Applied Machine Learning

Amit Kapoor
@amitkaps

Bargava Subramanian
@bargava

“WHAT I CANNOT CREATE, I DO NOT UNDERSTAND ”

Richard Feynman

Approach

- Understand the ML Process.
- Code it to learn it.
- Play with code.

Agenda

Module 1: Linear Models
Module 2: Model Evaluation
Module 3: Tree-Based Models
Module 4: Unsupervised Learning

SCHEDULE

1000 - 1130: Session 1
1120 - 1130: Break
1130 - 1230: Session 2
1230 - 1330: Lunch
1330 - 1530: Session 3
1530 - 1540: Break
1540 - 1700: Session 4

“IT’S TOUGH TO MAKE PREDICTIONS, ESPECIALLY ABOUT THE FUTURE.”

Yogi Berra

WHAT IS MACHINE LEARNING (ML)?

“[Machine learning is the] field of study that gives computers the ability to learn without being explicitly programmed.”
Arthur Samuel

“Machine learning is the study of computer algorithm that improve automatically through experience”
Tom Mitchell

ML Problems

» “Is this cancer?”

» “What is the market value of this house?”

» “Which of these people are friends?”

» “Will this person like this movie?”

» “Who is this?”

» “What did you say?”

» “How do you fly this thing?”.

