Introduction to Machine Learning for Public Policy

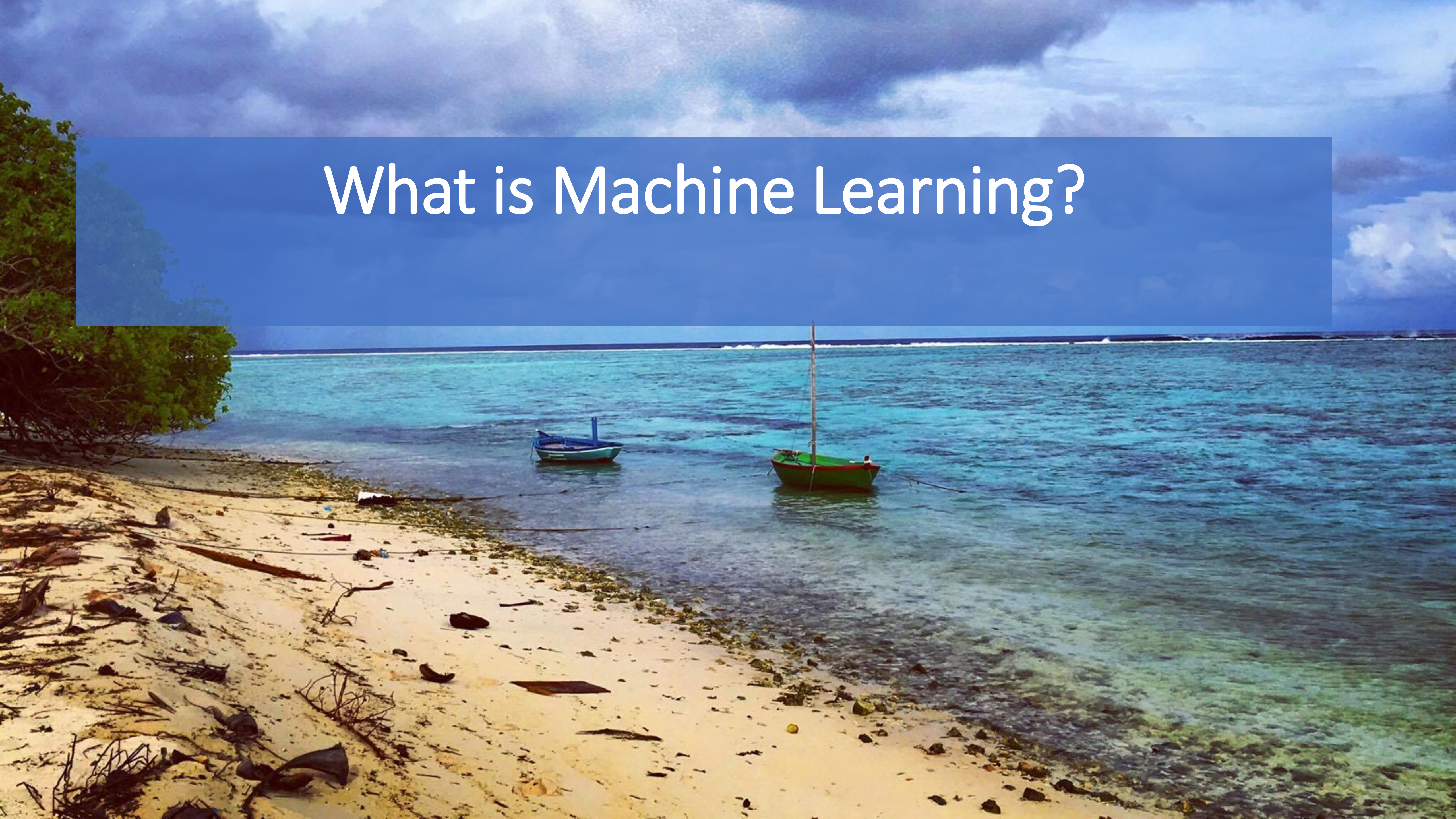
Jonathan Hersh, PhD (Chapman Argyros School of Business)

12/2/2020

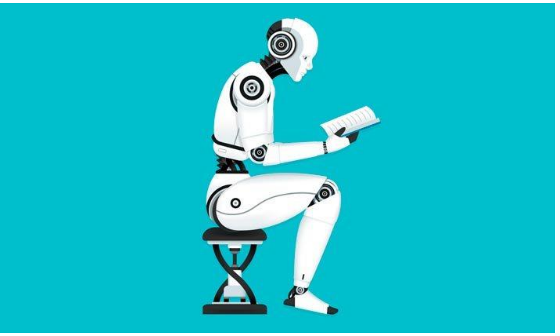
Outline

1. **What is Machine Learning?**
   - Machine learning versus econometrics
2. **Why Machine Learning for Public Policy**
   - Big data requires it
   - Non-linear relationships
   - Better forecasts/econometrics
   - Anomaly detection
3. **Some Basic Machine Learning Concepts**
   - Supervised vs Unsupervised learning
   - Testing/Training Sets
   - Bias-Variance Tradeoff

What is Machine Learning?



Public Conception of Machine Learning

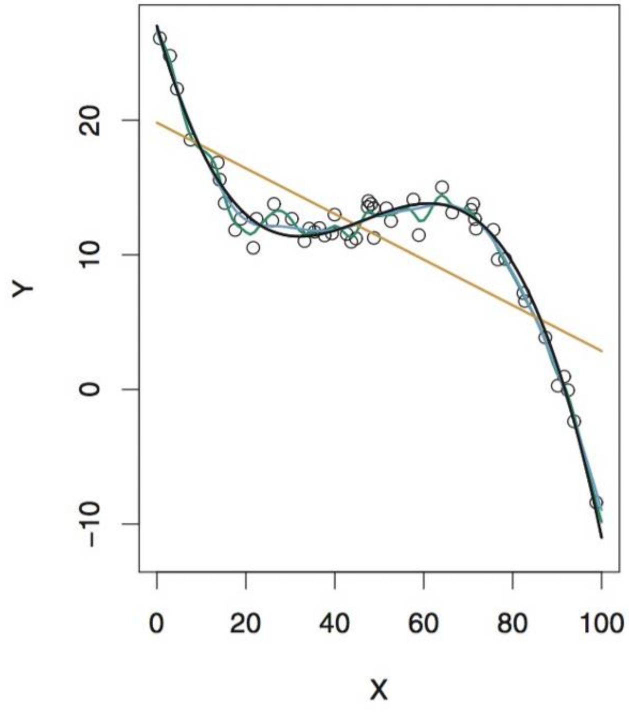


Reality (90% of the time)

Target or Output

Input data

$$\hat{y} = \hat{f}(x)$$

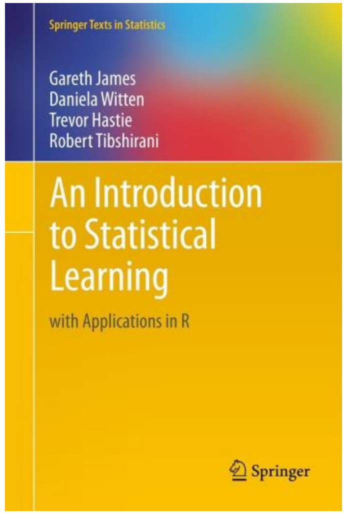


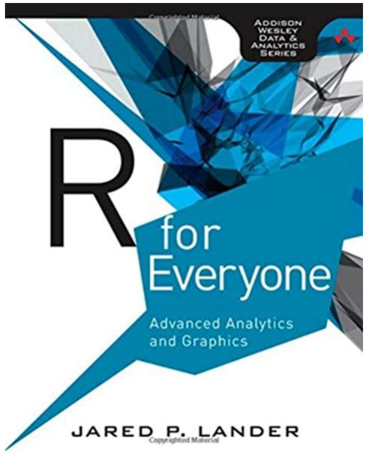
Machine Learning Versus Econometrics

- **Machine Learning**
  - Developed to solve problems in computer science
  - Prediction/classification
  - Desire: goodness of fit
  - Huge Datasets! (Terabytes) Thousands of variables!
  - Whatever works

- **Econometrics**
  - Developed to solve problems in economics
  - Explicitly testing a theory
  - "Statistical significance" more important than model fit
  - Small datasets Few dozen variables
  - "It works in practice, but what about theory?"

