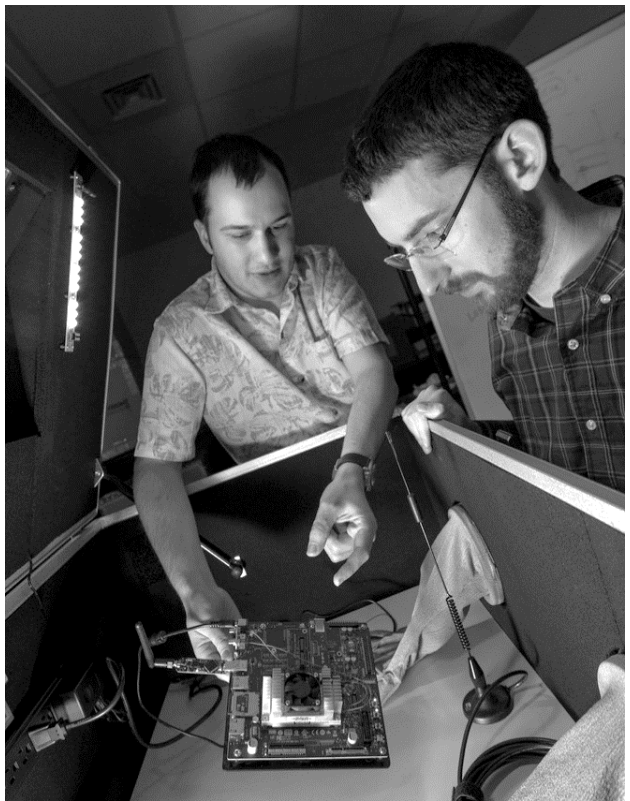


VIRGINIA TECH NATIONAL SECURITY INSTITUTE

Introduction to Machine Learning

DATAWorks 2023

Stephen Adams

Associate Research Professor

Assistant Director, Intelligent Systems Division

Virginia Tech National Security Institute

April 25, 2023

NATIONAL SECURITY INSTITUTE
VIRGINIA TECH

Overview

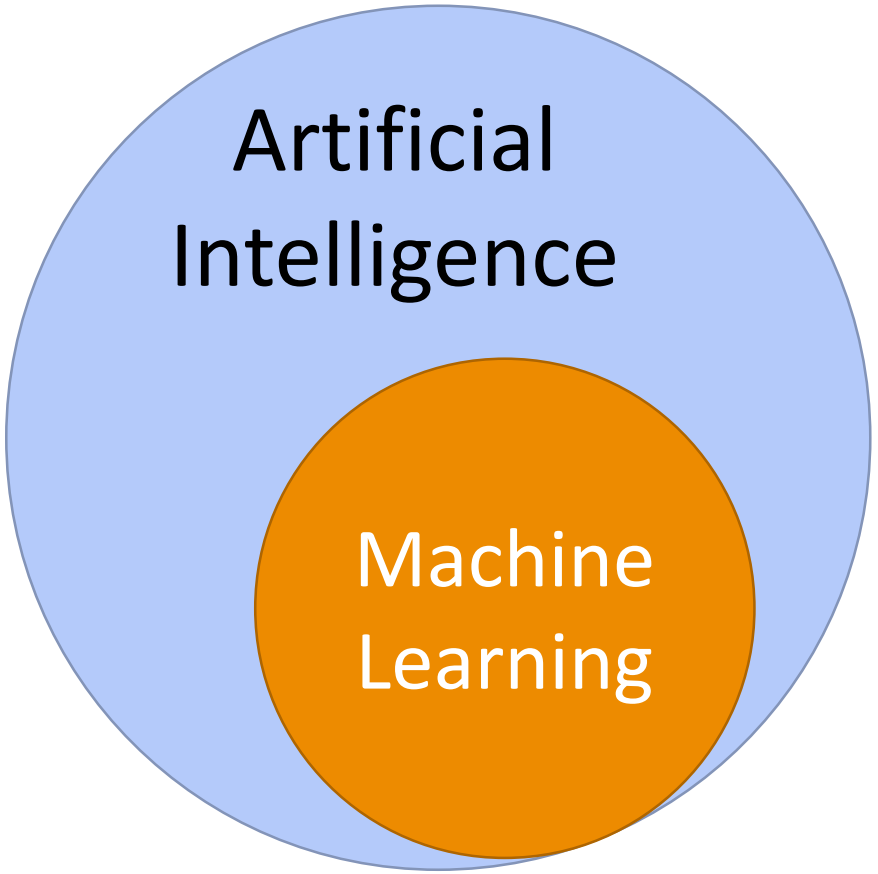
Lectures

- Introduction to Machine Learning
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Demonstrations

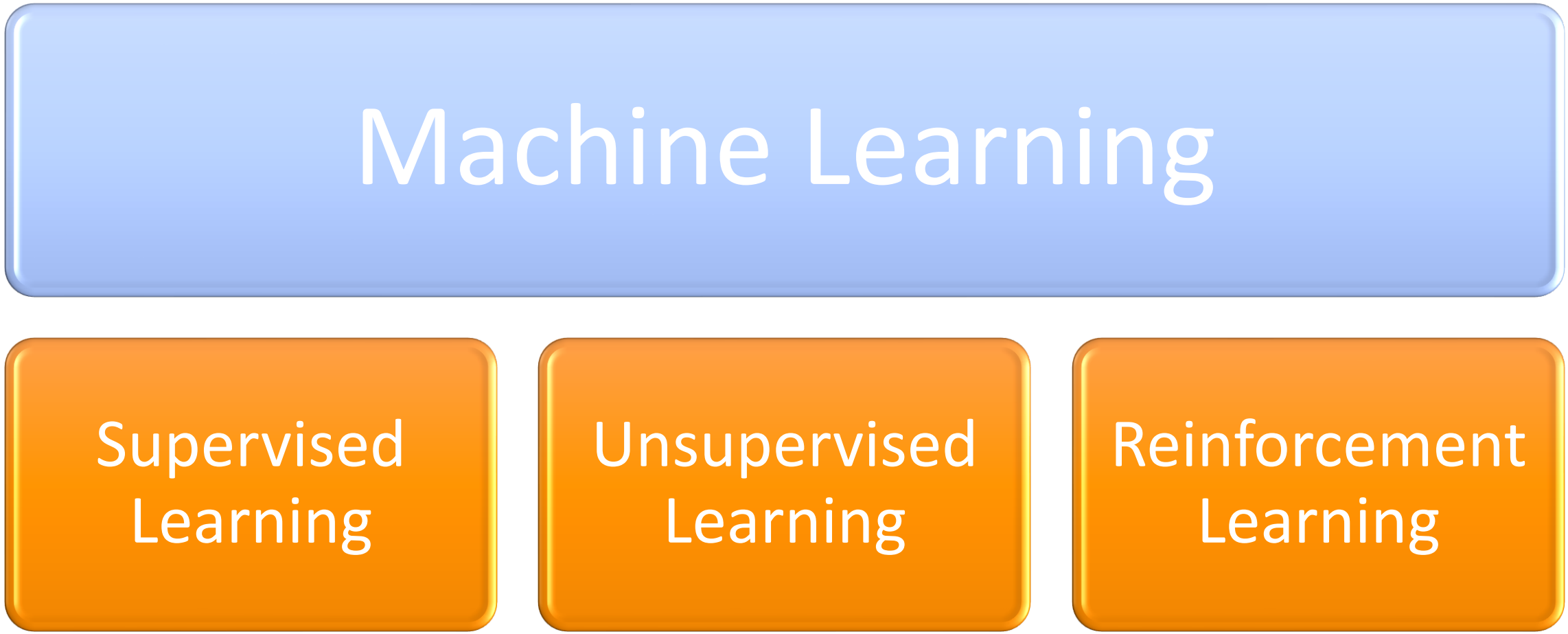
- Python
  - Numpy
  - Pandas
  - Matplotlib
  - Seaborn
  - Statsmodels
  - Scikit learn
  - Gymnasium
- Jupyter notebooks