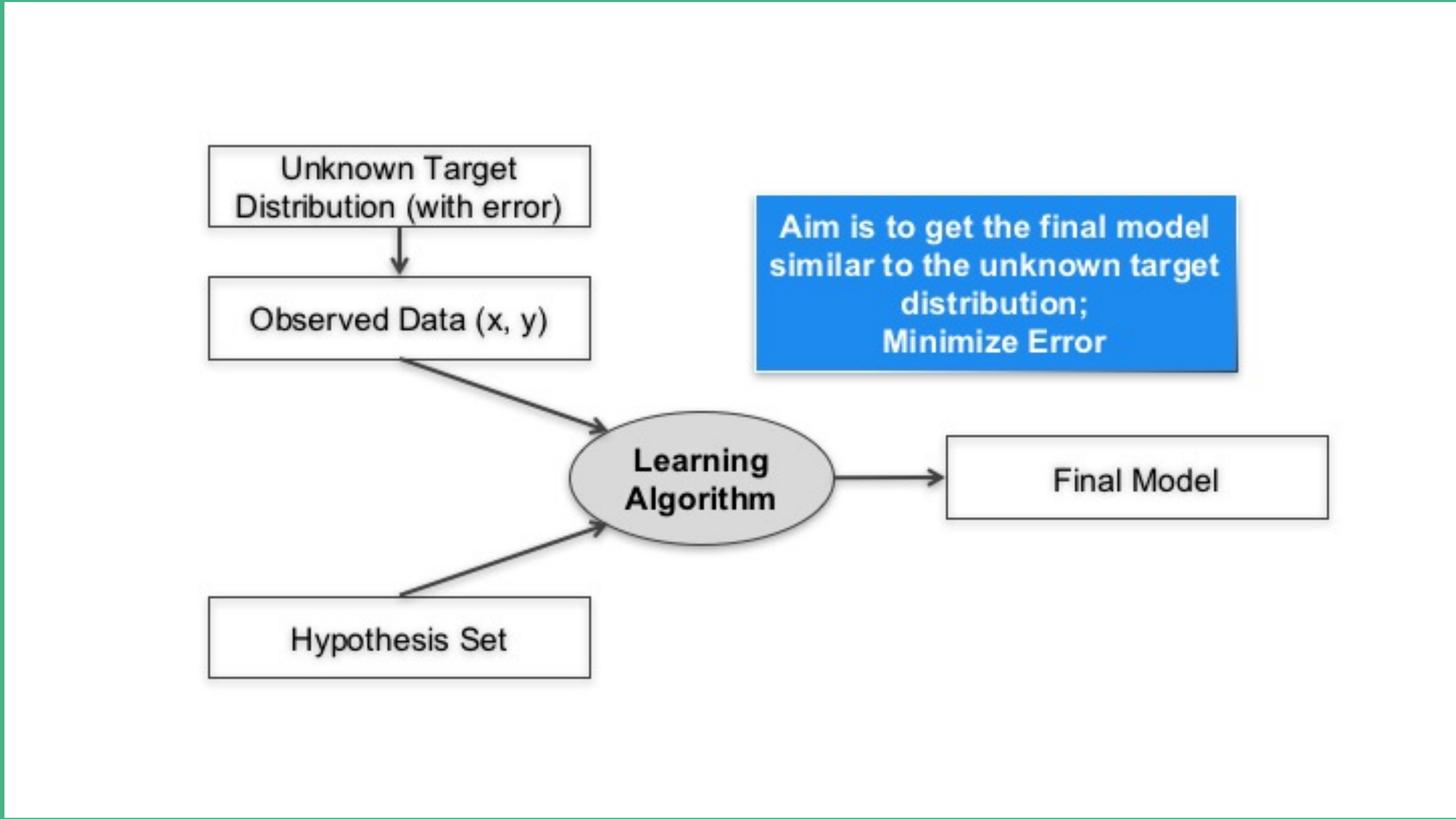
ML in use everyday

- Search
- Photo Tagging
- Spam Filtering
- Recommendation
- ...

BROAD ML APPLICATION

- Database Mining e.g. Clickstream data, Business data
- Automating e.g. Handwriting, Natural Language Processing, Computer Vision
- Self Customising Program e.g. Recommendations