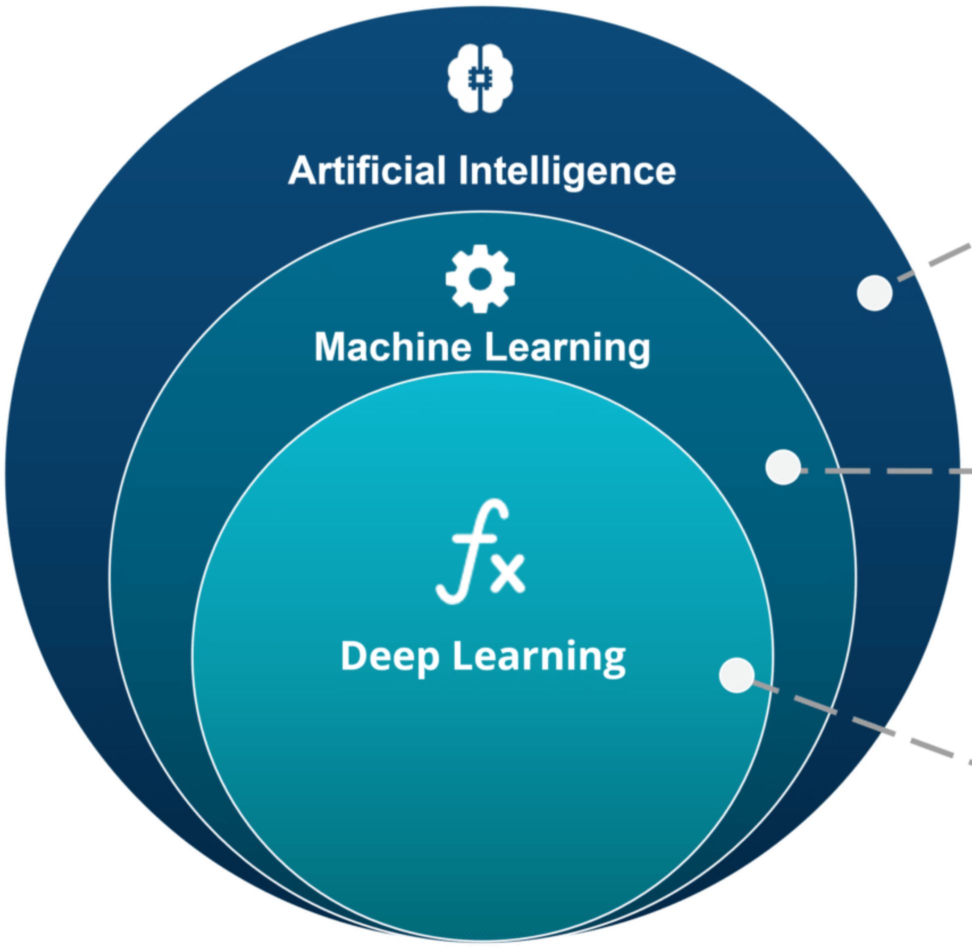
Today – Brief Introduction to Machine Learning





- **Cross-Validation** [Chapter 2 ISLR]
- **Ridge Regression** [Chapter 6 ISLR]
- **Lasso Regression** [Chapter 6 ISLR]
- **Decision Trees** [Chapter 8 ISLR]
- **Introduction to R** [R for Everyone]

Machine Learning Versus Artificial Intelligence



ARTIFICIAL INTELLIGENCE
A technique which enables machines to mimic human behaviour

MACHINE LEARNING
Subset of AI technique which use statistical methods to enable machines to improve with experience

DEEP LEARNING
Subset of ML which make the computation of multi-layer neural network feasible

Why Machine Learning?

Arguments for Using Machine Learning for Public Policy

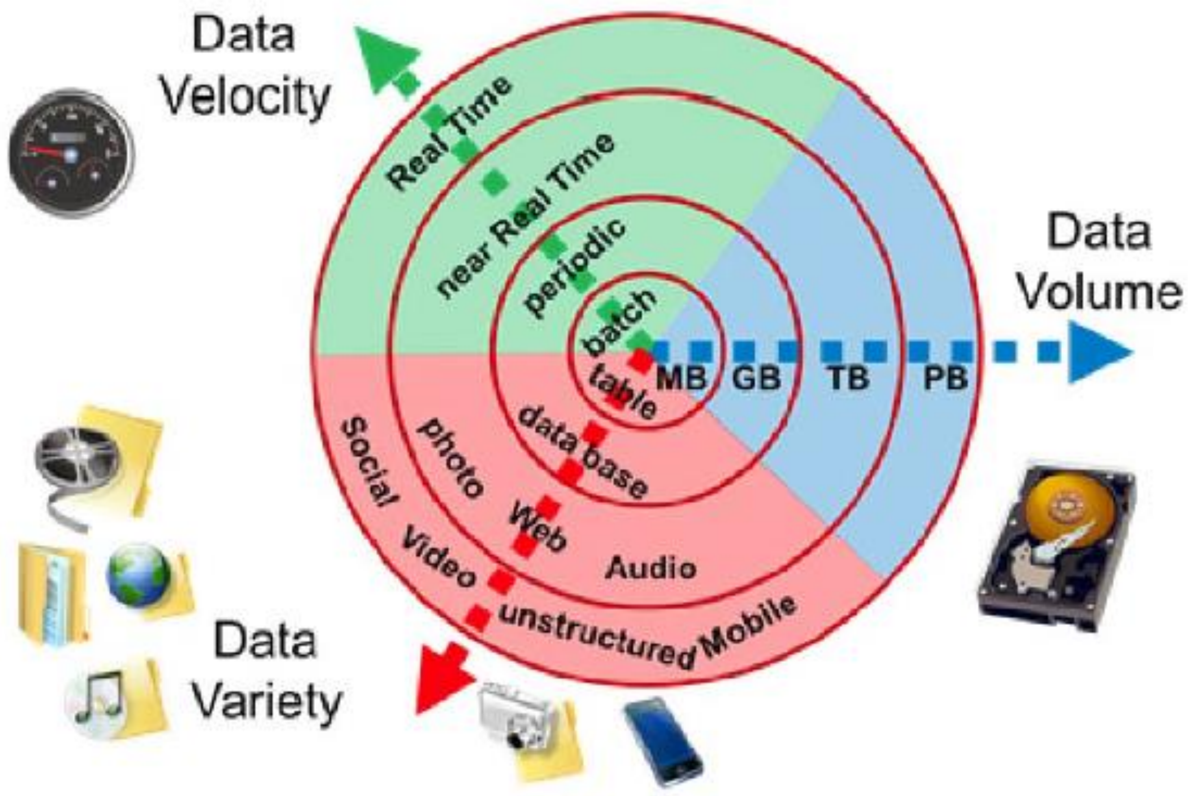
1. **Needed ML Big Data (models with 100+ variables)**
   - “Unstructured” data e.g. satellite imagery, text