ML vs AI

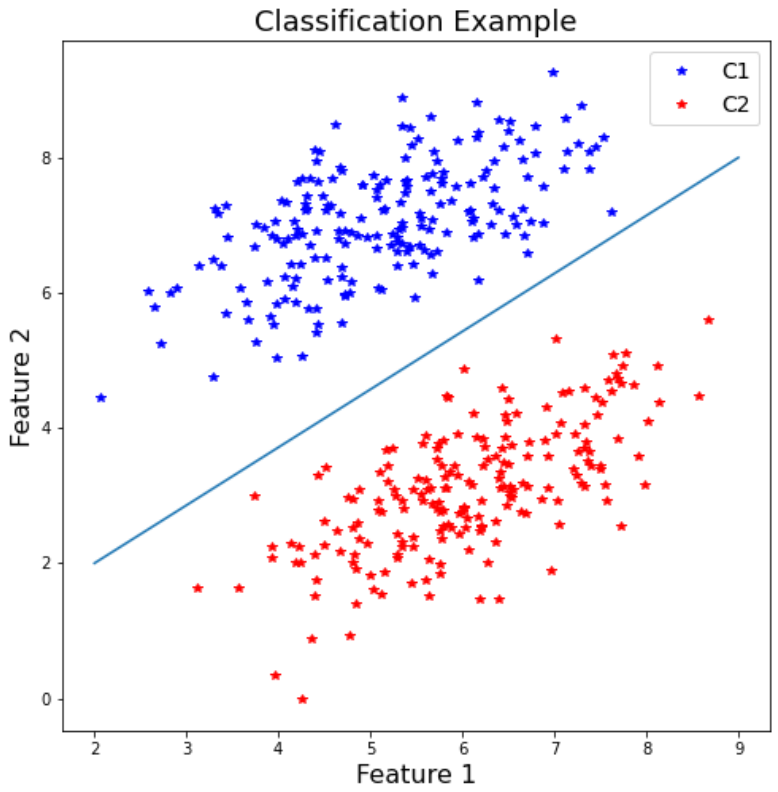


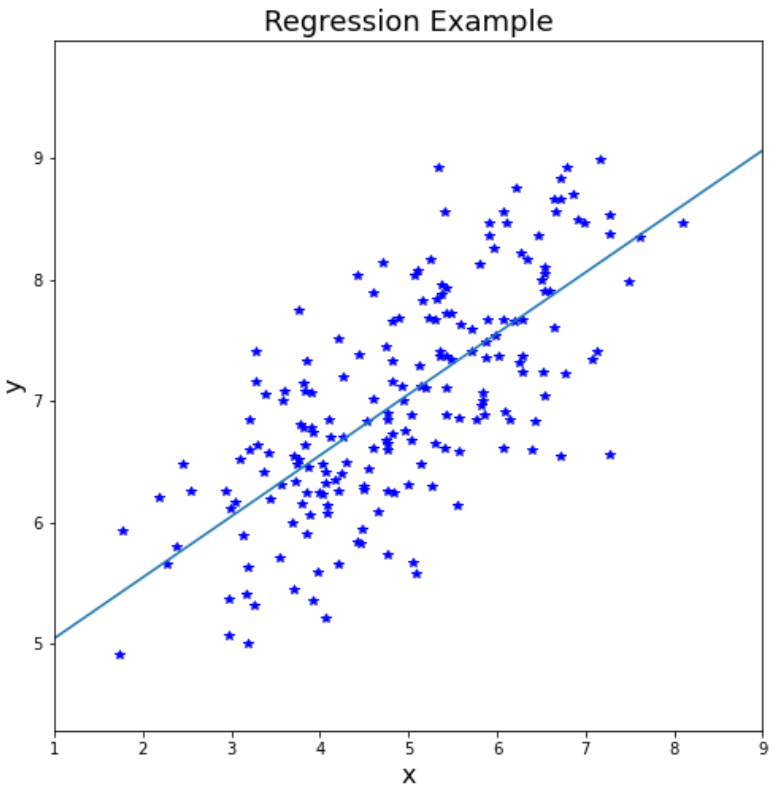
- Artificial intelligence – any system that makes a decision using a computer
  - Rule-based systems
- Machine learning – learns how to make decisions from data or experience
  - ChatGPT

