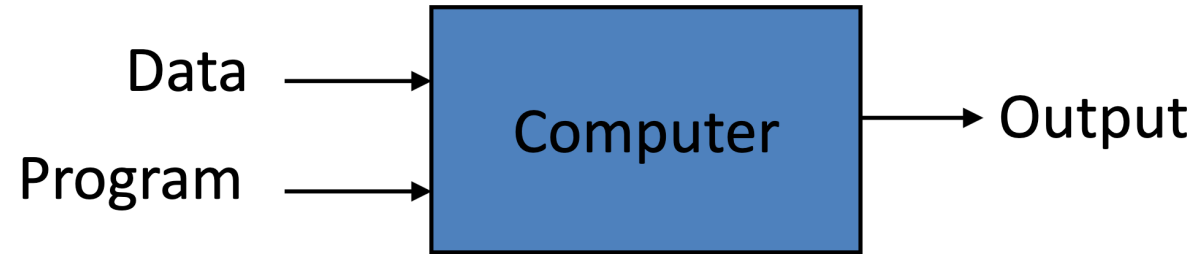


Machine Learning with scikit-learn

<https://github.com/nuitrcs/Python-scikit-learn-2023>

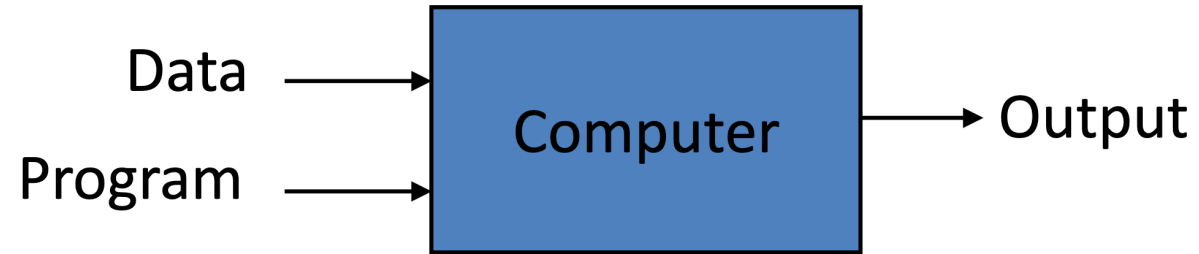
What is Machine Learning?

Traditional Programming

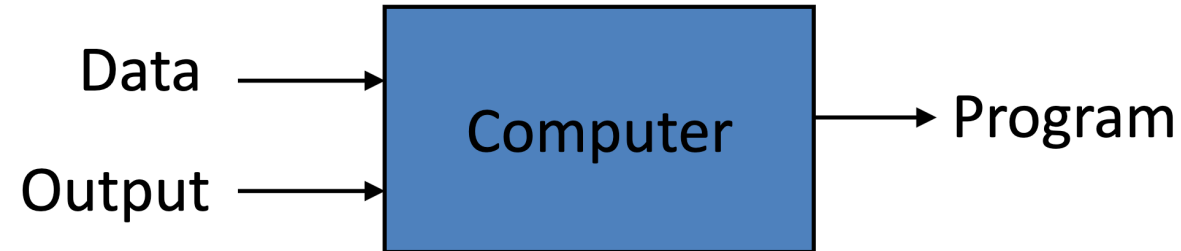


What is Machine Learning?

Traditional Programming



Machine Learning



When Do We Use Machine Learning?