2. **Can learn non-linear relationships**

3. **Better forecasts / econometrics**

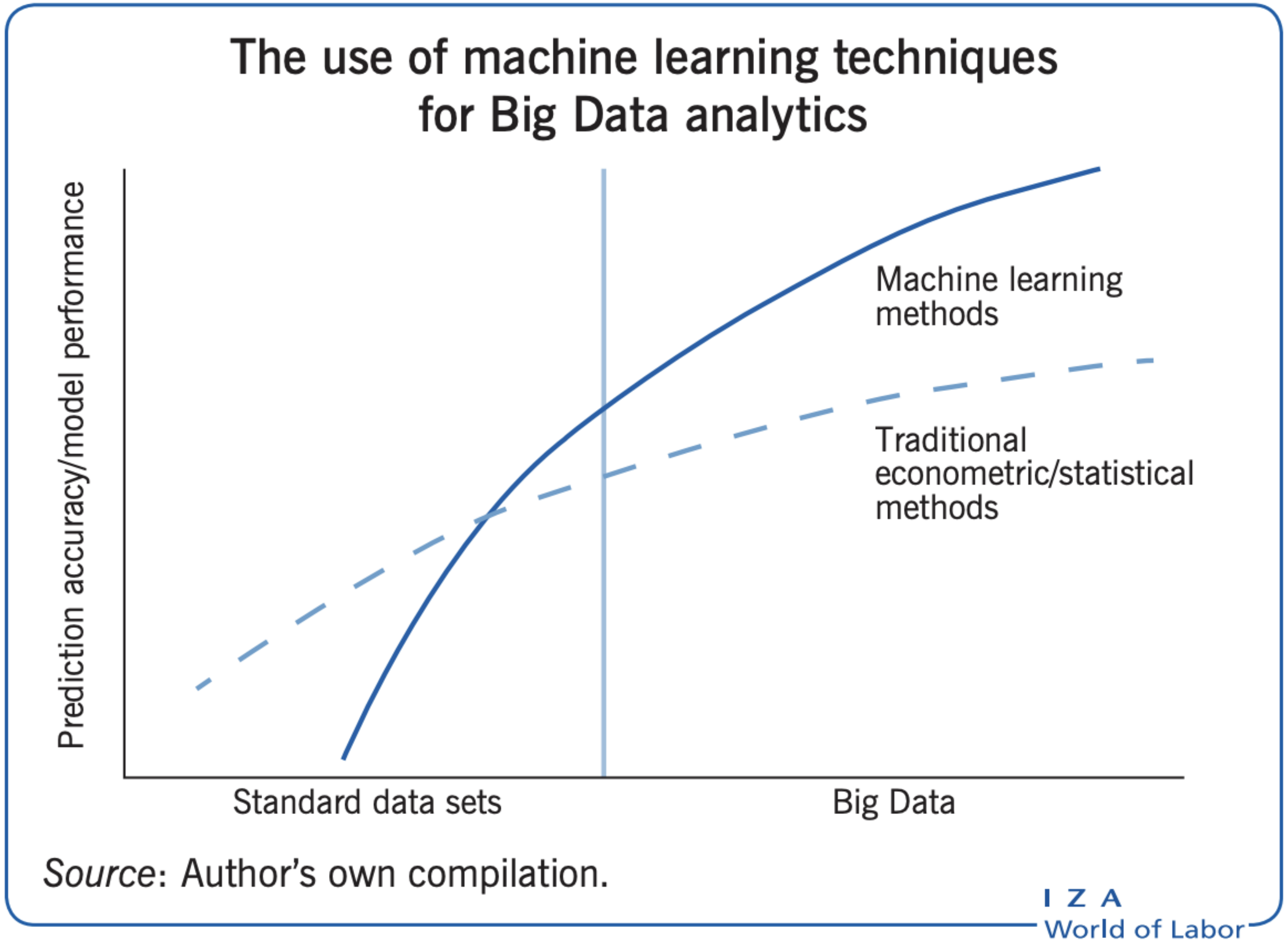
4. **Anomaly detection (for fraud detection)**

What is Big Data?



- **Big data is Data with Three "v's"**
  - High <u>volume</u>
  - High <u>variety</u>
  - High <u>velocity</u>

Why ML? (Big Data Needs Machine Learning)



- Machine Learning models continue to improve given more data (both # of variables and # of observations)
- Bigger datasets: bigger gain from machine learning vs econometrics

Why ML? (To Use Satellite Imagery ”Big Data”)

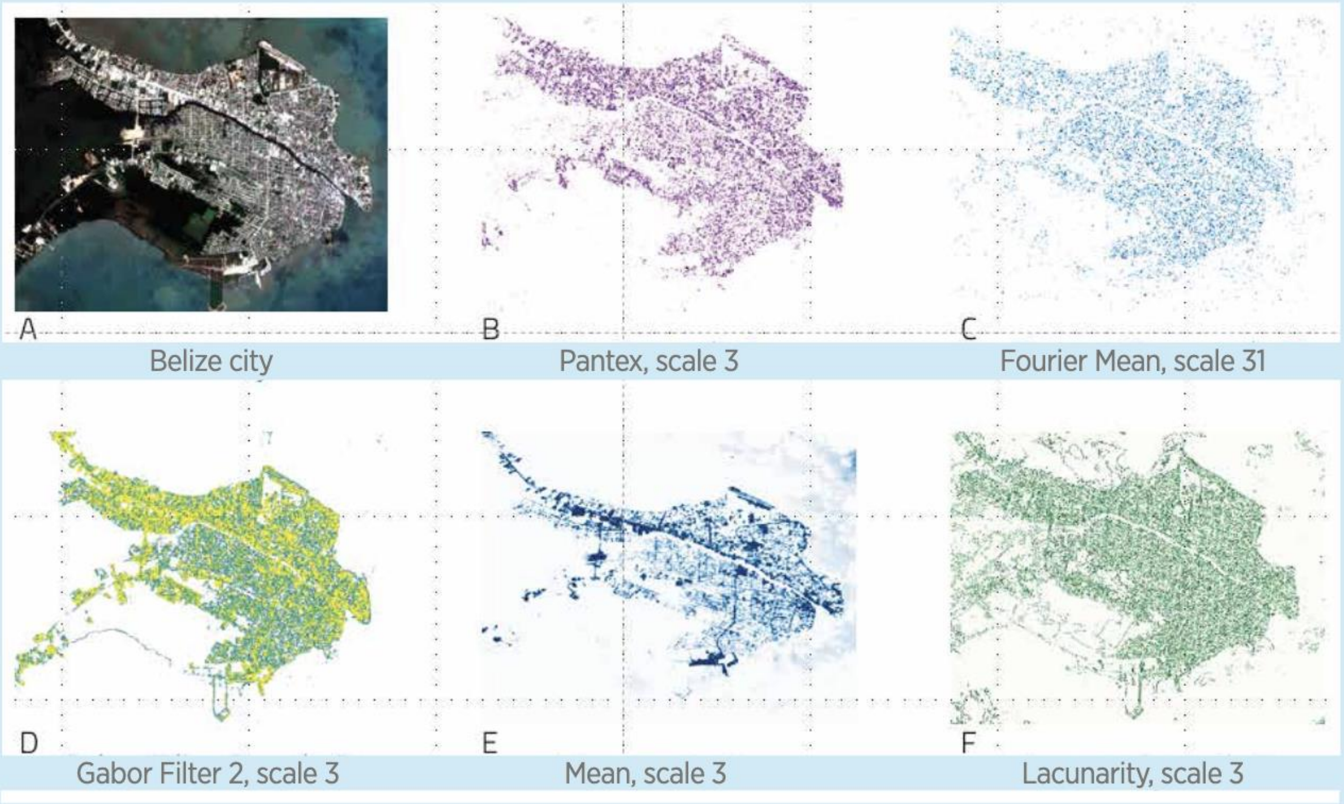
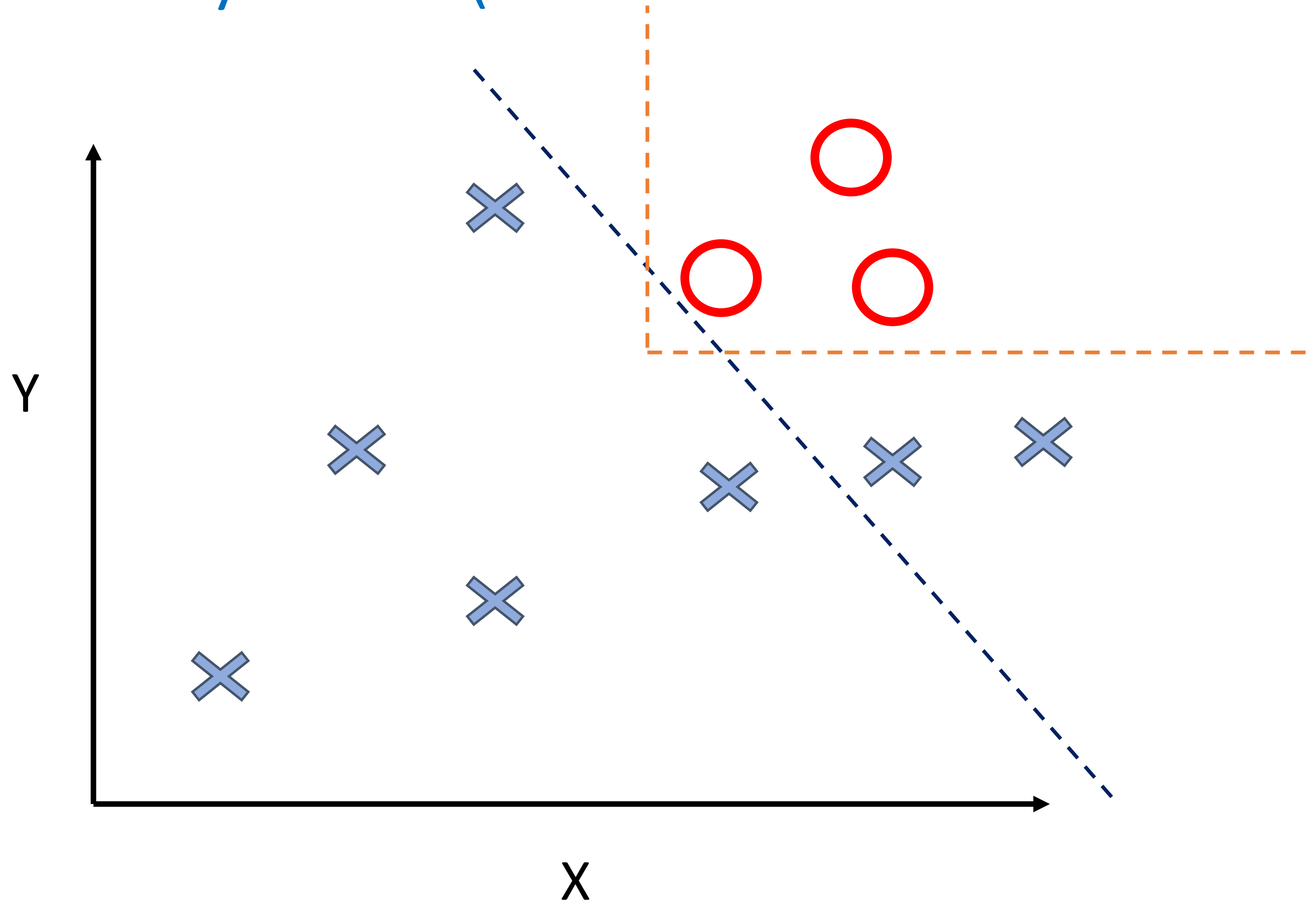