Machine Learning

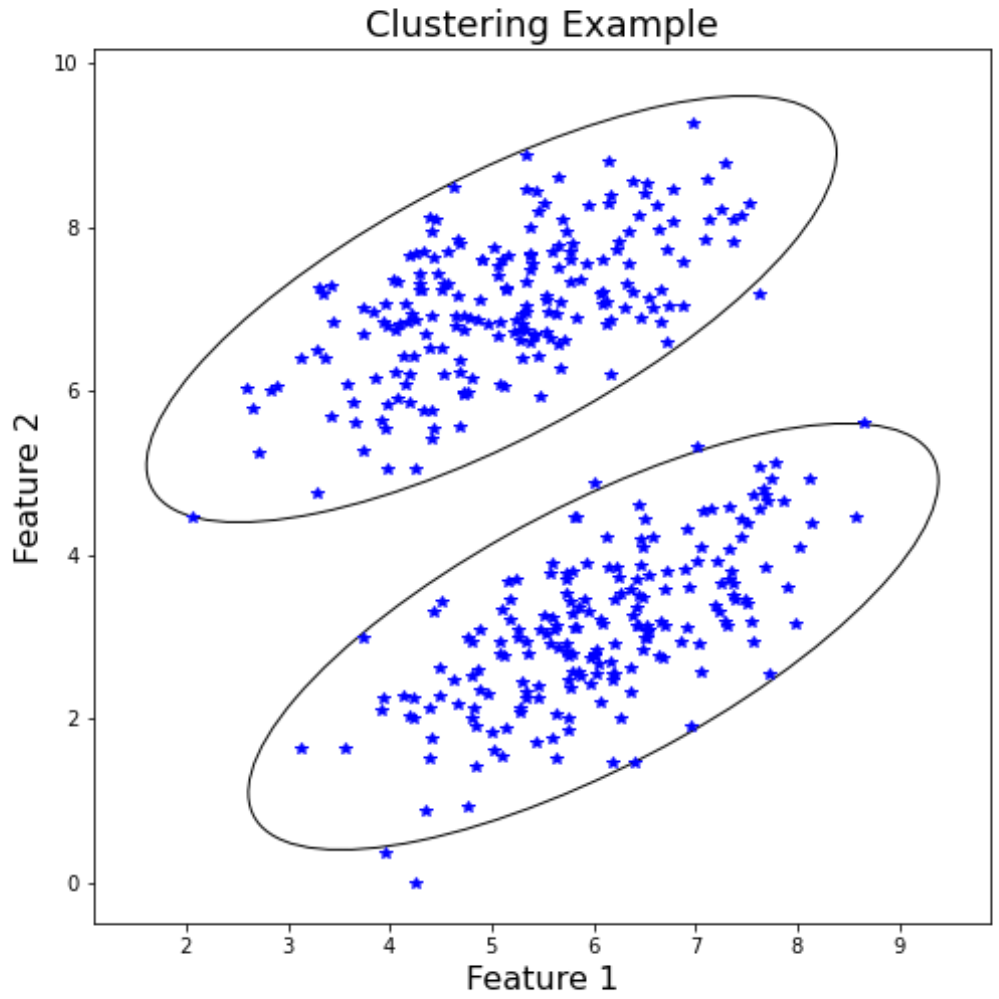


Supervised Learning

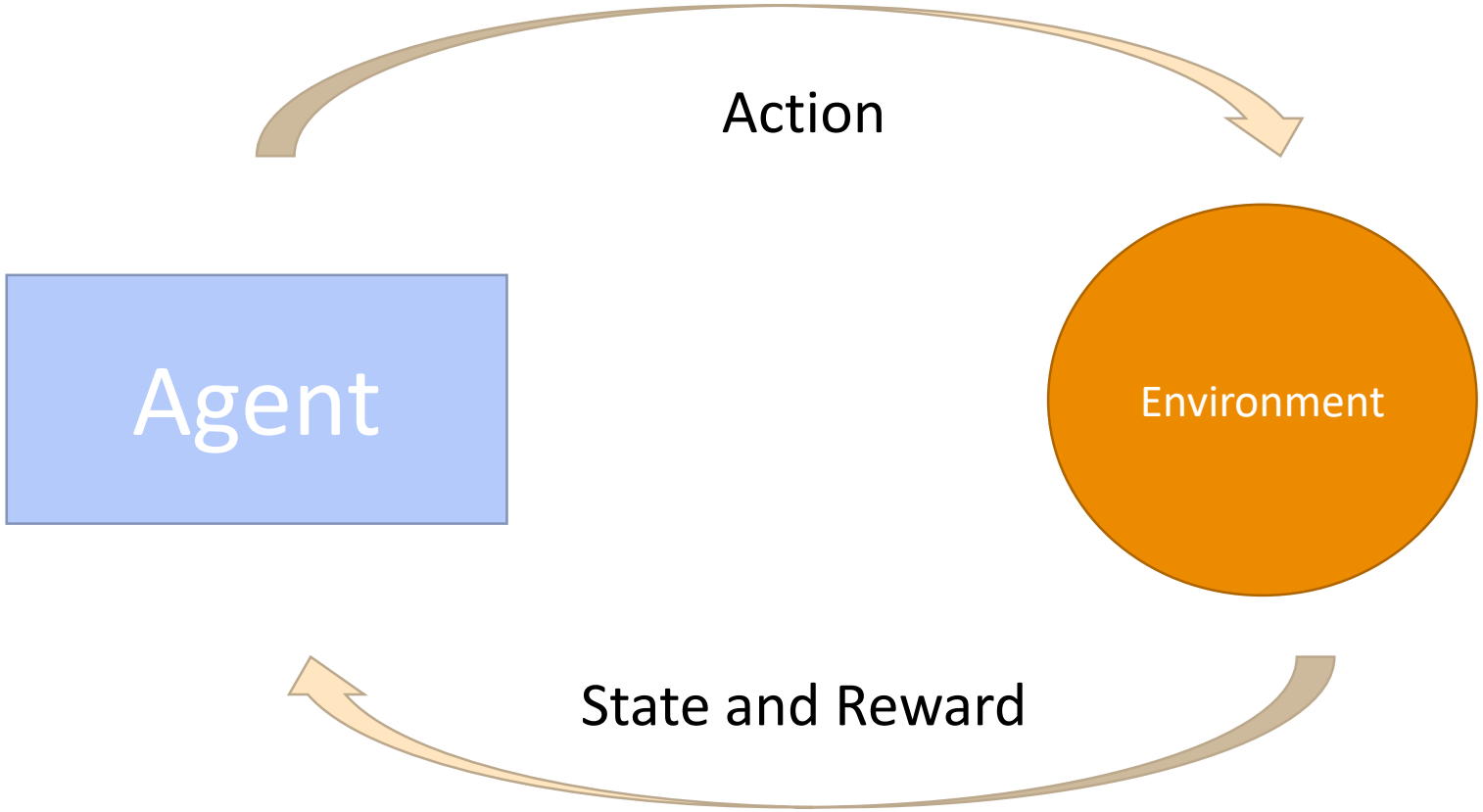




Unsupervised Learning

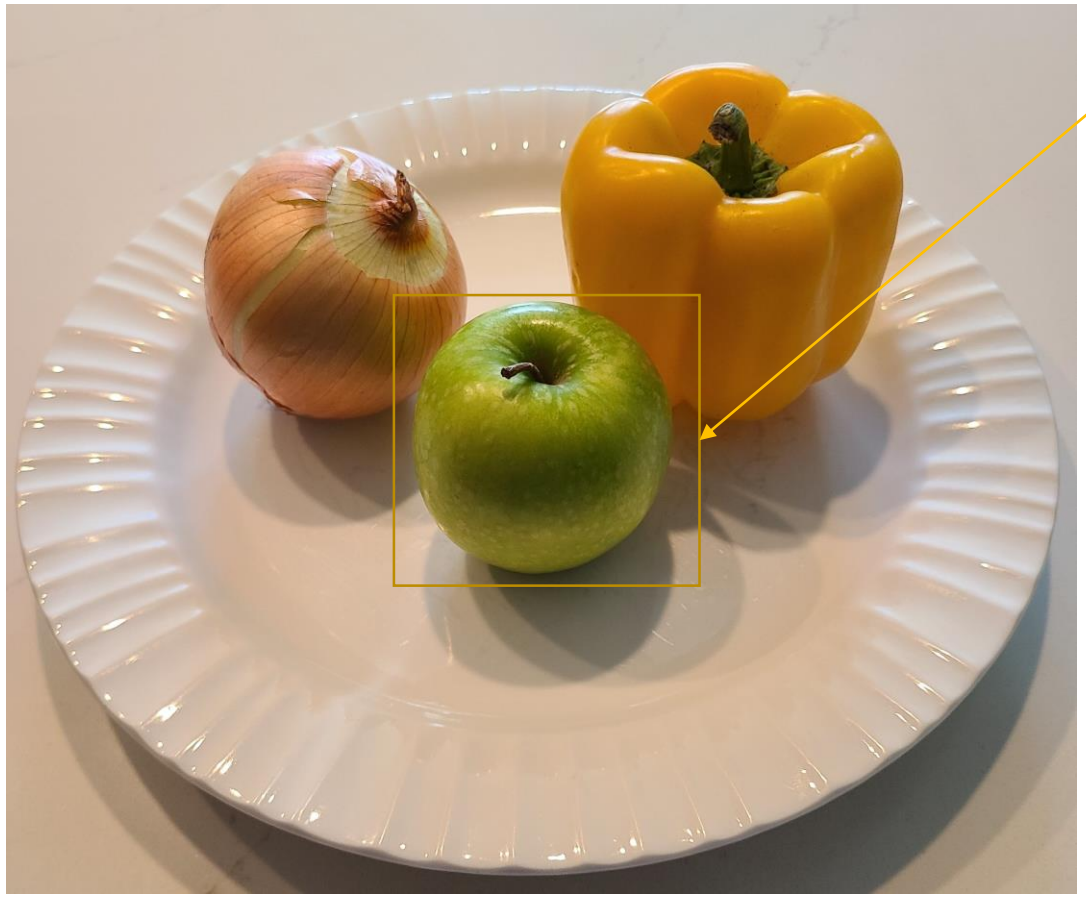


Reinforcement Learning



Machine Learning Applications

Computer Vision



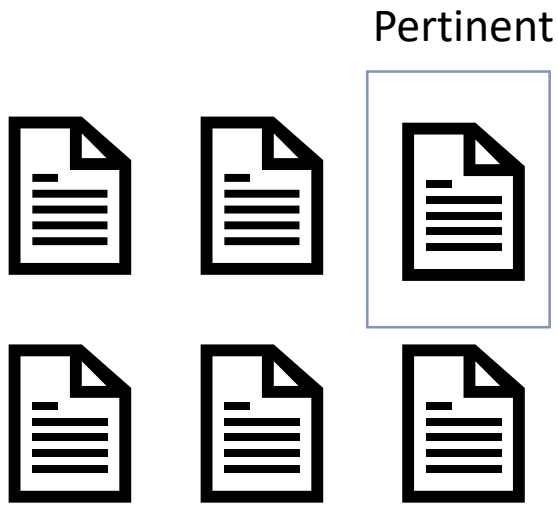
https://www.youtube.com/watch?v=h_gQHshUtGc

- Image classification
- Image segmentation
- Video
- Bounding boxes
- Multi-label
- Feature extraction
- Convolution neural networks

Natural Language Processing

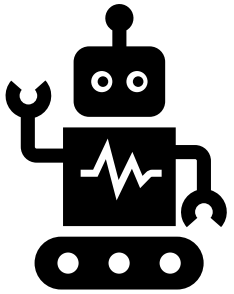
The Introduction to Machine Learning Short Course at DATAWorks 2023 will cover supervised learning, unsupervised learning, and reinforcement learning.

- Proper nouns
- Adjectives



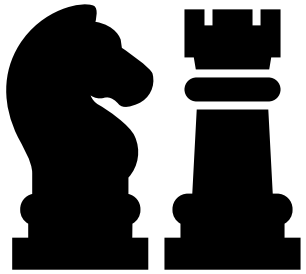
https://www.youtube.com/watch?v=7vQM4pRqg5c

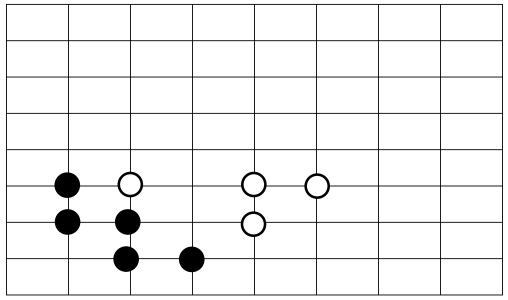
Open AI's ChatGPT



- Topic modeling and clustering
- Document classification
- Part-of-speech identification
- Translation
- Sentence/document completion
- Generative modeling
- Bayesian models
- Recurrent neural networks

