements
- metrics
- data preparation
- models



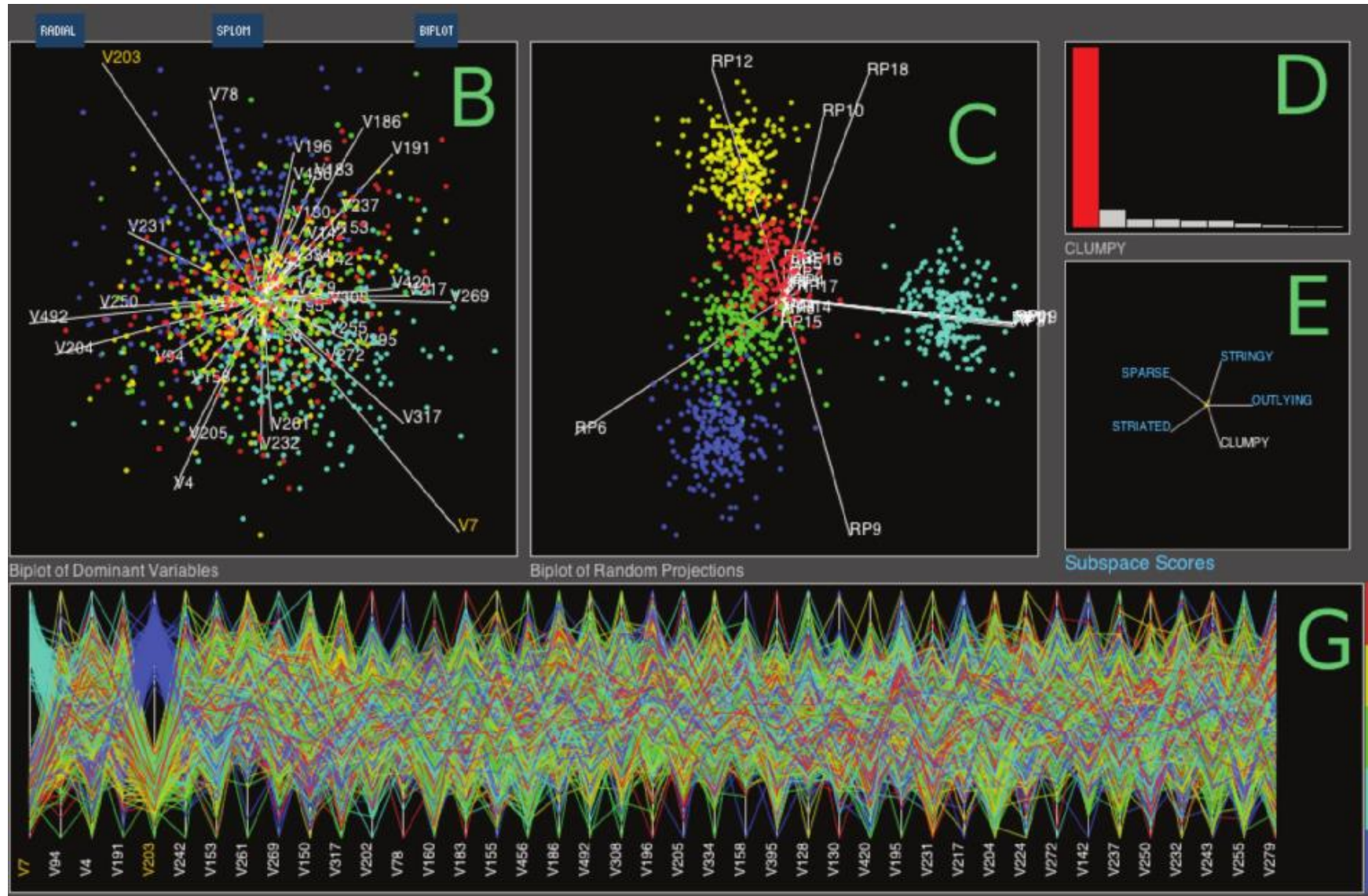
Dimensionality