ML THOUGHT PROCESS



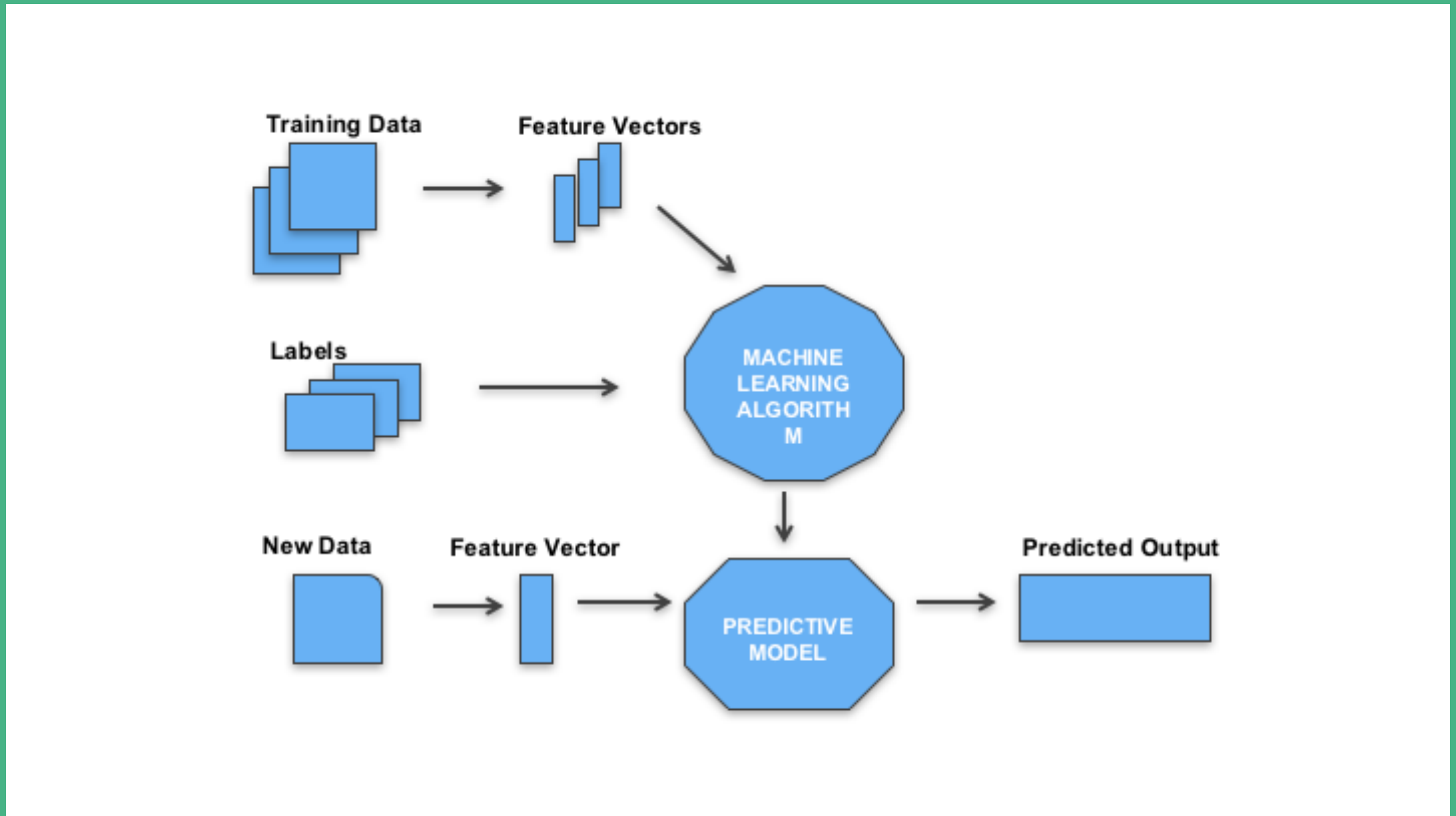
Learning Paradigm

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Online Learning

SUPERVISED LEARNING

» Regression

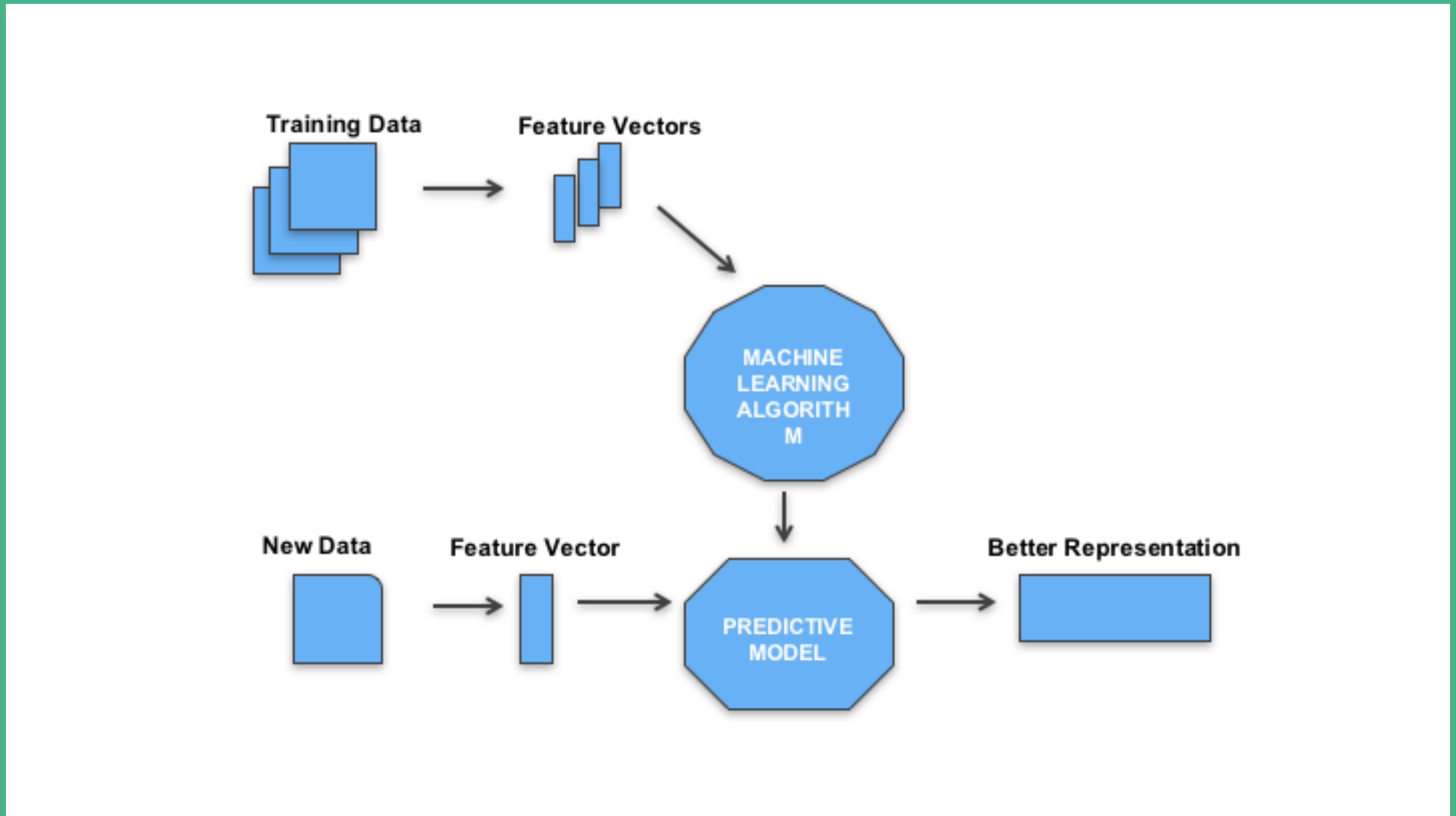
» Classification



Unsupervised Learning

» Clustering