Games



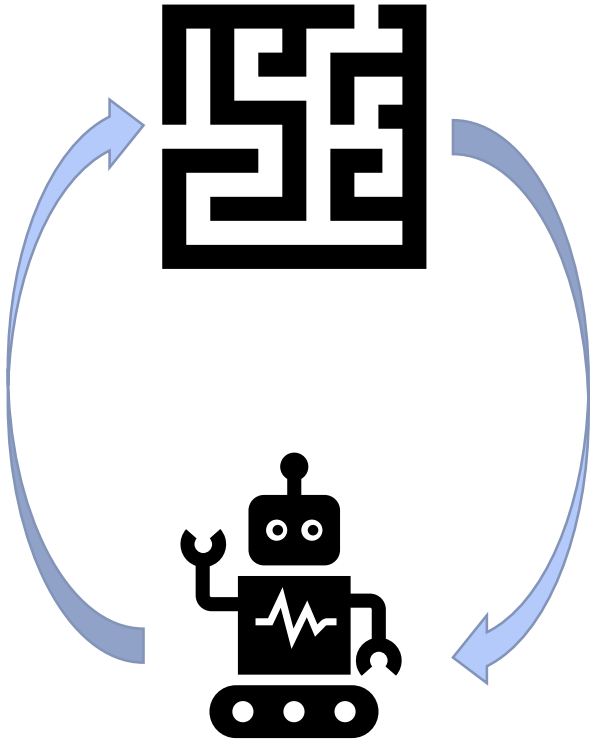


Alpha Go

https://www.youtube.com/watch?v=8tq1C8spV_g

Alpha Star

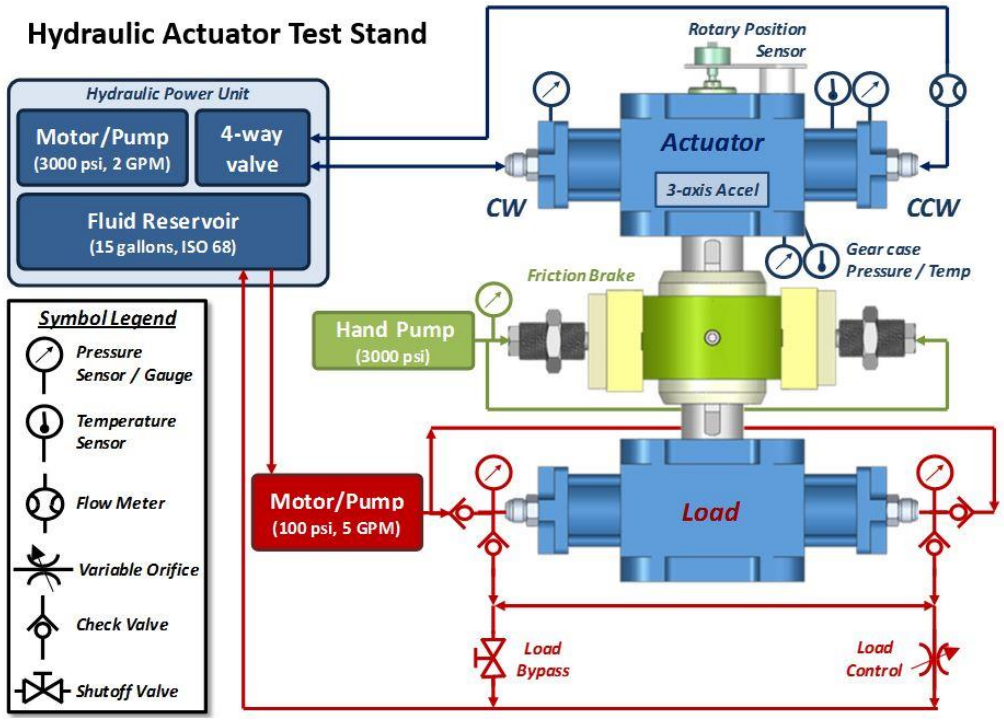
https://www.youtube.com/watch?v=UuhECwm31dM

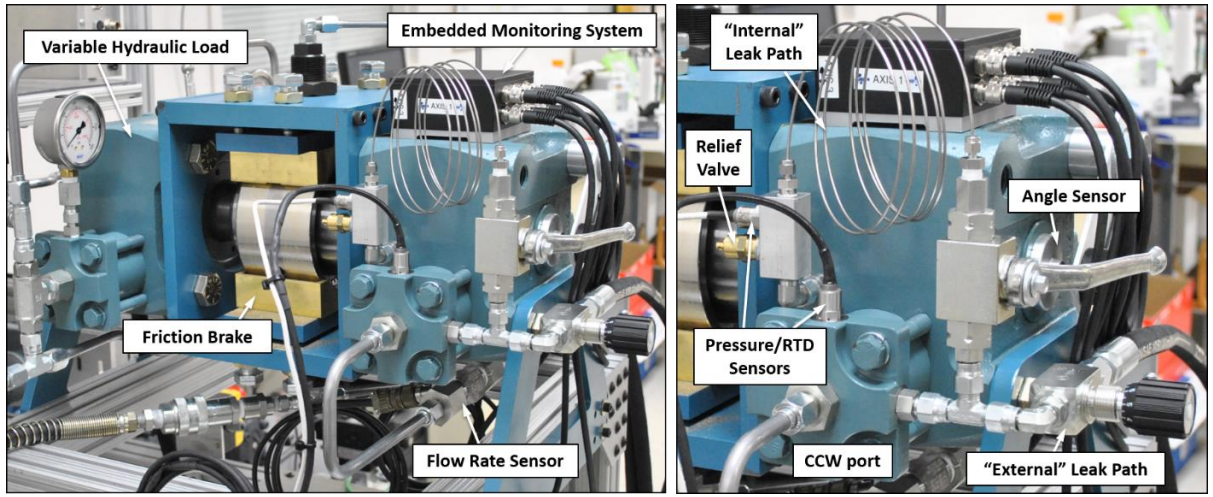


Action

State and Reward

Condition Monitoring





Adams, Stephen, et al. "Condition based monitoring for a hydraulic actuator." *Annual Conference of the PHM Society*. Vol. 8. No. 1. 2016.

Other applications

My research:

- Cyber threat detection
- Credit card fraud
- Control of cyber-physical systems
  - Manufacturing
  - Infrastructure
- Sports analytics

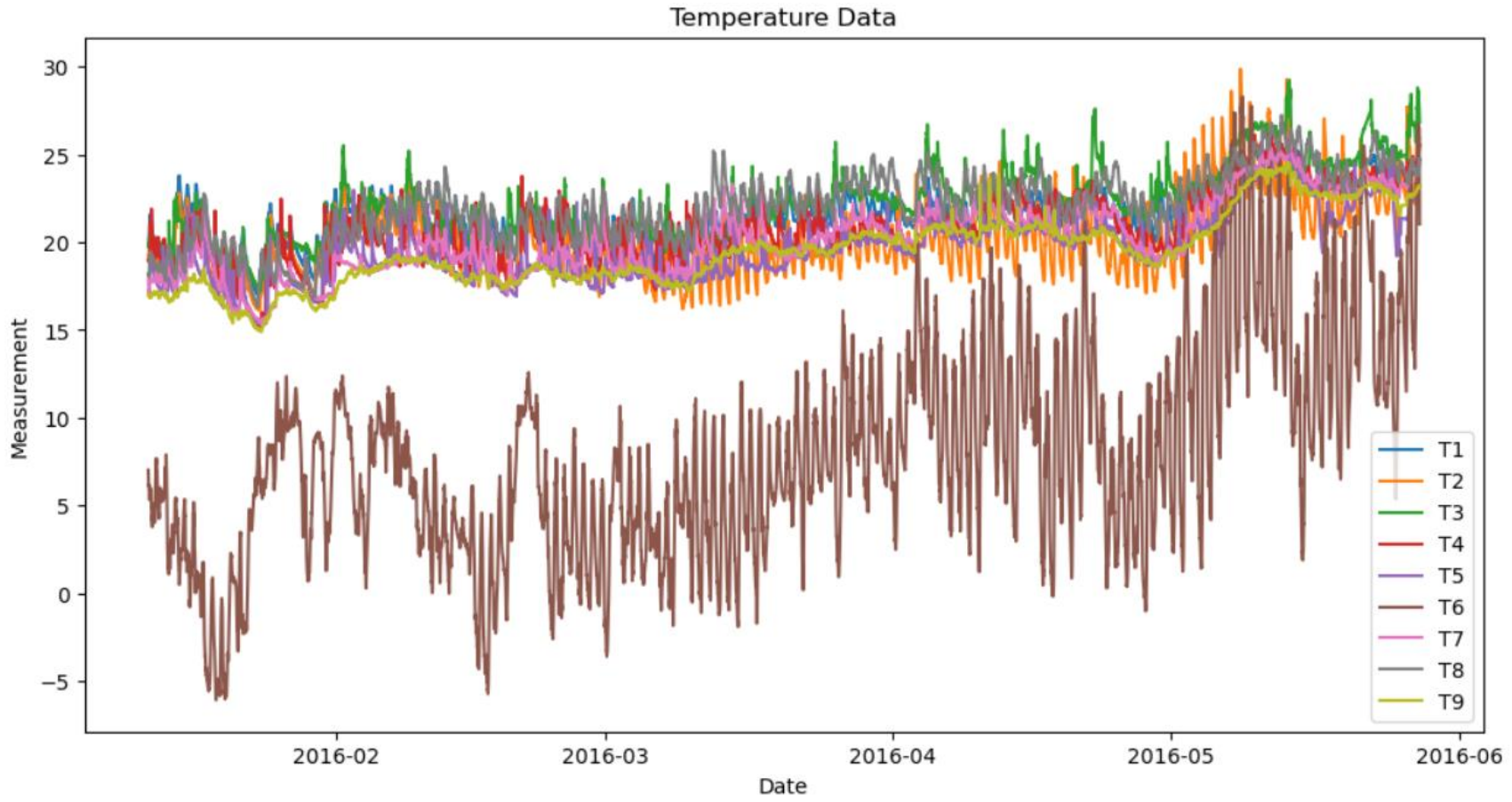
Others?