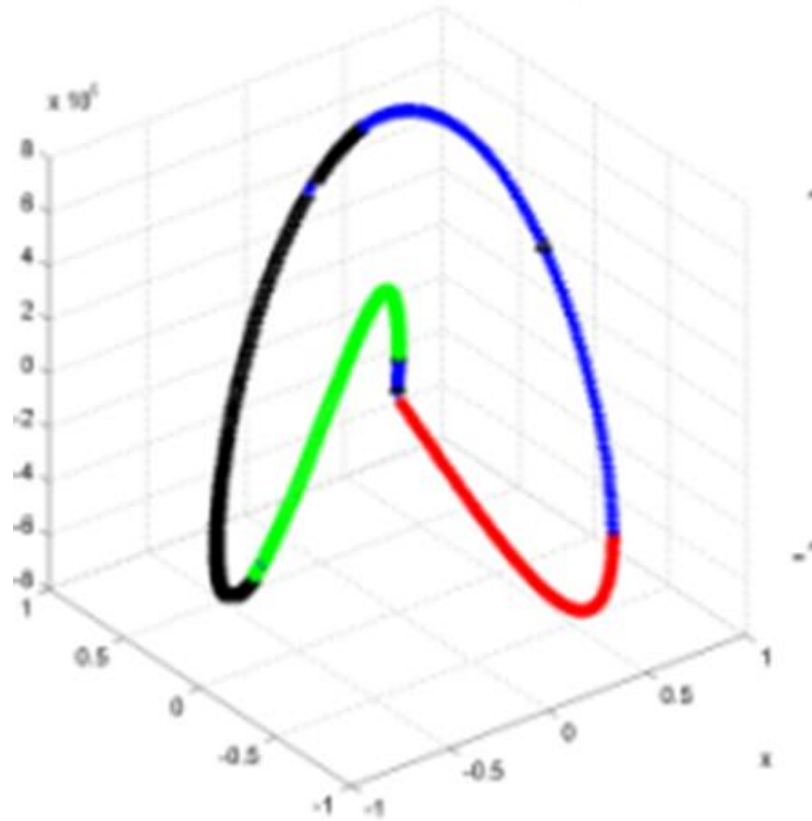
- data
- model
- search
- sampling
- overfitting