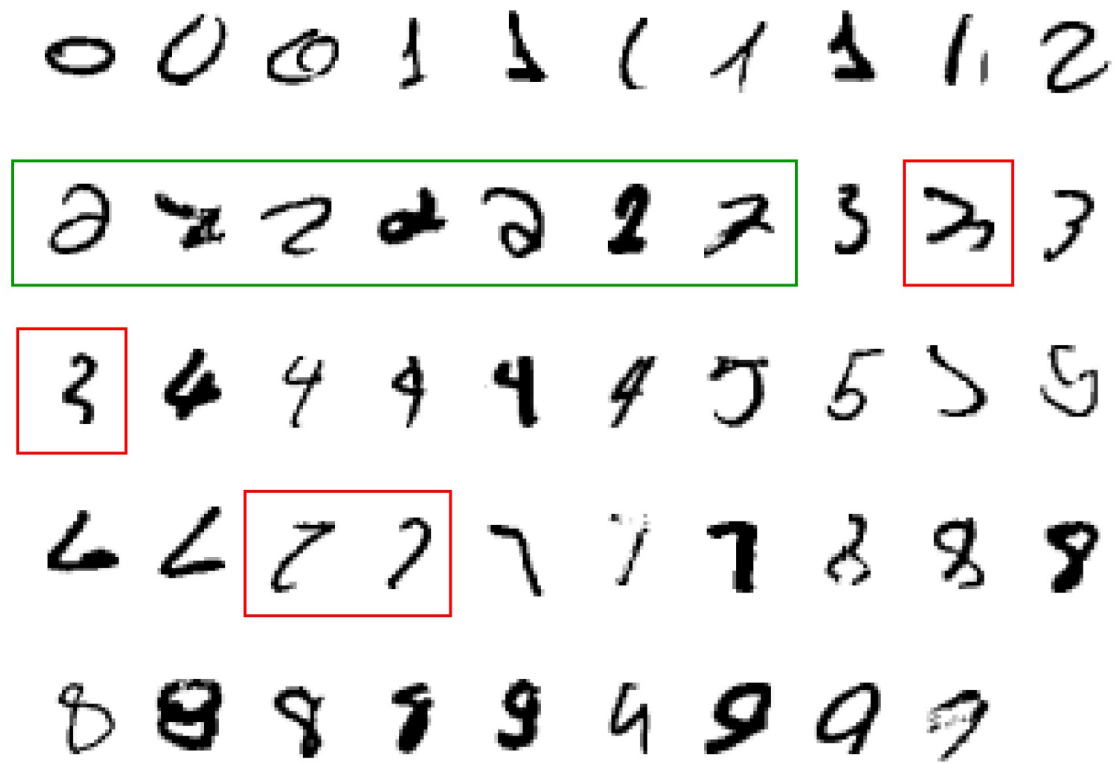
ML is used when:

- Human expertise does not exist (navigating on Mars)
- Humans can't explain their expertise (speech recognition)
- Models must be customized (personalized medicine)
- Models are based on huge amounts of data (genomics)

Learning isn't always useful:

There is no need to “learn” to calculate payroll

A classic example of a task requiring ML



What makes a “2” ?

Some more examples of ML tasks

Recognizing patterns:

- Facial identities or facial expressions
- Handwritten or spoken words
- Medical images

Generating patterns:

- Generating images or motion sequences

Recognizing anomalies:

- Unusual credit card transactions
- Unusual patterns of sensor readings in a nuclear power plant

Prediction:

- Future stock prices or currency exchange rates

Types of ML Problems

	<i>Supervised Learning</i>	<i>Unsupervised Learning</i>
<i>Discrete</i>	classification or categorization	clustering
<i>Continuous</i>	regression	dimensionality reduction

Supervised Learning

When you already know the “answers”

And want to predict future “answers”