***Figure 1** Belize City contextual features*
*Displayed in **A** is the true color (Blue, Green, and Red) Sentinel 2 image for Belize City, **B** is the Pantex measure derived for this area at scale 3 (30m), **C** is the Mean Fourier transform at scale 31 (310m), D) is the second Gabor Filter at scale 3 (30m), **E** is the Mean brightness at scale 3 (30m), and **F** is the Lacunarity measure at scale 3 (30m).*

- Satellite Imagery variables too high dimensional for traditional econometric models

- Mapping Poverty in Belize Using Satellite Imagery https://publications.iadb.org/publications/english/document/Mapping-Income-Poverty-in-Belize-Using-Satellite-Features-and-Machine-Learning.pdf

Why ML? (Can learn Nonlinear relationships)



- **Example: classify "O"s separate from X's**

Econometrics: $y = X * \beta$

Machine Learning: regression tree

Model	Accuracy
Econometrics	80%
Machine Learning	100%

Why ML? (Better Forecasts For Fiscal Crises)

Predicting Fiscal Crises: A Machine Learning Approach

Klaus-Peter Hellwig[1]

International Monetary Fund, Asia Pacific Department

This version: July 27, 2020

Abstract: This paper assesses the ability of econometric and machine learning techniques to predict fiscal crises out of sample. We show that the standard econometric approach used in policy applications cannot outperform a heuristic rule of thumb derived from unconditional historical averages. Elastic net and tree ensemble methods (random forest, gradient boosted trees) deliver significant improvements in accuracy. Performance of machine learning techniques improves, particularly for developing countries, when expanding the set of potential predictors from a small set, preselected manually from the literature, to a large set (748 variables) and relying on algorithmic variable selection techniques. There is considerable agreement across learning algorithms in the set of selected predictors: Results confirm the importance of external sector stock and flow variables found in the literature but also point to demographics and the quality of governance as important predictors of fiscal crises. Fiscal variables appear to have less predictive value, and public debt matters only to the extent that it is owed to external creditors.

Why ML? (Better Forecasts of Inflation)