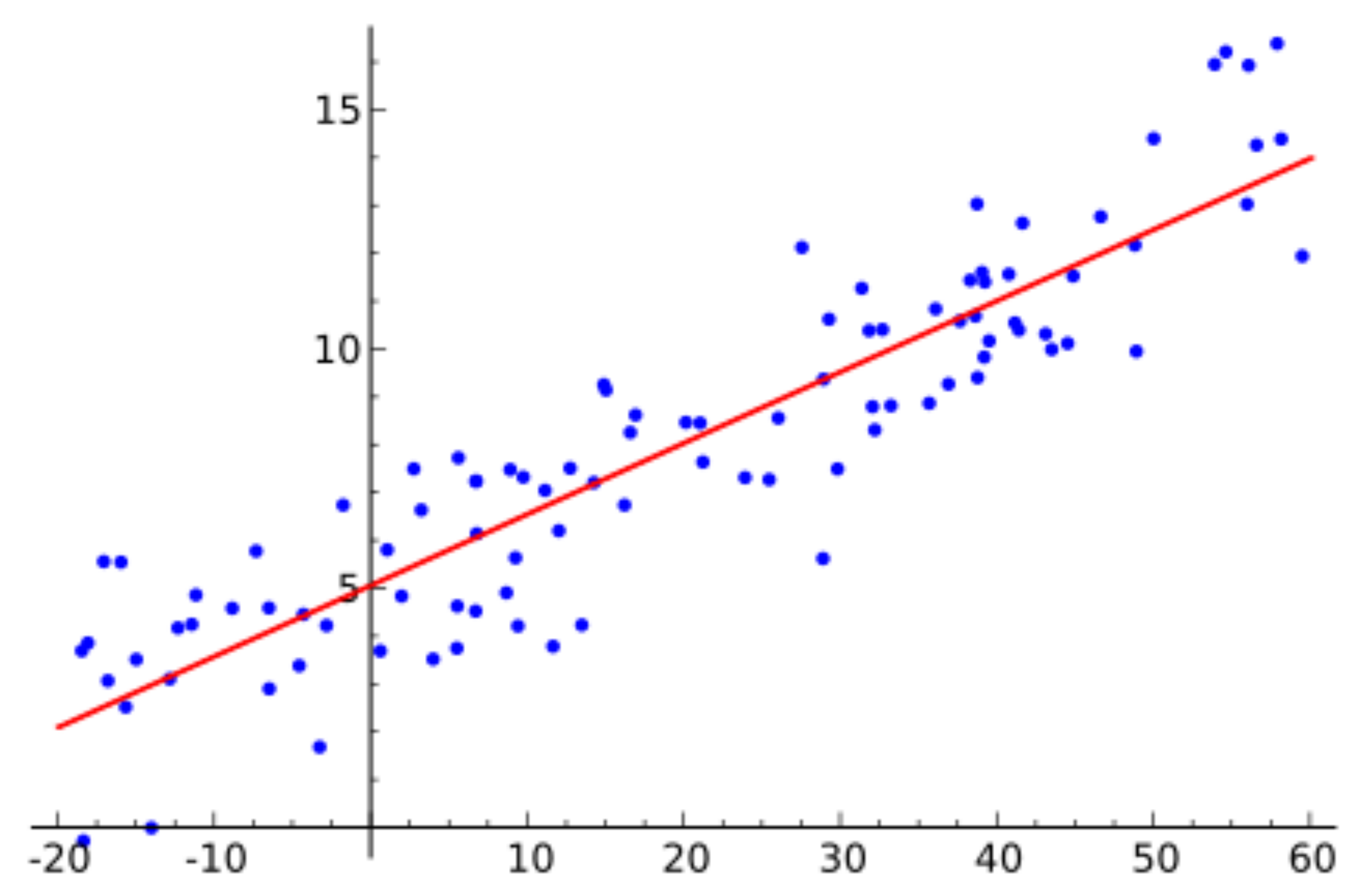
» Dimensionality Reduction



ML Pipeline

- » Frame: Problem definition
- » Acquire: Data ingestion
- » Refine: Data wrangline
- » Transform: Feature creation
- » Explore: Feature selection
- » Model: Model creation & assessment
- » Insight: Communication