- Learn a discrete **Classification**
- Learn a continuous **Regression**

Unsupervised Learning

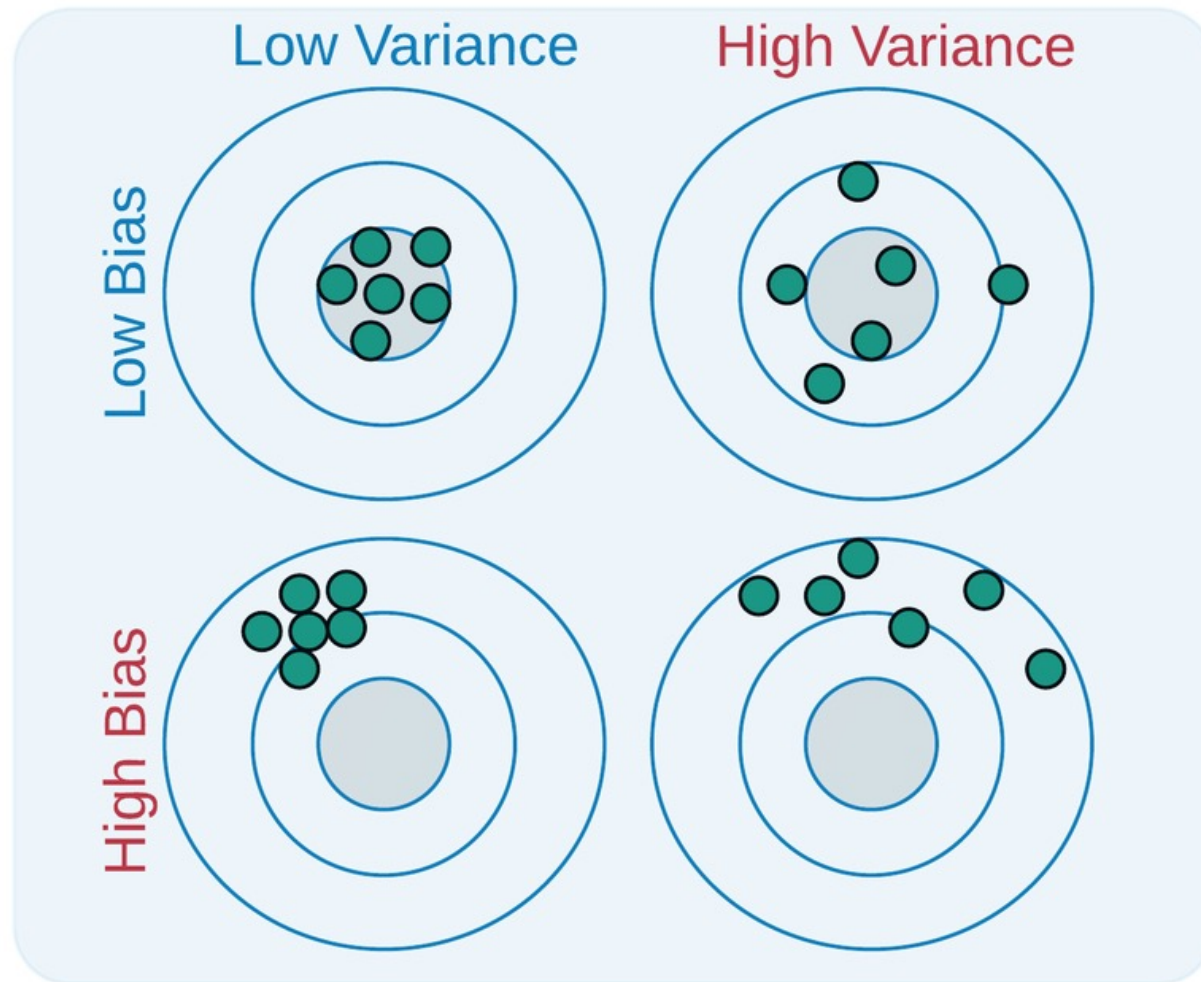
When you don't know any labels

But want to predict / learn useful labels

- Learn a discrete **Clustering**
- Learn a continuous **Dimension Reduction**

Model Optimization

Types of ML Errors



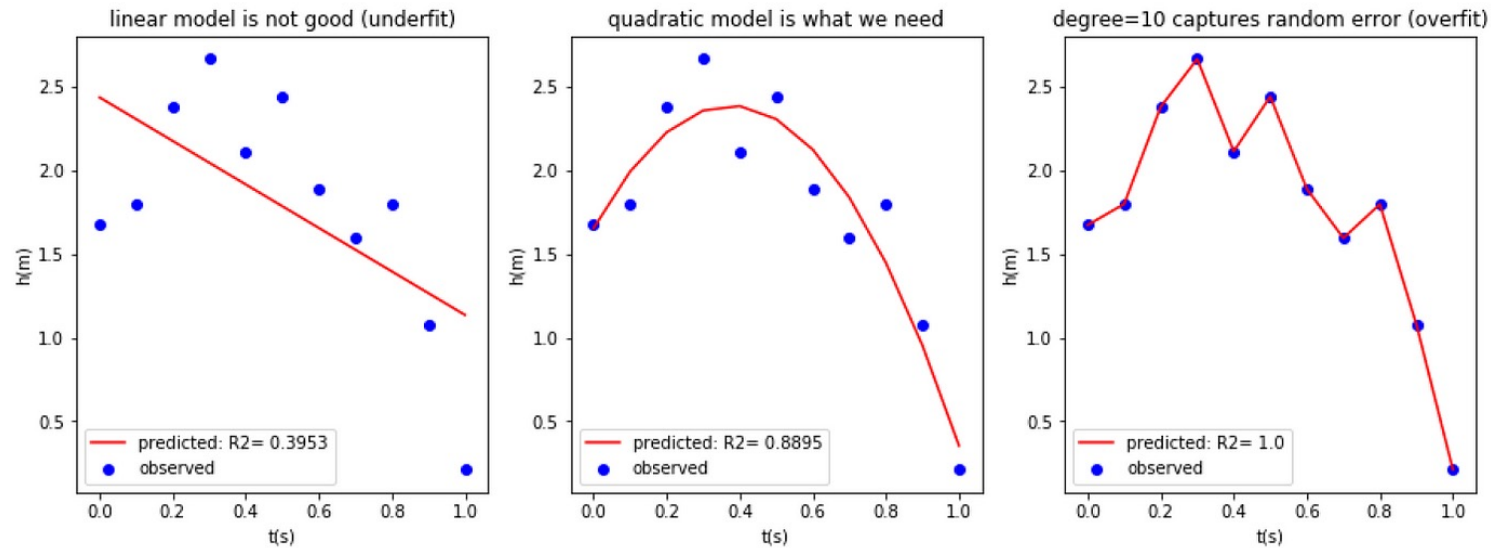