Dimensionality



Dimensionality: 6. Data Dimensionality

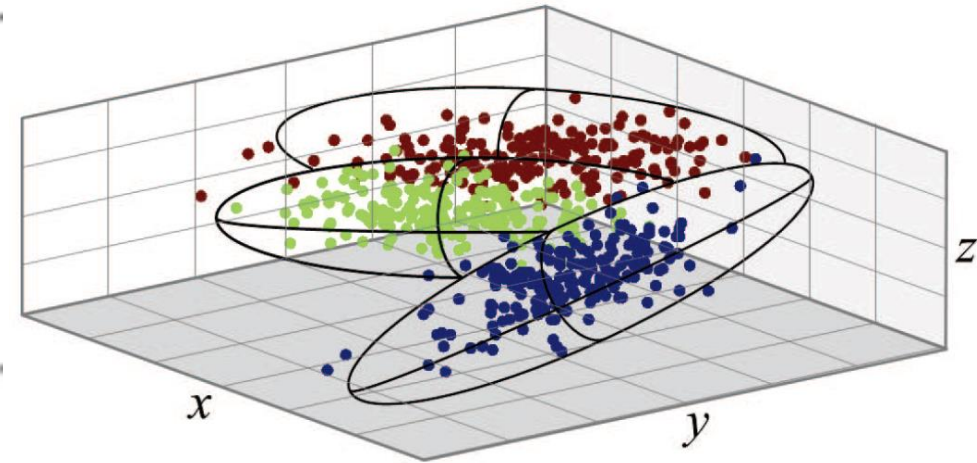


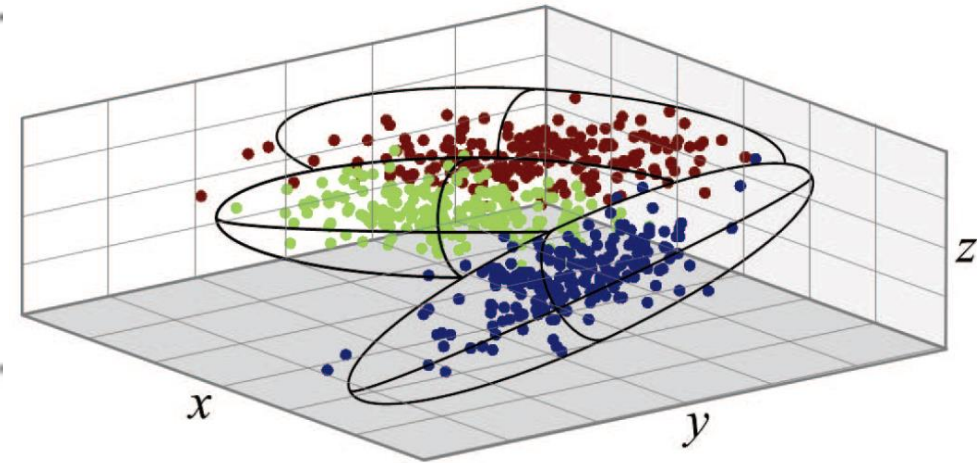
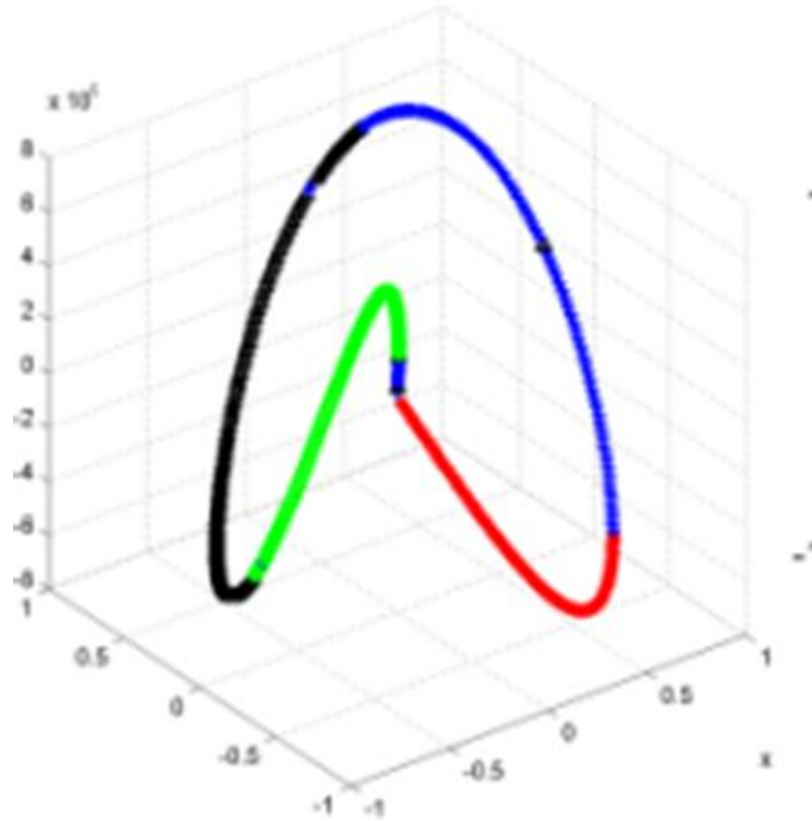