Limitations of ML

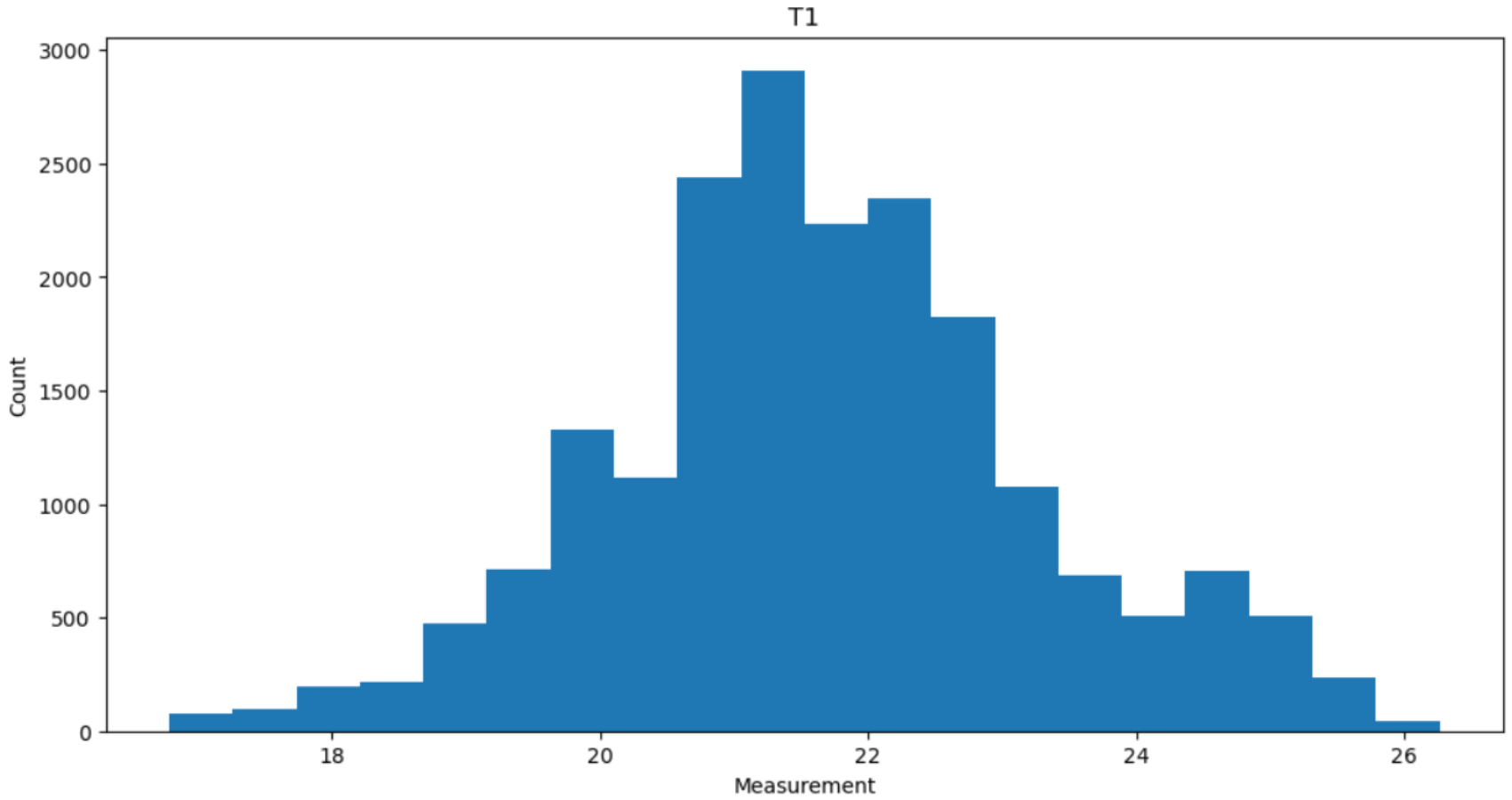
- Requires data (or experience)
  - Collection
  - Labeling
  - Storage
- Generally need a large amount of data
- Computation
- Generalization
  - May make decisions a human would not make
- Can be confidently wrong
- Black-box

Exploratory Data Analysis

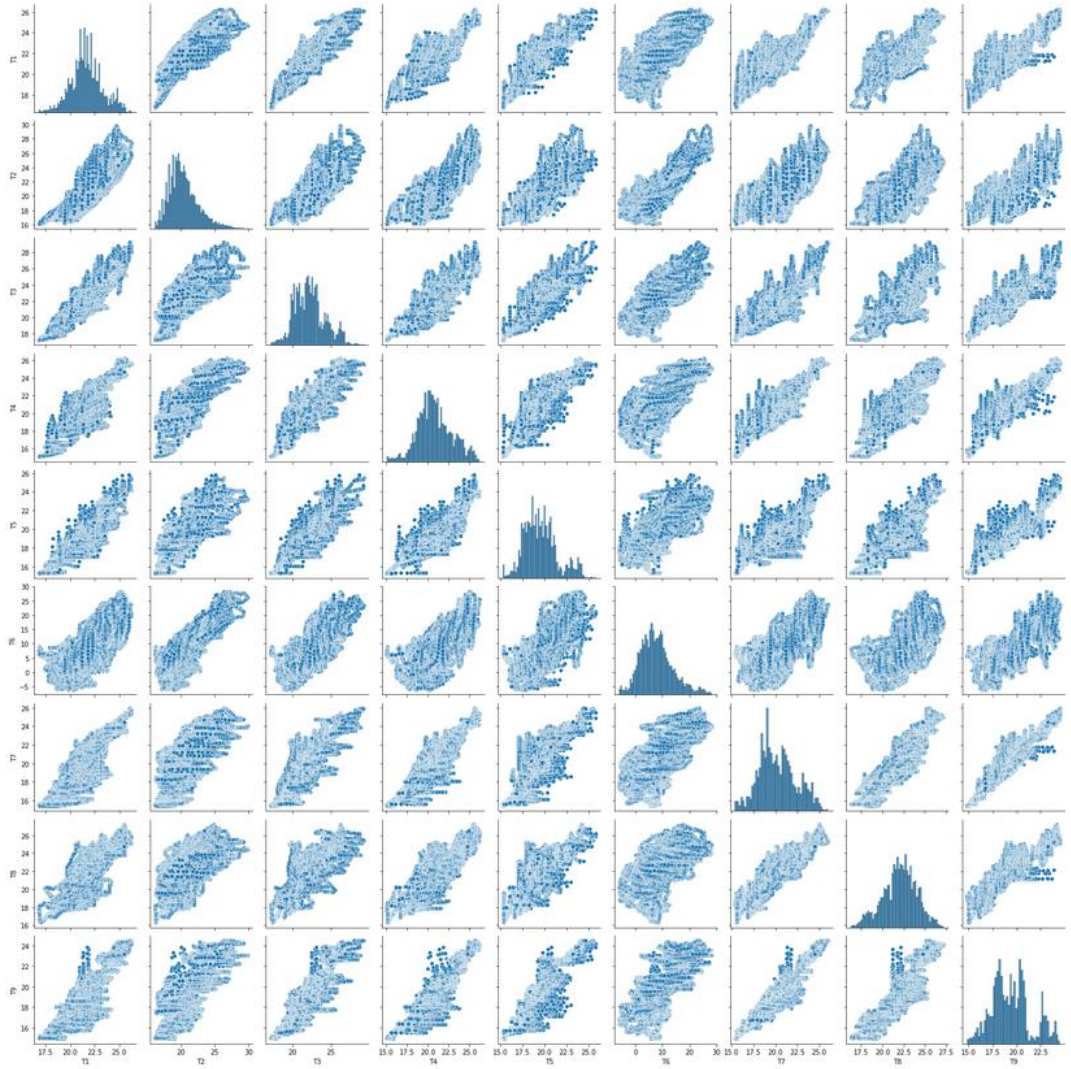
Plots


Temperature Data
Measurement
30
25
20
15
10
5
0
−5
2016-02
2016-03
2016-04
2016-05
2016-06
Date
T1
T2
T3
T4
T5
T6
T7
T8
T9