Bias:

Systematic prejudice in the model
Simple model = High bias

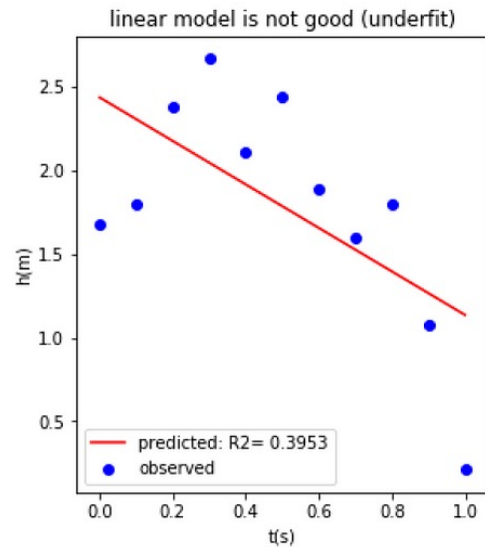
Variance:

Change in the model's prediction,
when the dataset is changed a
little bit
Complex model = High variance

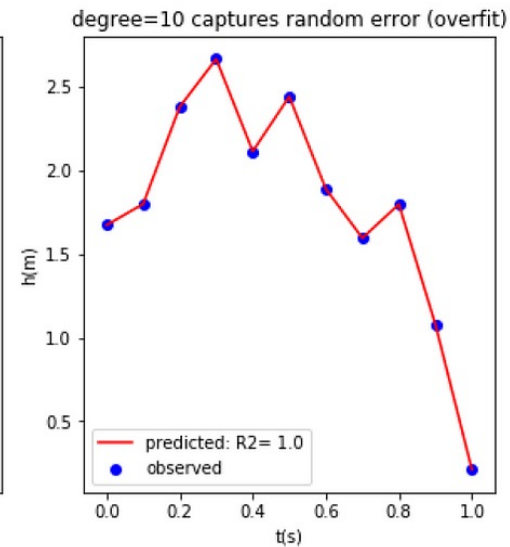
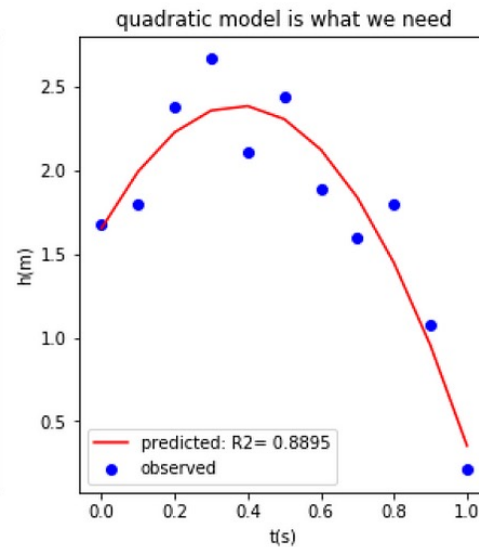
Model Optimization *Bias-variance trade-off*