Manifolds

&

Clusters



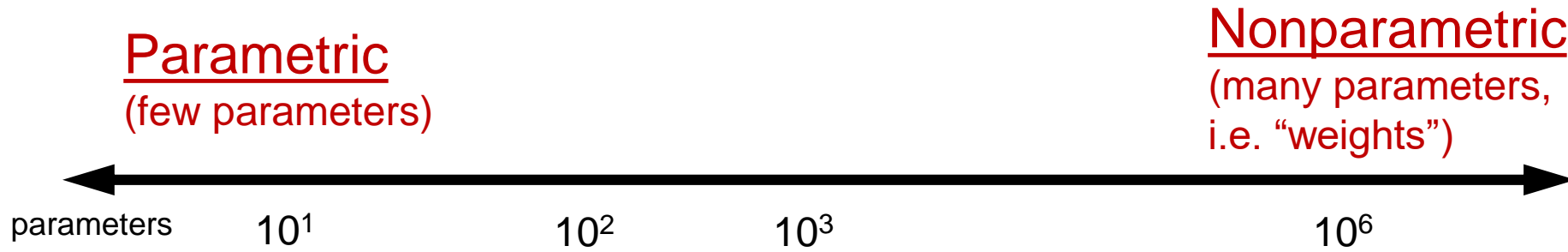


Manifolds

&

Clusters

Dimensionality: 7. Model Dimensionality



Dimensionality: 7. Model Dimensionality