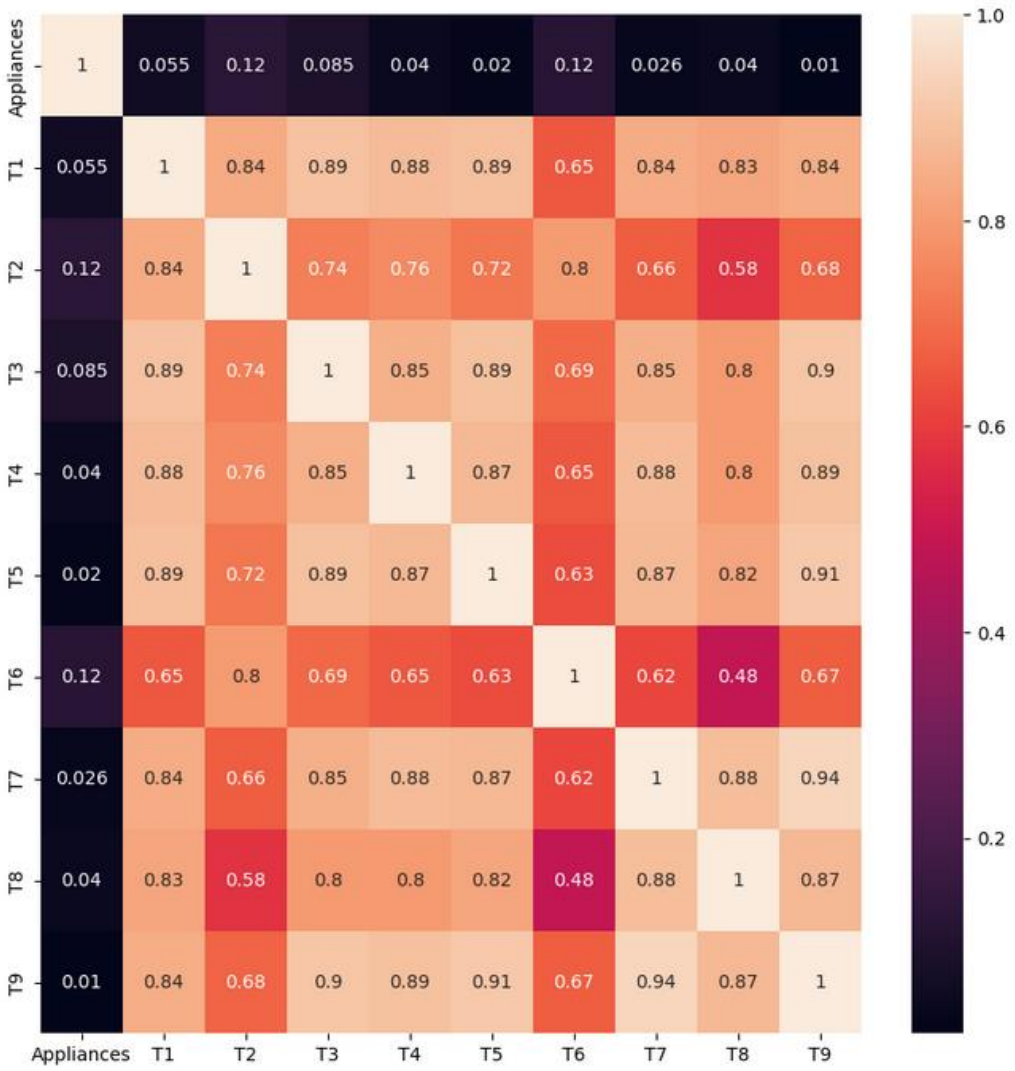
Distributions



Pair Plots

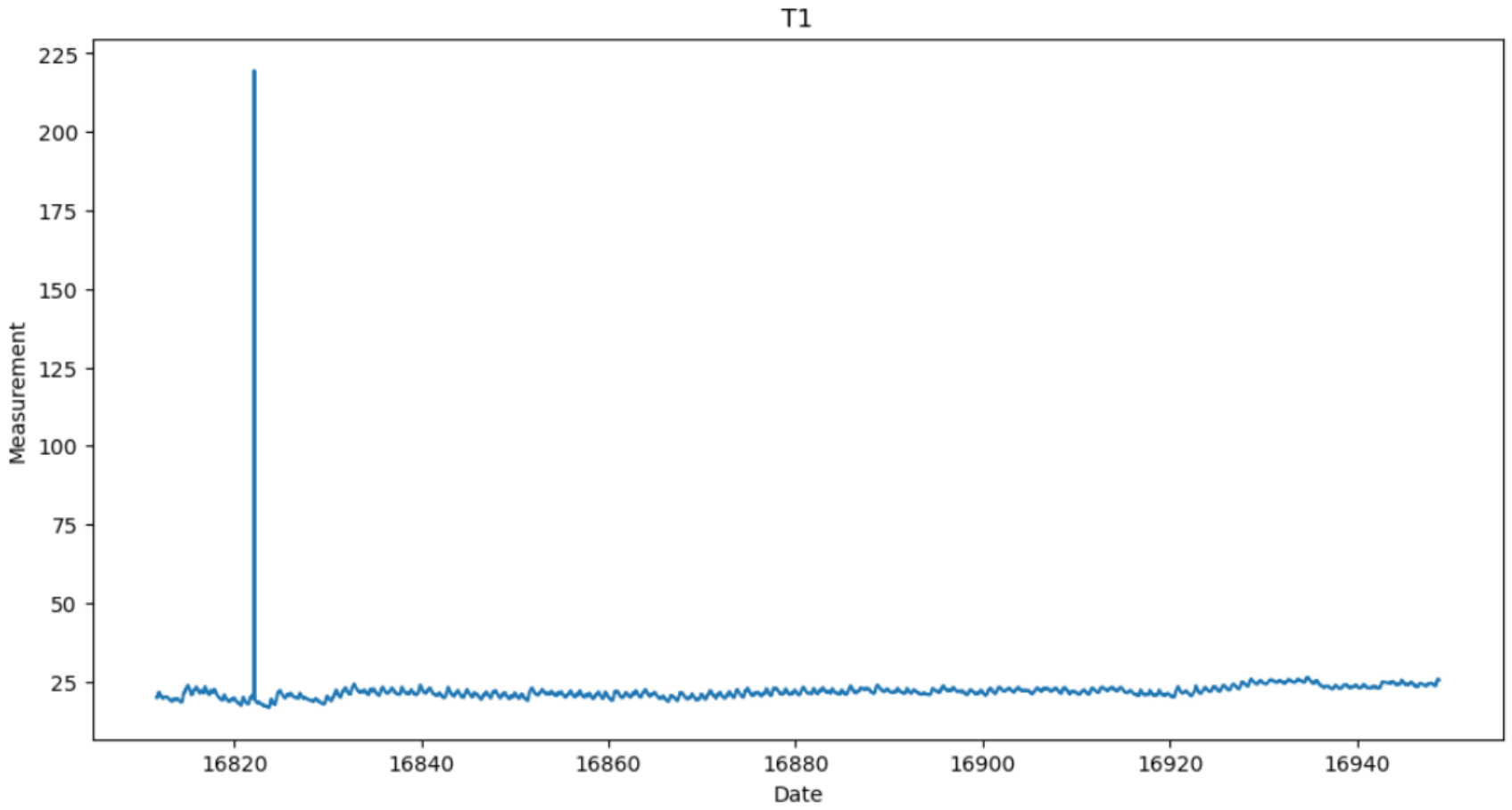


Correlation

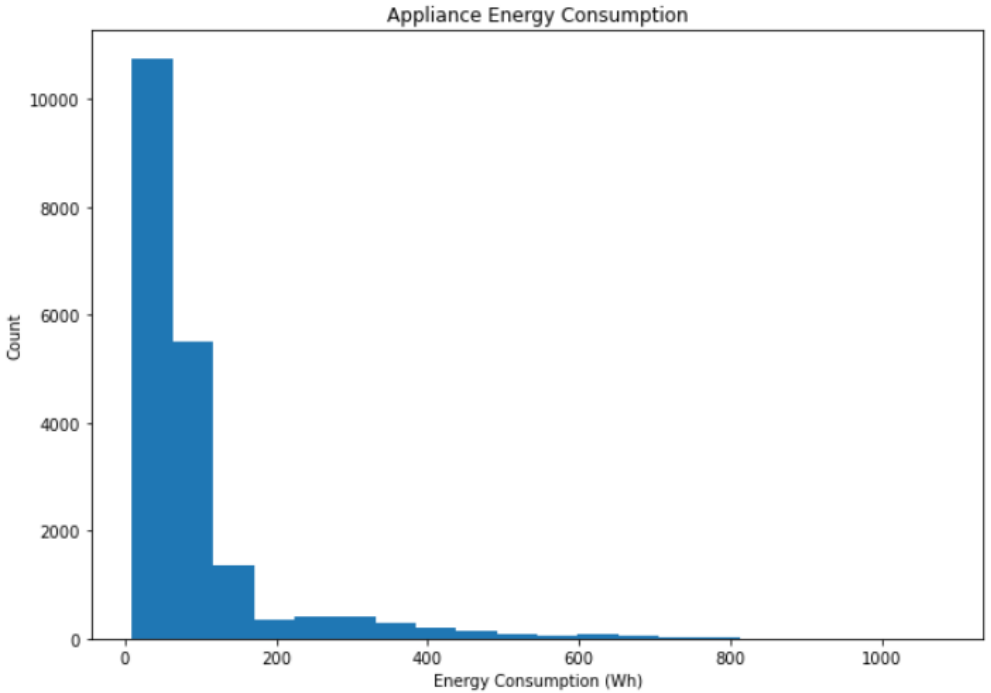


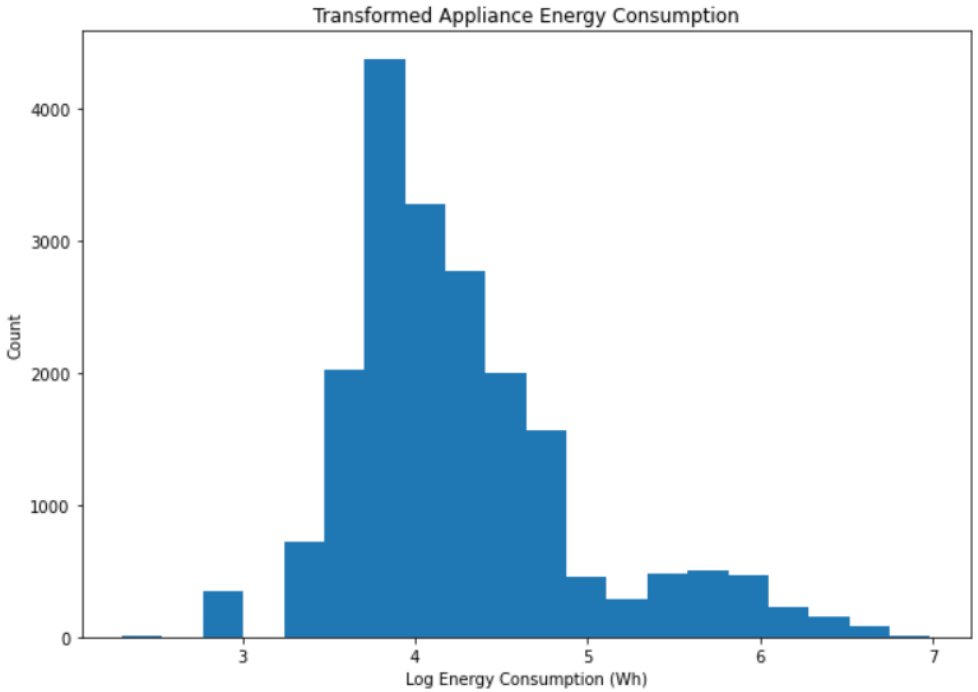
NATIONAL SECURITY INSTITUTE
VIRGINIA TECH

Outliers


T1
225
200
175
150
125
100
75
50
25
Measurement
16820
16840
16860
16880
16900
16920
16940
Date

Transformations