Model Optimization *Bias-variance trade-off*



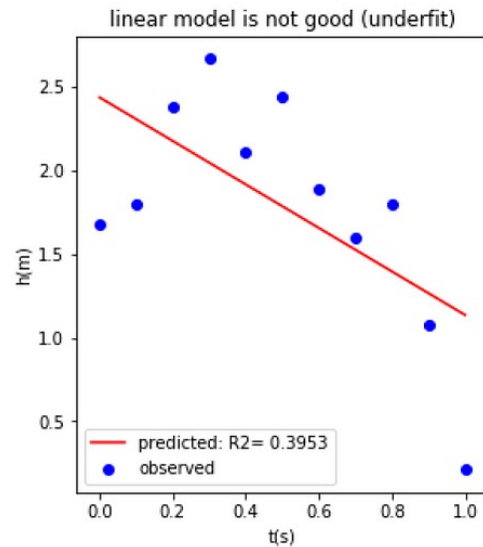
High Bias
Low Variance



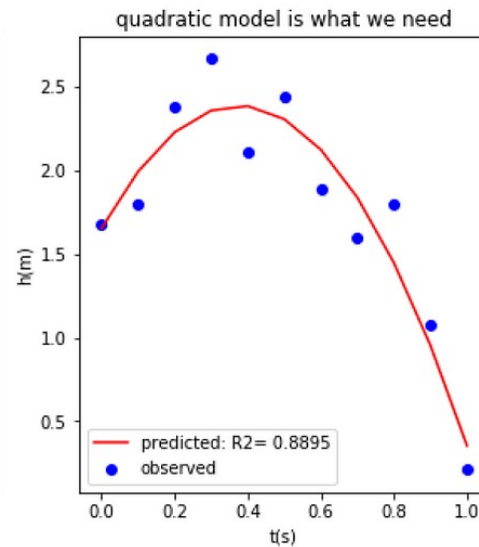
Low Bias
High Variance

Model Optimization *Bias-variance trade-off*

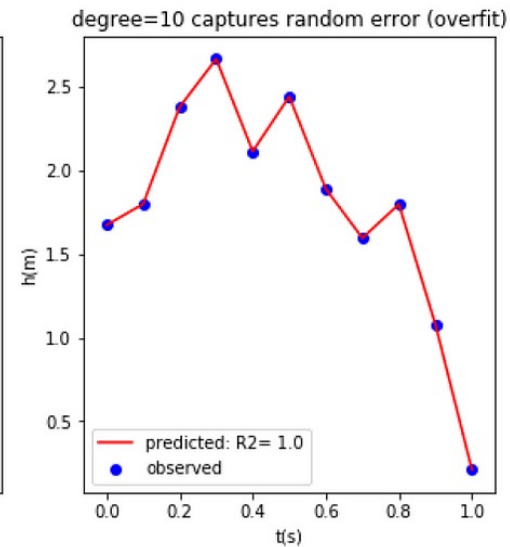
Under-fitting



High Bias
Low Variance