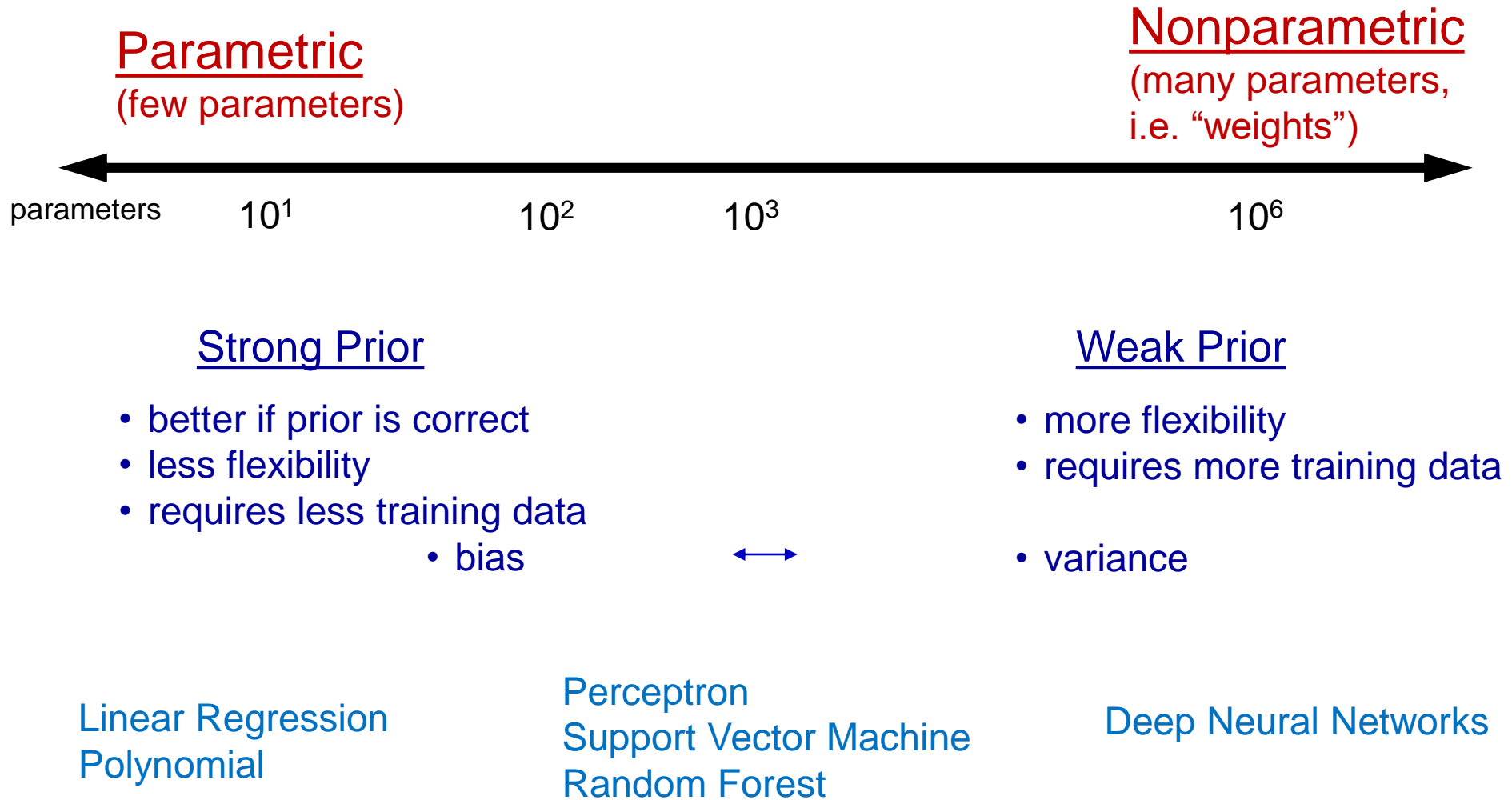


Few model parameters



Many model parameters

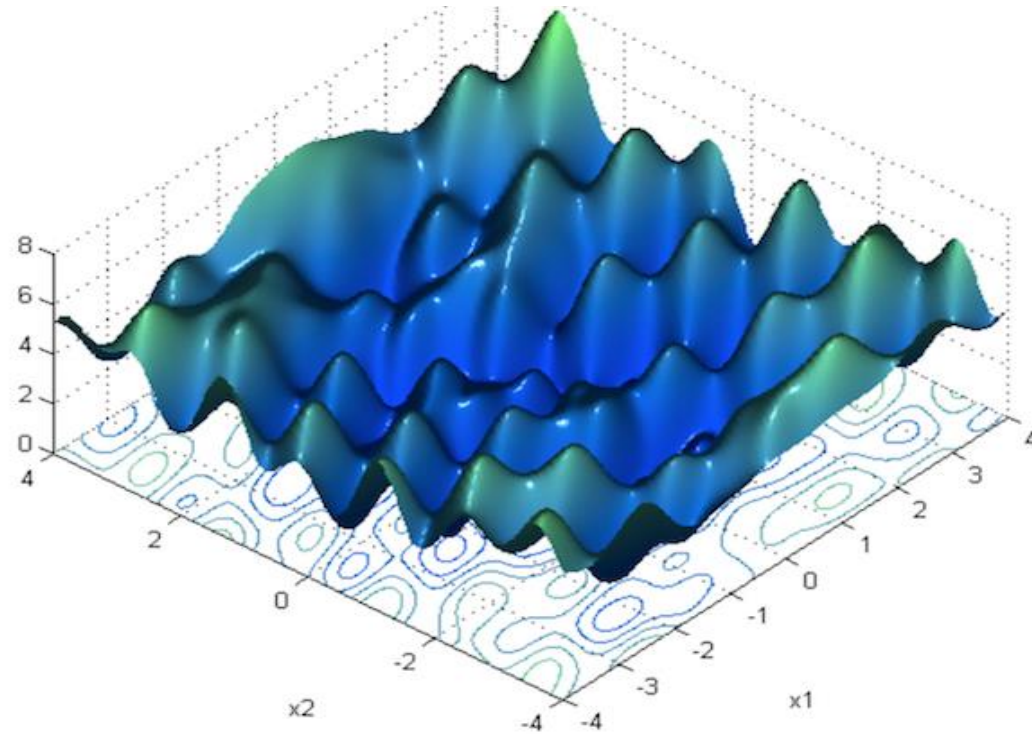
Dimensionality: 7. Model Dimensionality



And this shows only two dimensions!