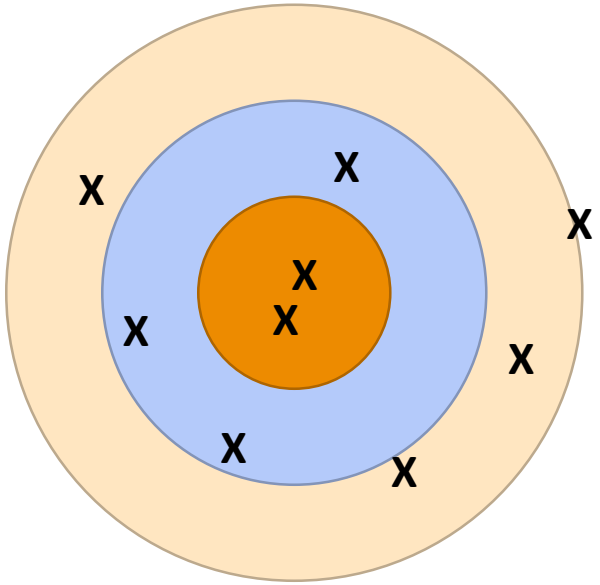


Modeling Considerations

Bias vs Variance

$$\text{Error} = \text{Bias}^2 + \text{Variance} + \text{Irreducible Error}$$

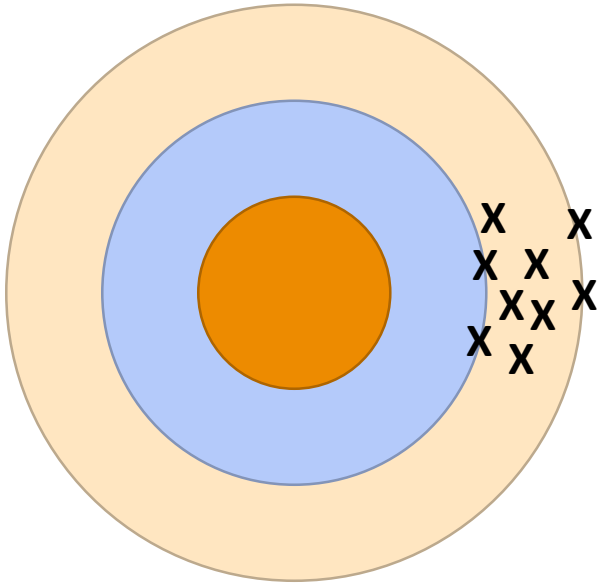
<u>Variance</u>



Bias vs Variance

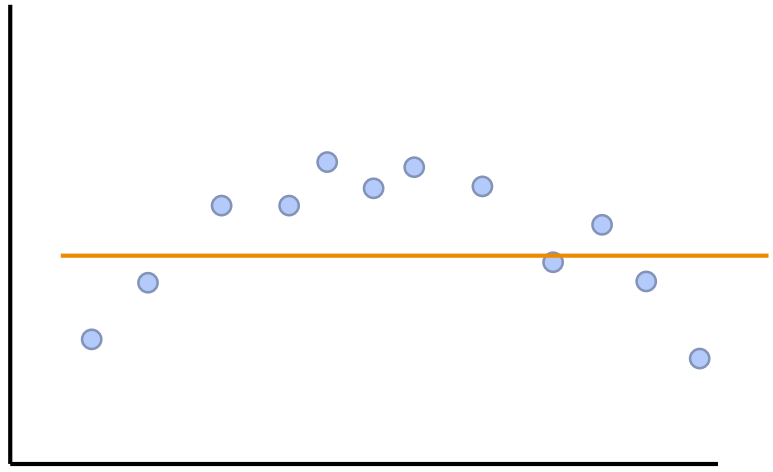
$$\text{Error} = \text{Bias}^2 + \text{Variance} + \text{Irreducible Error}$$