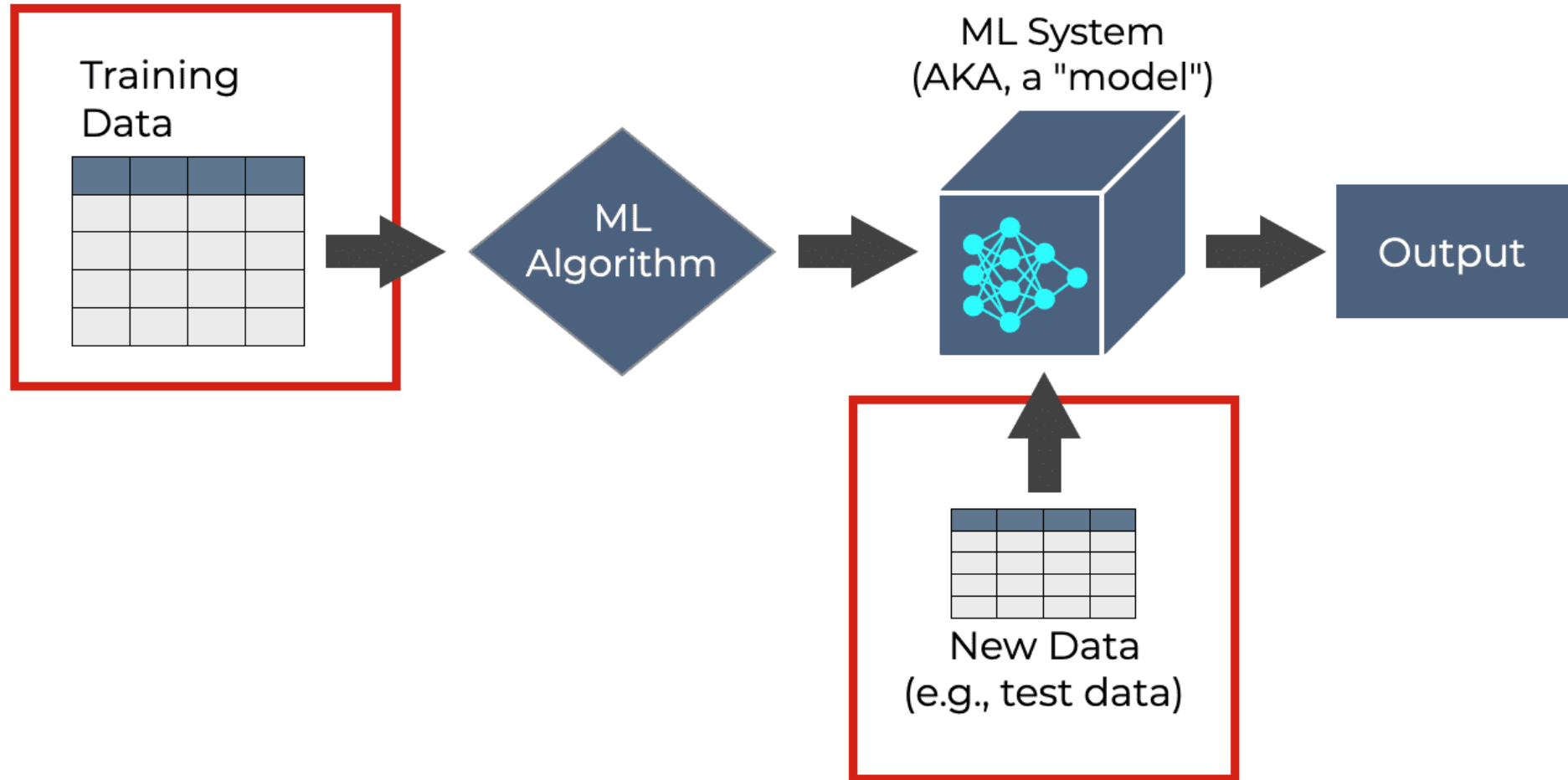
Over-fitting



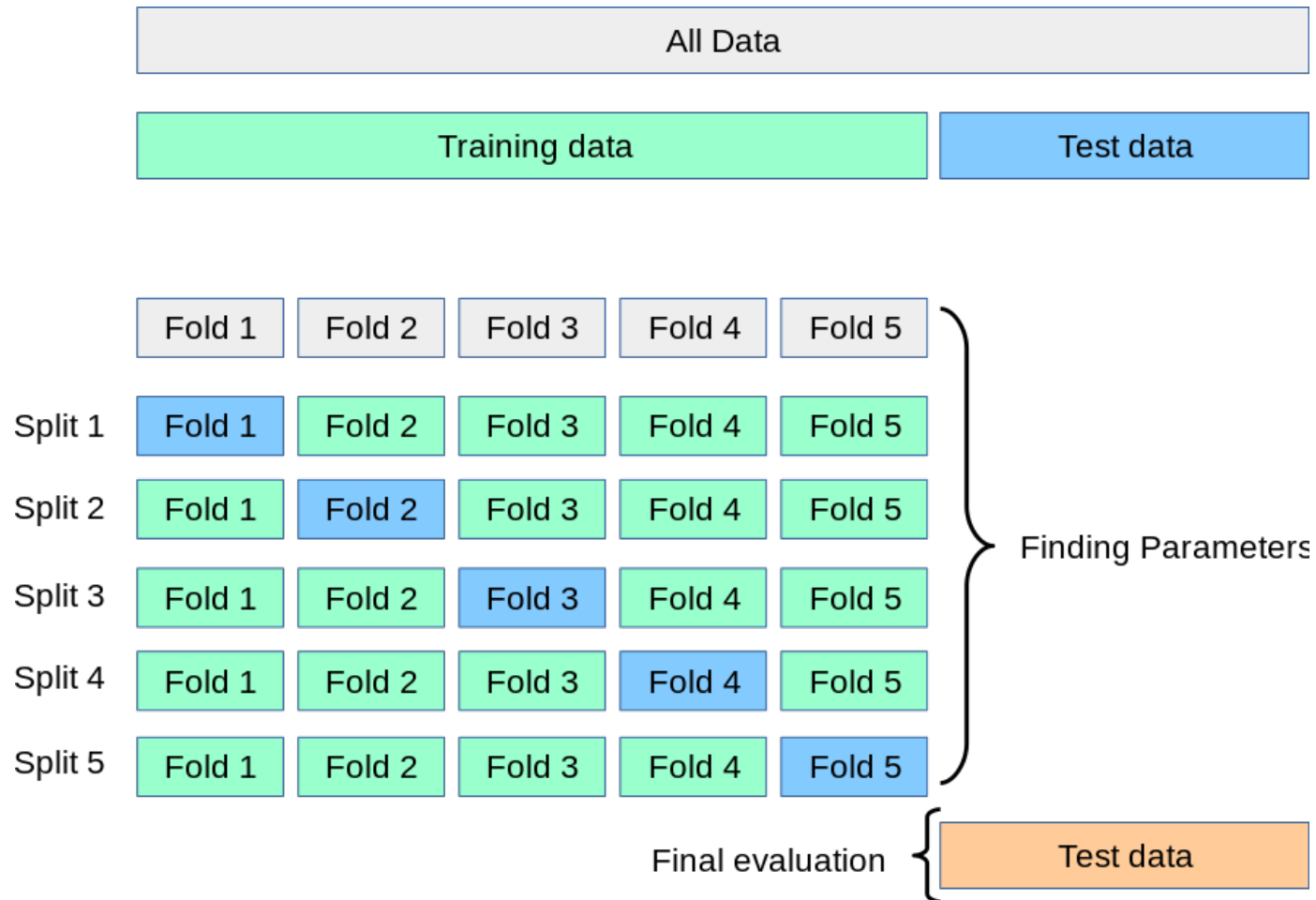
Low Bias
High Variance

Model Validation

Train – Test split



Model Validation *k-fold cross-validation*



ML workflow



1. Should I use ML on this problem?

Is there a pattern to detect?
Can I solve it analytically?
Do I have data?



2. Gather and organize data.

Preprocessing, cleaning, visualizing.



3. Establishing a baseline.



4. Choosing a model, loss, regularization, ...



5. Optimization (could be simple, could be a PhD...).



6. Hyperparameter search.



7. Analyze performance & mistakes and iterate back to step 4 (or 2).