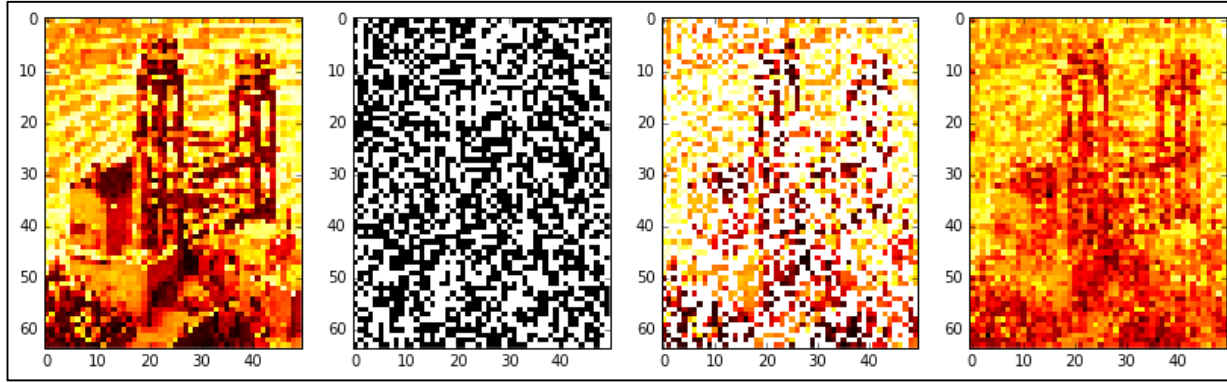
Search Strategies

- Grid search
 - systematically vary parameters under your control
- Hill climbing
 - keep improving what you've got
- Random restart
 - periodically try something new
- Parallelize
 - multiple teams
- Evolutionary
 - combine modular parts that contribute to better solutions



https://ml4a.github.io/ml4a/how_neural_networks_are_trained/

Dimensionality: 9. Sampling



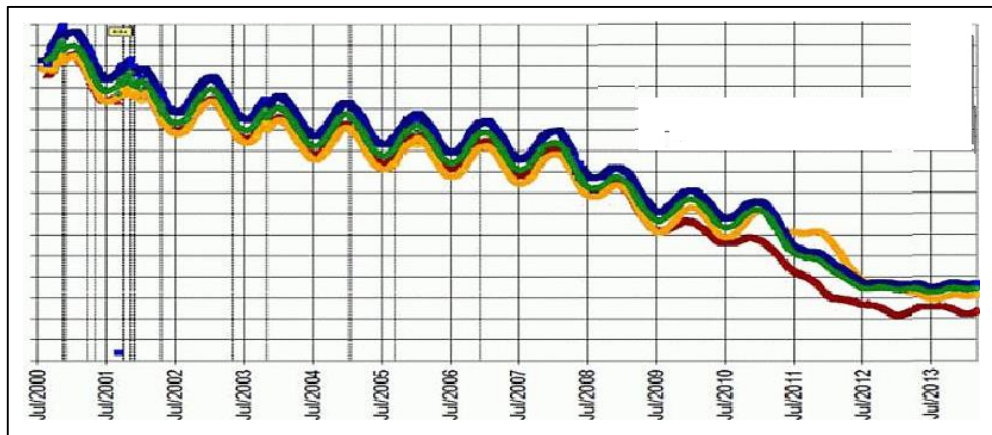
ML

- Signal & Noise

- Sample size
- Error bars
- Coverage

Management

- Bias/Variance: Does that data point make sense?
- Targeted sampling.