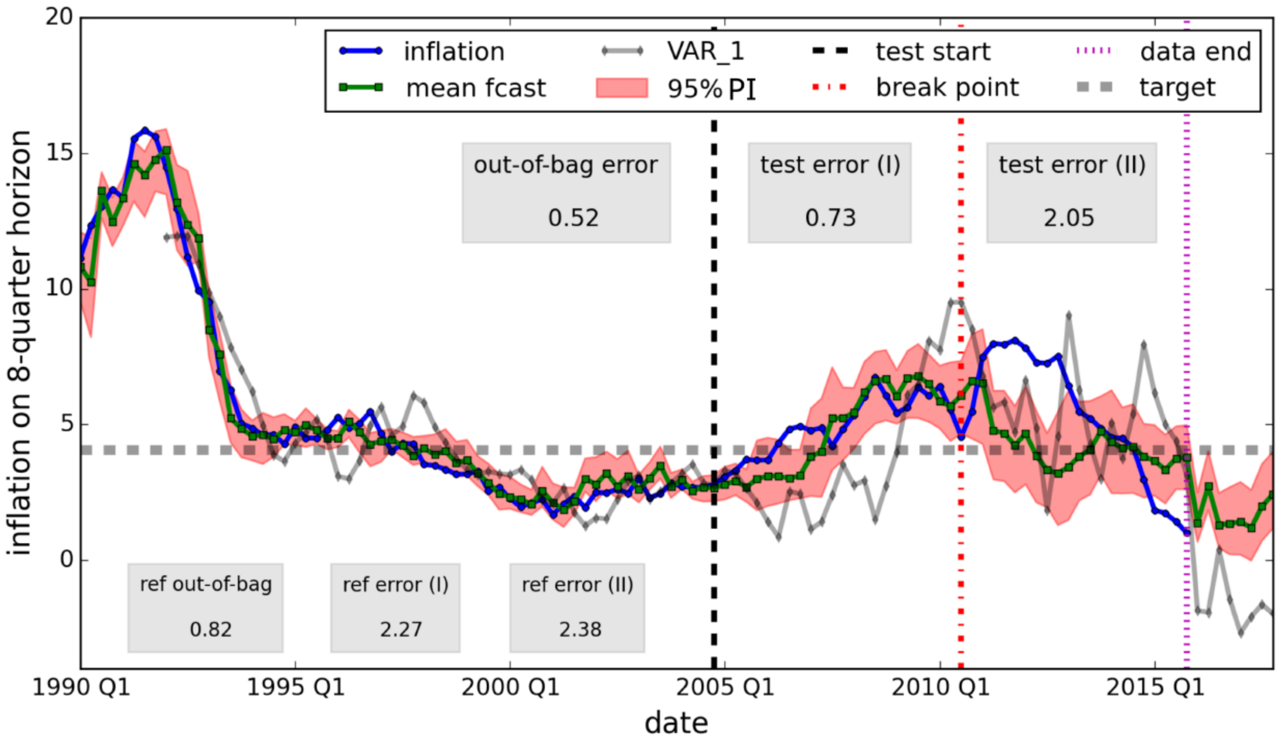
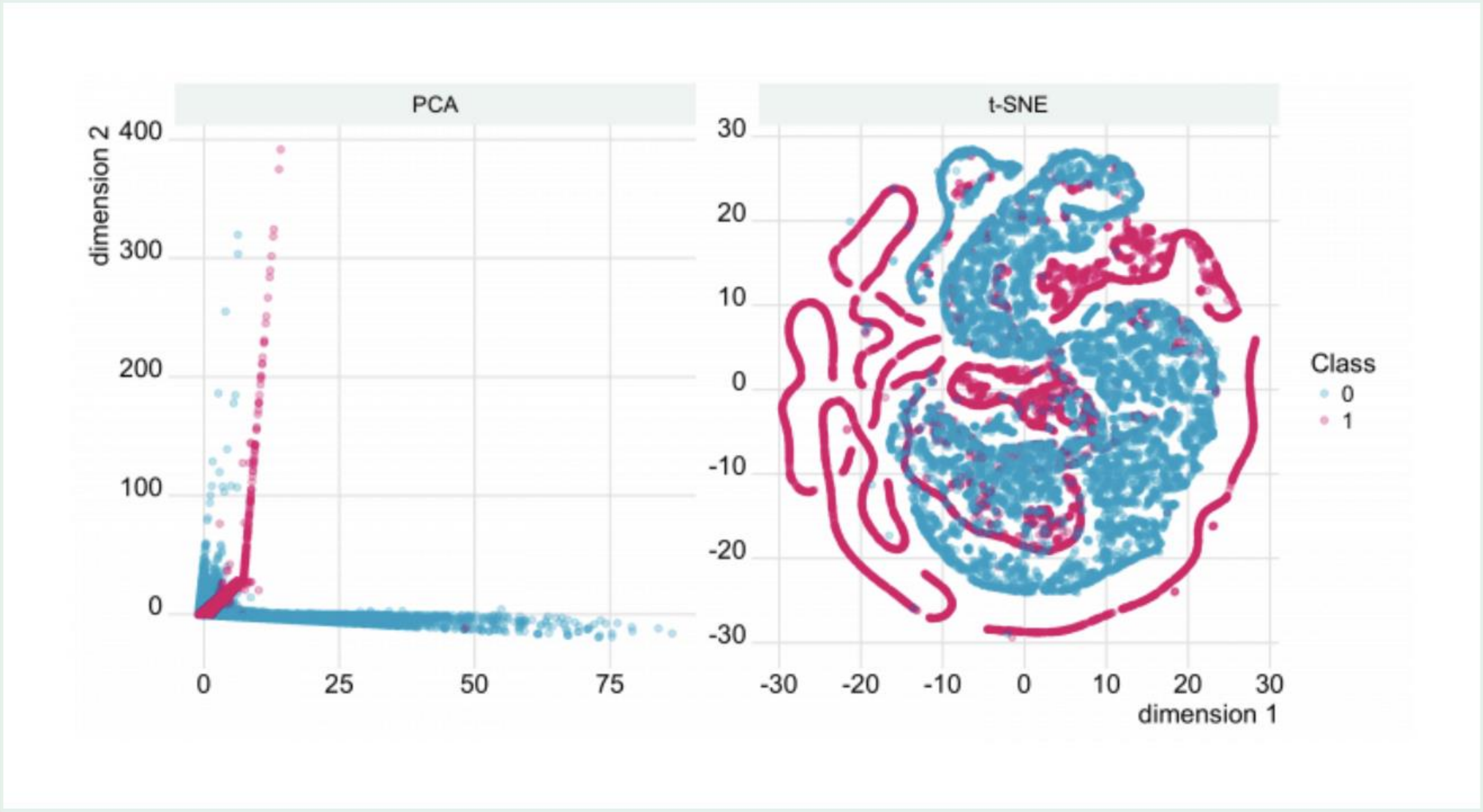


Figure 16: Averaged bootstrapped SVM-dFFANN projections (green line) for two-year changes in CPI (blue line). The shaded band indicates the 95% prediction interval (PI) across bootstrapped models. The vertical dashed lines separates the initial training, pre-crisis, post-crisis and post-data periods, respectively. Model and VAR_1 reference errors are given in the boxes. Sources: BoE, ONS, BIS, World Bank and authors' calculations.

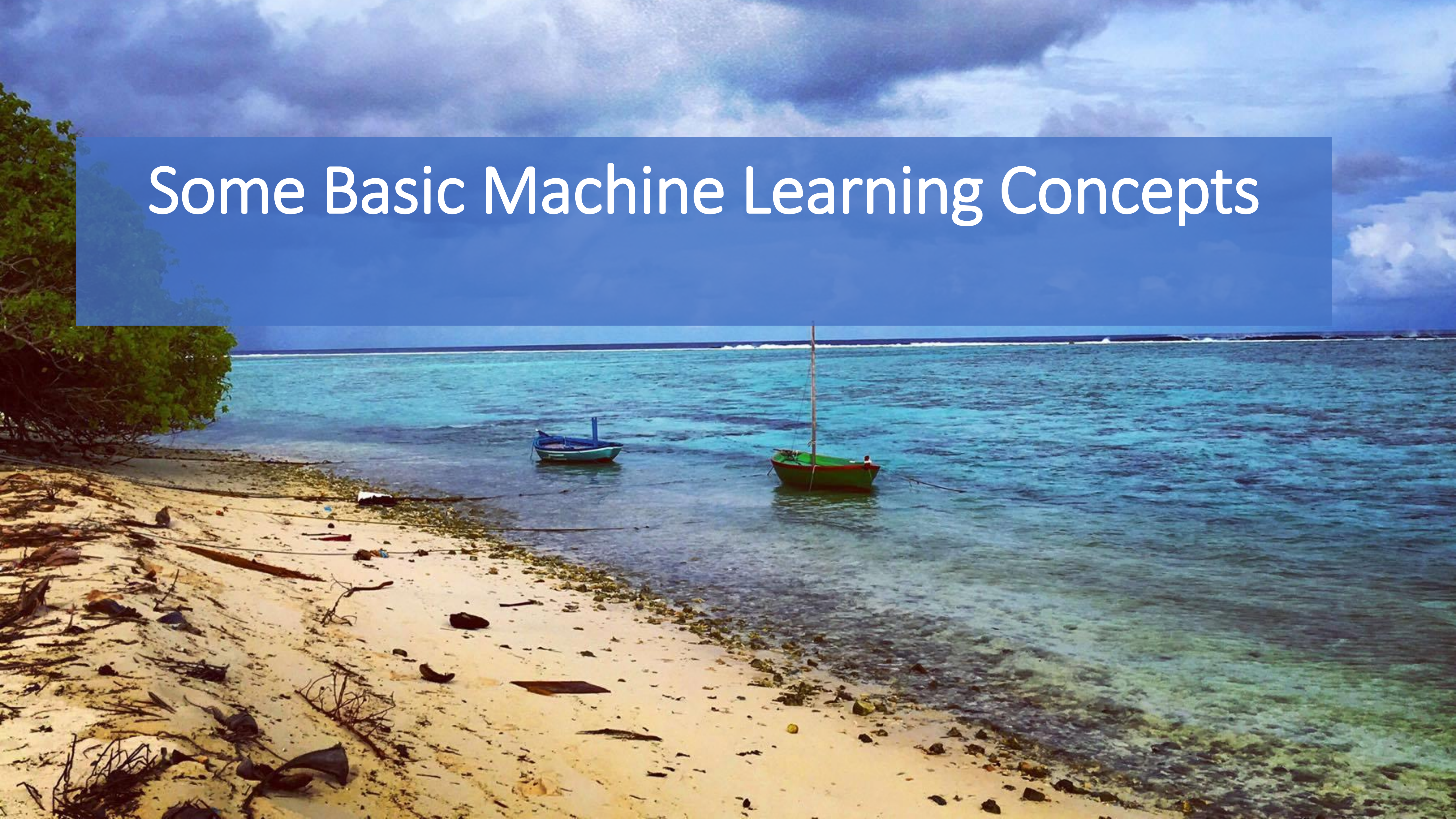
Why ML? (Anomaly Detection Aka Fraud Alerts)



Dimensionality reduction techniques in fraud analytics. The plots show the first two dimensions of PCA (left) and t-SNE (right) for fraudulent (Class = 1) and regular (Class = 0) transactions.