Bias



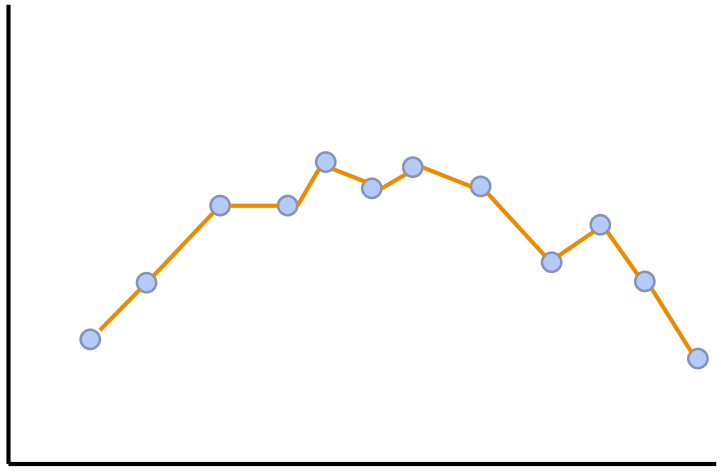
Underfitting vs Overfitting

Underfitting



High Bias

Overfitting

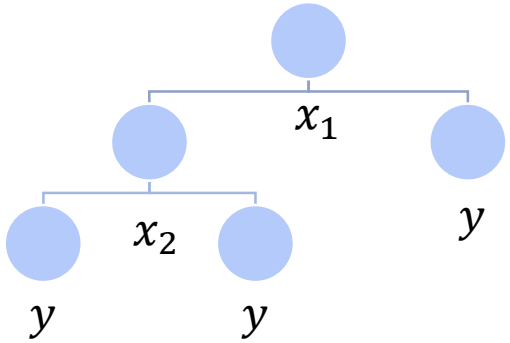


High Variance

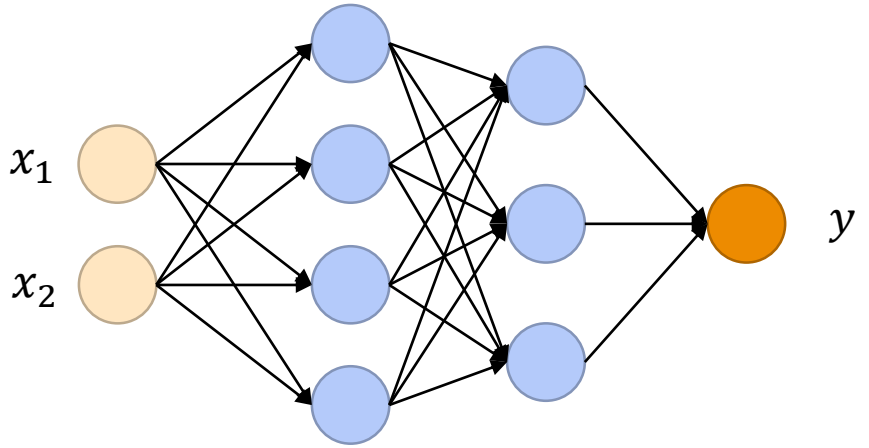
Model Complexity

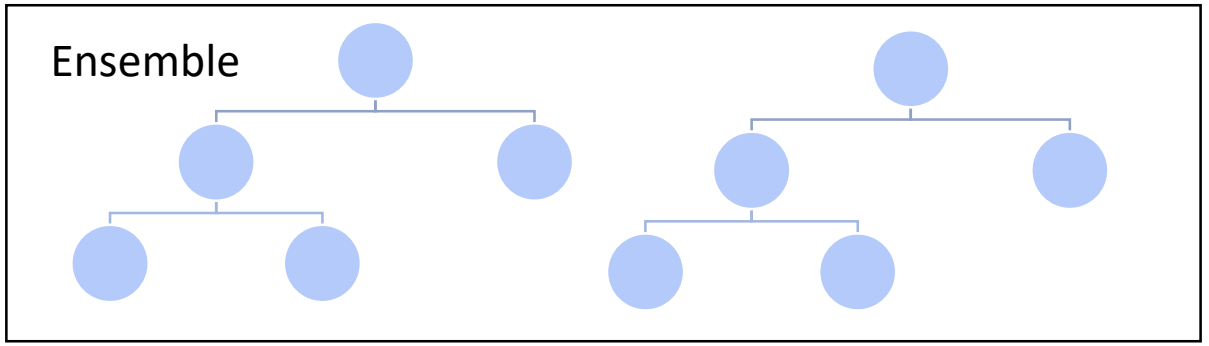
Low Complexity

$$y = wx + c$$

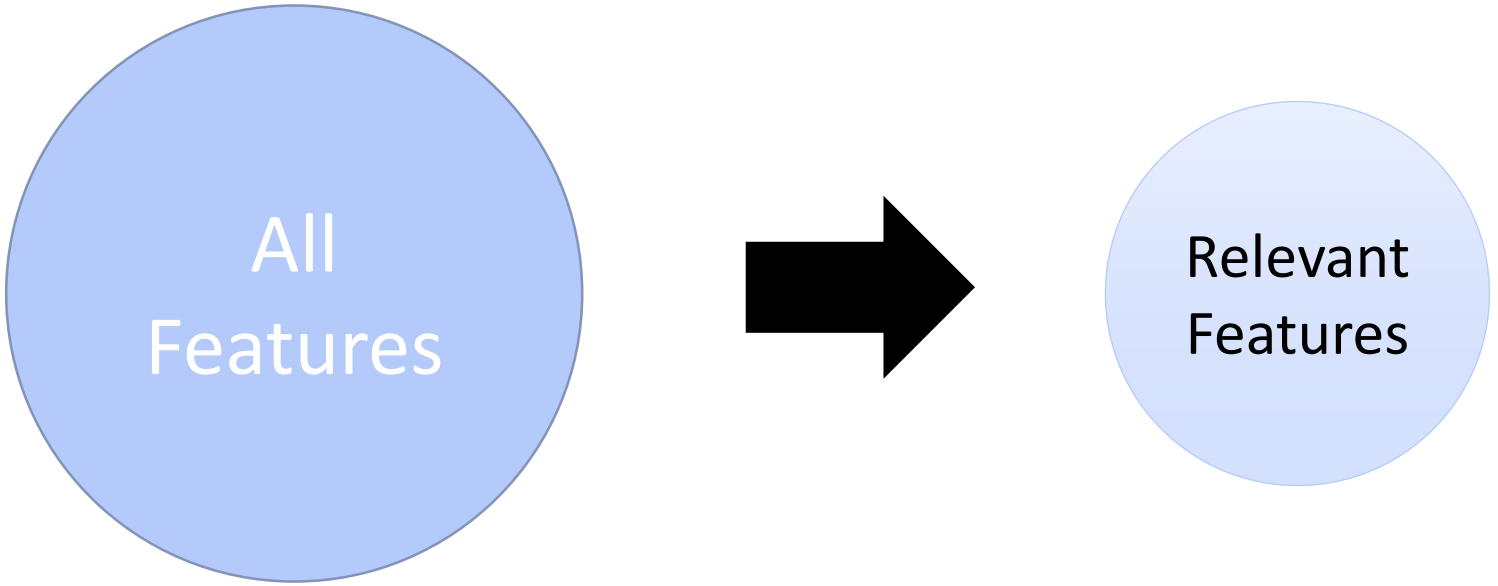


High Complexity



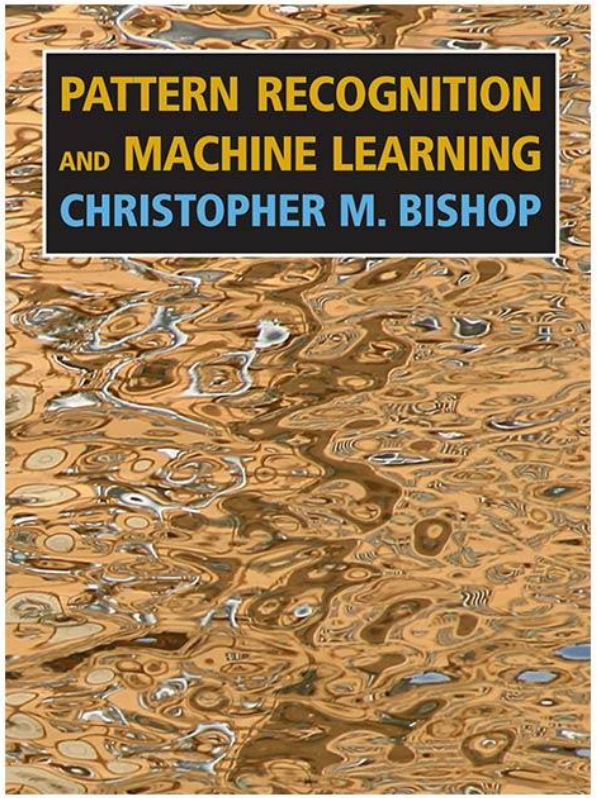


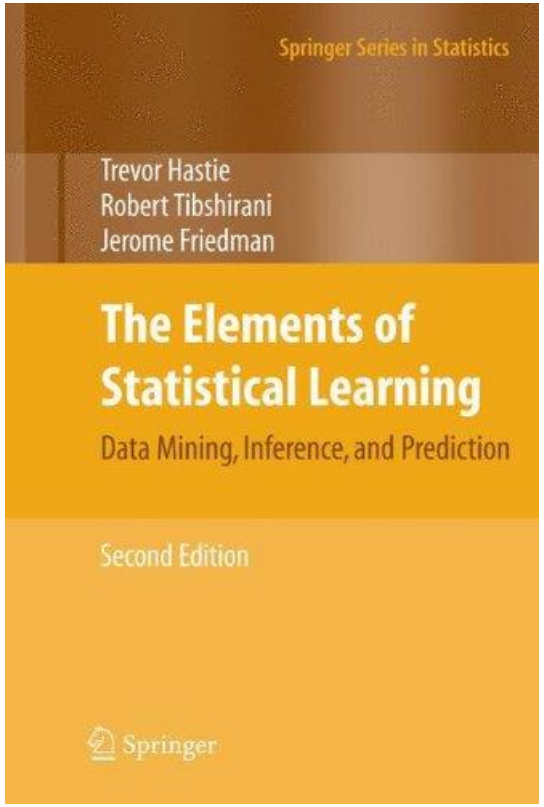
Feature Selection

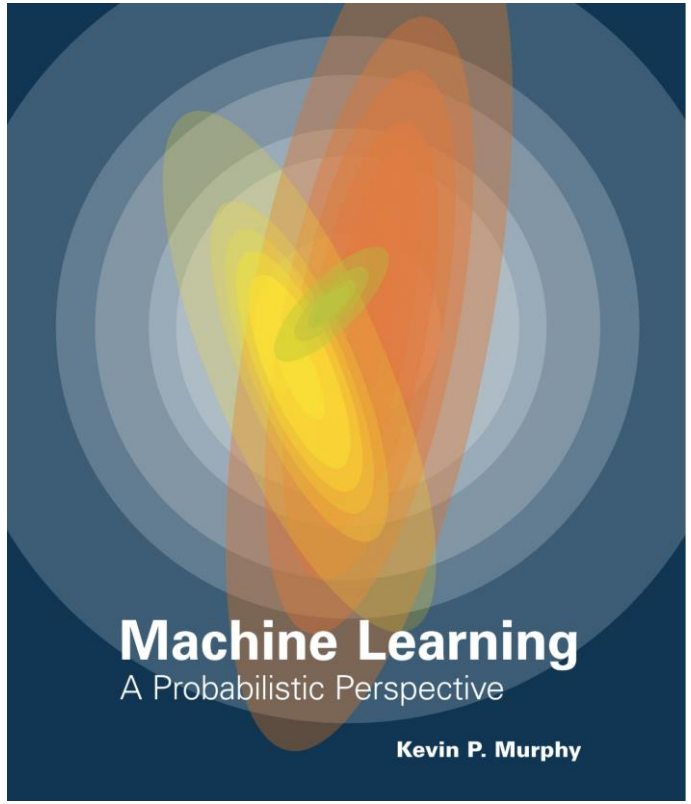


Resources

Books


PATTERN RECOGNITION
AND MACHINE LEARNING
CHRISTOPHER M. BISHOP


Springer Series in Statistics
Trevor Hastie
Robert Tibshirani
Jerome Friedman
The Elements of Statistical Learning
Data Mining, Inference, and Prediction
Second Edition
Springer


Machine Learning
A Probabilistic Perspective
Kevin P. Murphy

Data Repositories

https://archive.ics.uci.edu/ml/index.php

Questions?