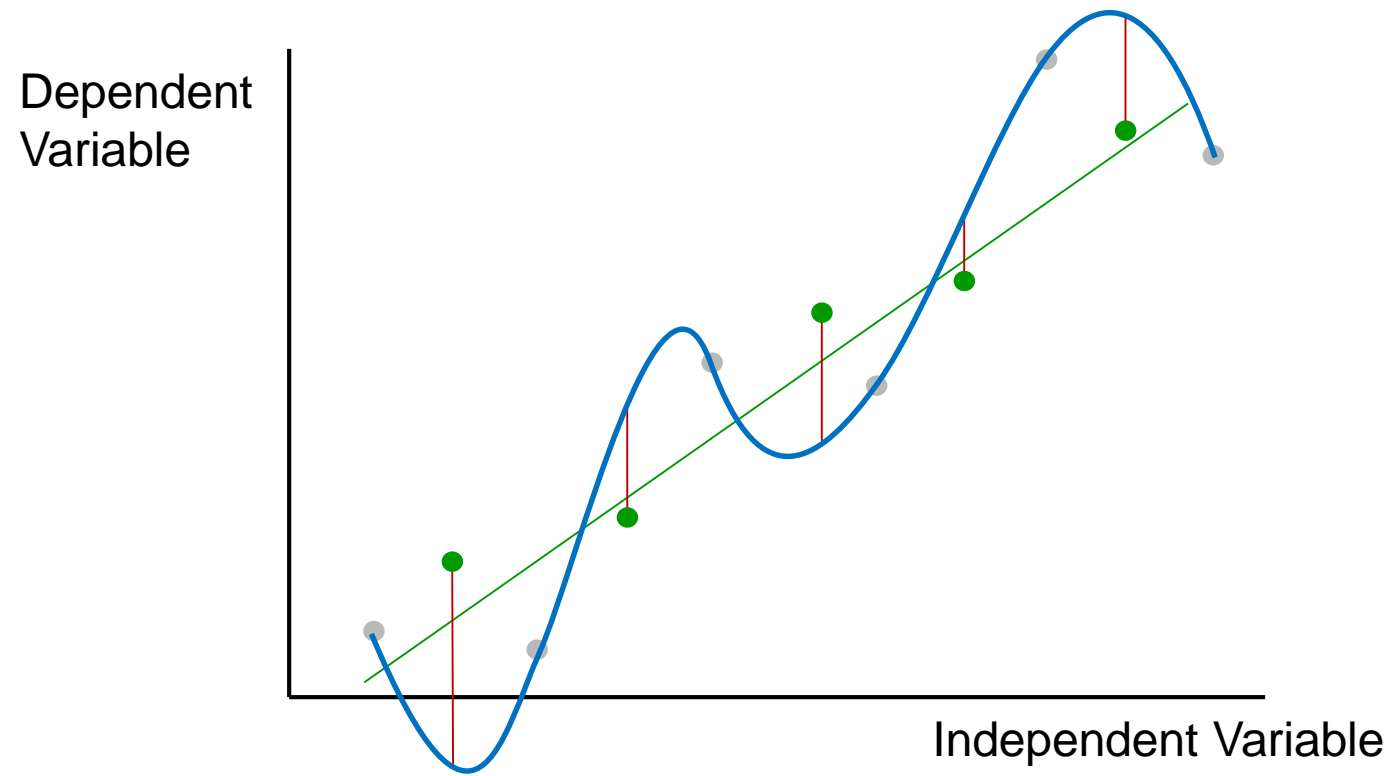


- Nonstationary Data

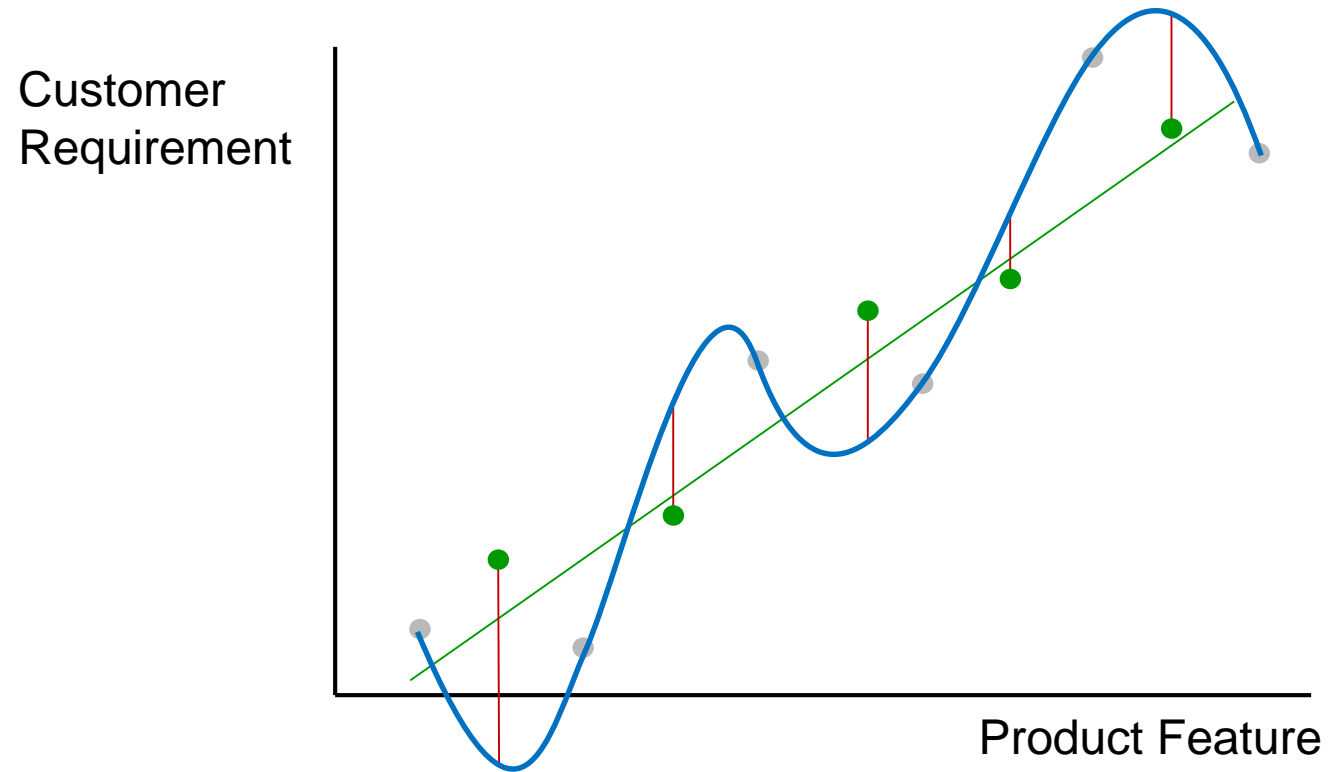
- Retraining in production

- Continual product revision

Dimensionality: 10. Overfitting

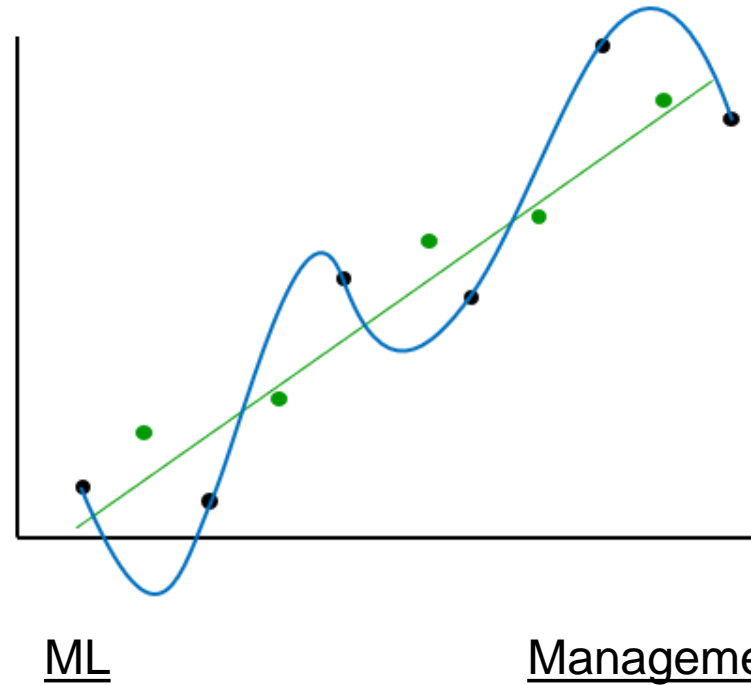


Dimensionality: 10. Overfitting



- Charter Customer
- Other Customers

Remedies to Overfitting



- Regularization

Bias: prefer smooth over wiggly.

Reminders of big picture: don't deep-end on subsets.

- Cross-Validation

Train on subset of available (*black*) data points; test on remainder.

Quiz: does a satisfaction of some initial requirements throw others out of whack?

Analogy Map

AI / ML Concept

Technical Meaning

Analogy in Project Management

Problem Formulation

- grokking data
- requirements
- metrics
- data preparation
- models



Dimensionality

- data
- model
- search
- sampling
- overfitting





Eric Saund

- *Research scientist in Cognitive Science and AI.*
- *Conversational Agents, Visual Perception, Cognitive Architectures.*
- *I build stuff.*

Projects

Papers

Curiosities

Links

Contact

<http://www.saund.org>

saund@alum.mit.edu

Conversation