[Source: https://shiring.github.io/machine_learning/2017/05/01/fraud](https://shiring.github.io/machine_learning/2017/05/01/fraud)

Some Basic Machine Learning Concepts



Supervised vs Unsupervised Learning

Supervised Learning:

- For every x_i we observe some y_i
- Ex: random forests to predict loan default (y_i) based on applicant characteristics (x_i)

Unsupervised Learning:

- We only observe x_i
- Ex: clustering loan applicants based on characteristics (x_i)

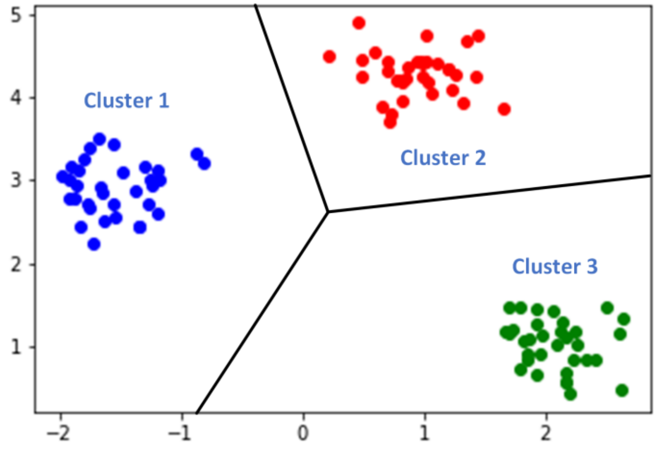
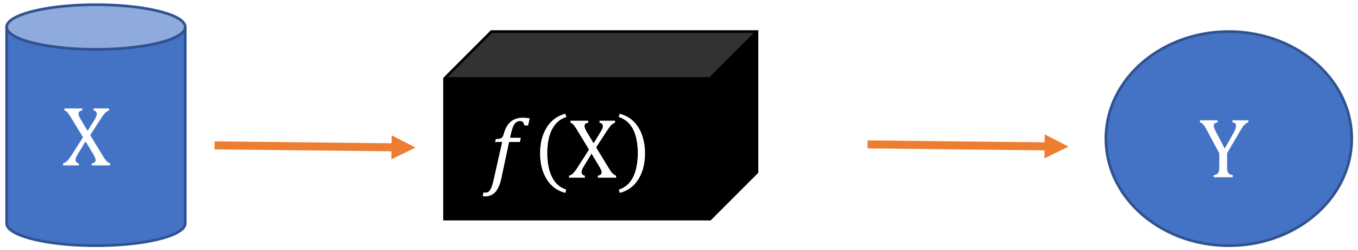


Fig.1. An Example Of Data Clustering

Supervised learning: learning $f(\mathrm{X})$ our predicted out given inputs

$\mathrm{Y} = f(\mathrm{X}) + \epsilon$



ϵ = "epsilon" (unexplained portion)

“Estimating” $\hat{f}(X)$

- $\mathrm{Y} = f(\mathrm{X}) + \epsilon$ is the true value
- We can only use data to “guess” at $f(\mathrm{X})$
- We call this guess $\hat{f}(X)$

How do we know when we’ve selected a “good” $\hat{f}(X)$?

- We reserve a portion of our data into a “test” set, estimate a model on the other part, and see how our model performs on this test set

Testing Training Data Subsets

Training set: (observation-wise) subset of data used to develop models

Testing/Training Split

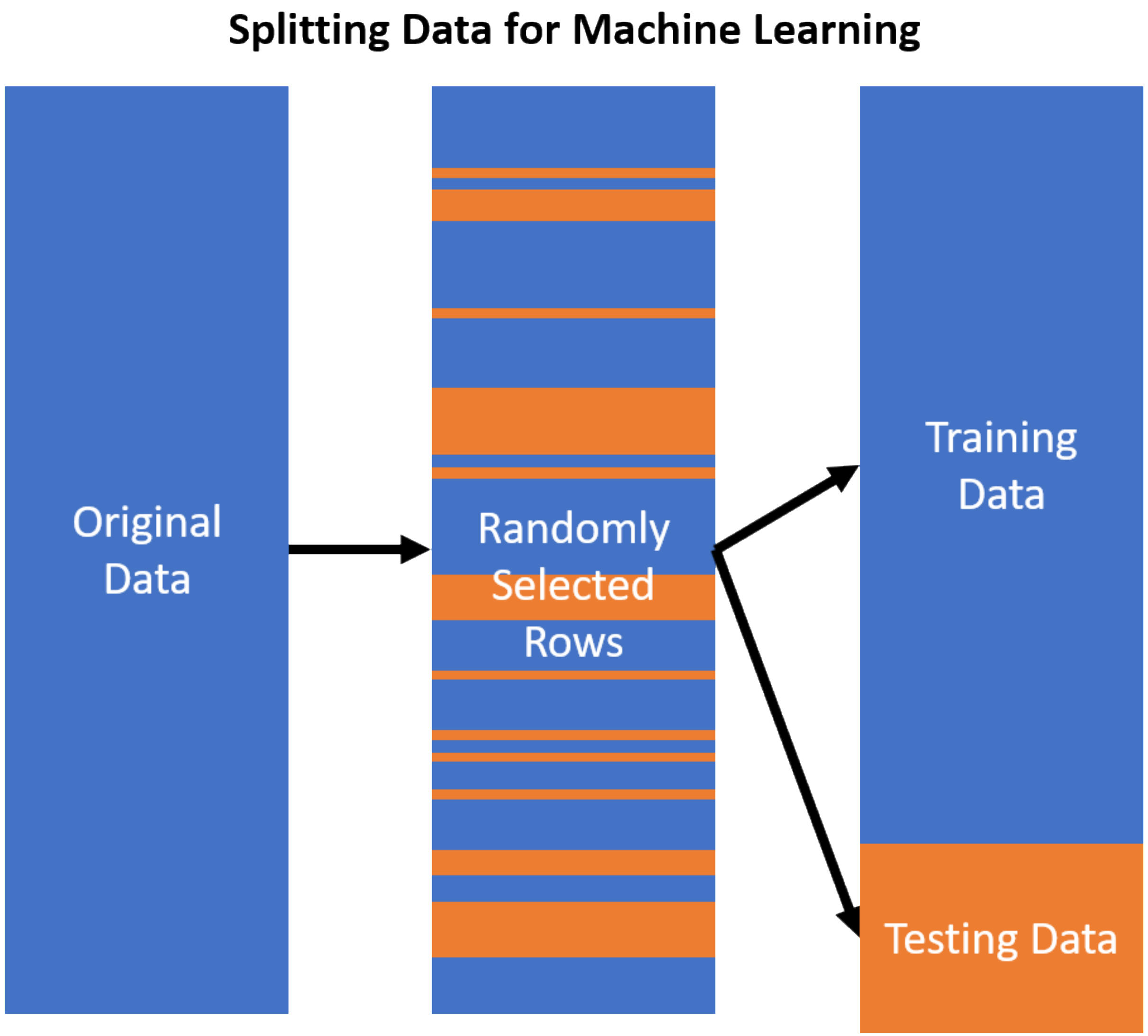
Training set: (observation-wise) subset of data used to develop models

Test or Validation set: subset of data used during intermediate stages to “tune” model parameters

Rule of thumb 75% training 25% test -ish

Randomly Selecting Rows for Test or Training Sets

- Observations are randomly selected into either testing or training splits of the data

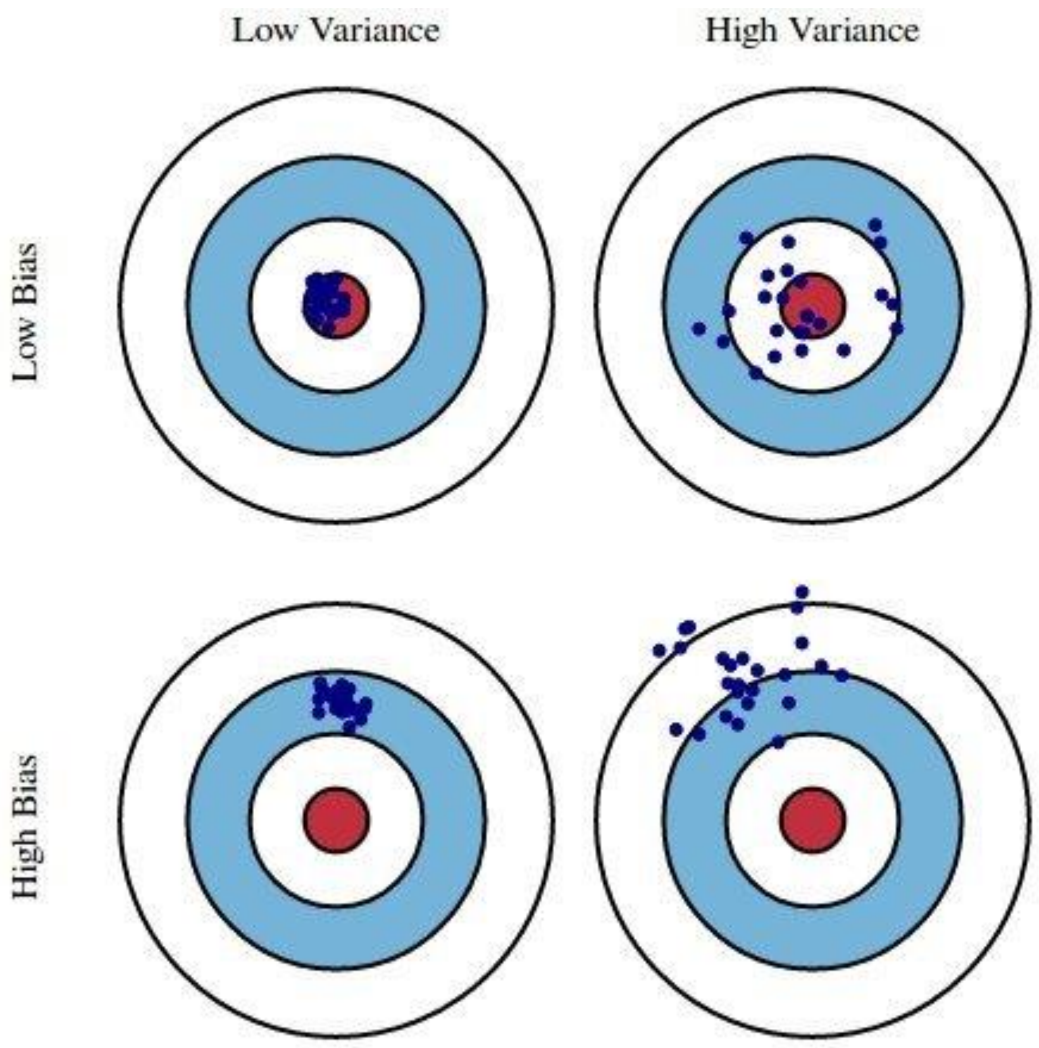


Bias and Variance

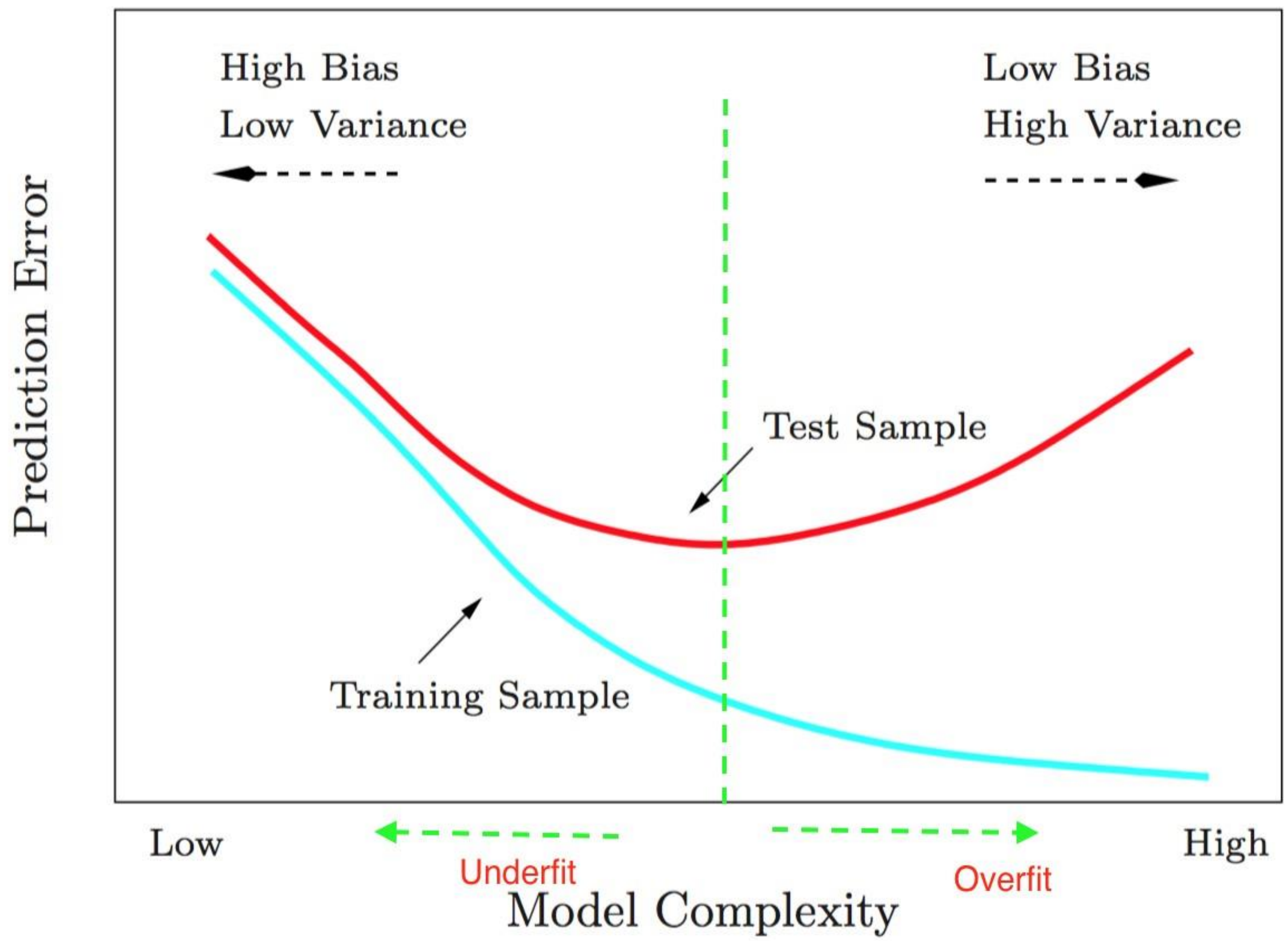
Bias: **Tendency of an in-sample statistic to over or under estimate the statistic in the *population***

Variance: **Tendency to noisily estimate a statistic**.
E.g., sensitivity to small fluctuations in the training dataset.

Bias-Variance Tradeoff



Bias-Variance Tradeoff

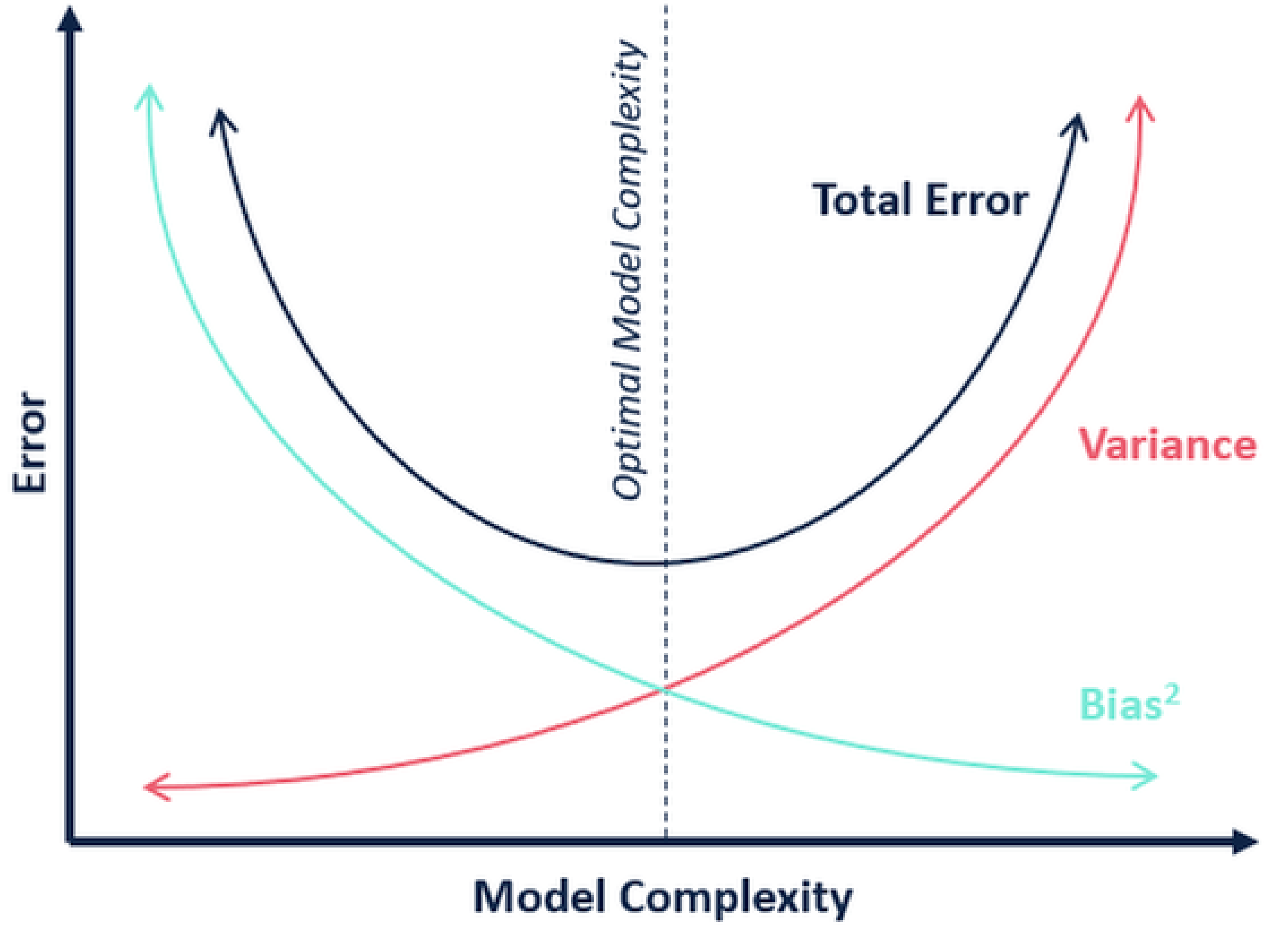


- Error in <u>Training</u> sample (~bias) ↓ as we ↑ model complexity (e.g. number of variables)

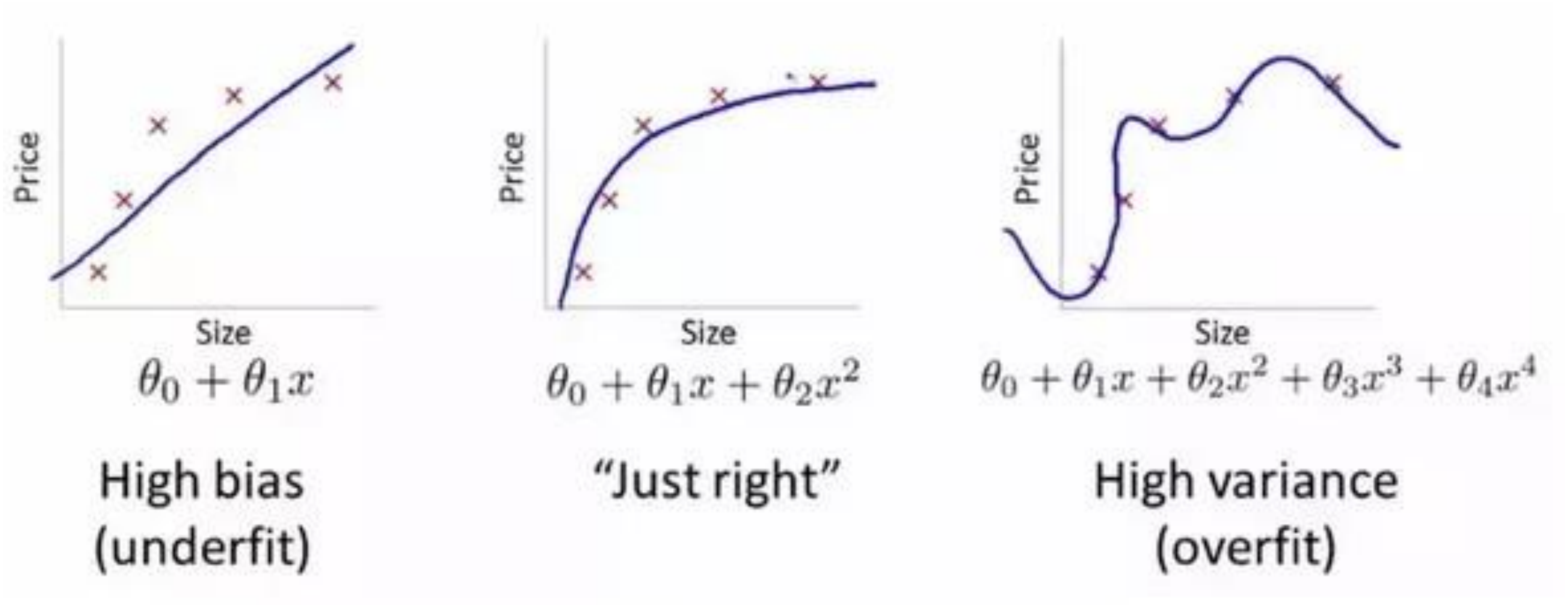
- Error in <u>Test</u> sample (~variance) ↑ as we ↑

- Key: finding optimal model complexity

Key: Finding Optimal Model Complexity



Optimal Model Complexity: Neither Underfit Nor Overfit



Assessing Model Accuracy: Mean Squared Error

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2$$

$\sum$ means we add up anything with i, starting at $i = 1$ to $i = n$

y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
5	5	0	$0^2 = 0$
6	7	-1	$-1^2 = 1$
9	8	1	$1^2 = 1$
10	1	9	$9^2 = 81$

Summary – Intro to Machine Learning

- **Machine Learning** is a set of methods developed to find robust patterns across datasets
- **Public Policy can benefit from machine learning.**
  - Big data requires it
  - Non-linear relationships
  - Better forecasts/econometrics
  - Anomaly detection
- **Remember these key concepts**
  - Supervised (Y,X) vs Unsupervised learning (just X)
  - Testing/Training Sets
    - (model -> train, see how it performs on test)
  - Bias-Variance Tradeoff
    - Bias – how far off model from true
    - Variance – precision of estimated model
    - Total error = bias^2 + variance