Linear Regression



LINEAR RELATIONSHIP

$$y_i = \alpha + \beta_1 x_1 + \beta_2 x_2 + ..$$

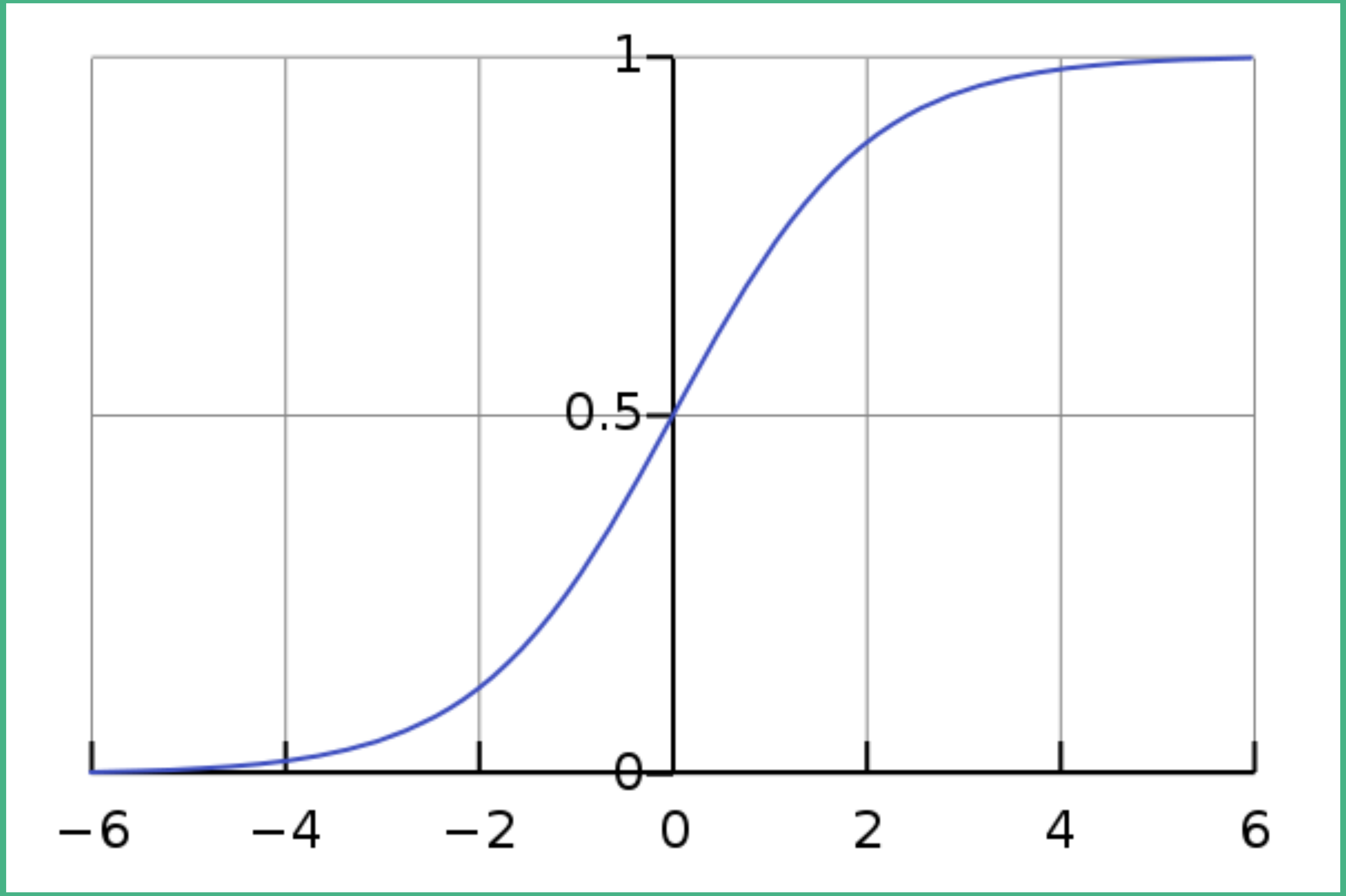
OBJECTIVE FUNCTION

$$\epsilon = \sum_{k=1}^{n} (y_i - \hat{y}_i)^2$$

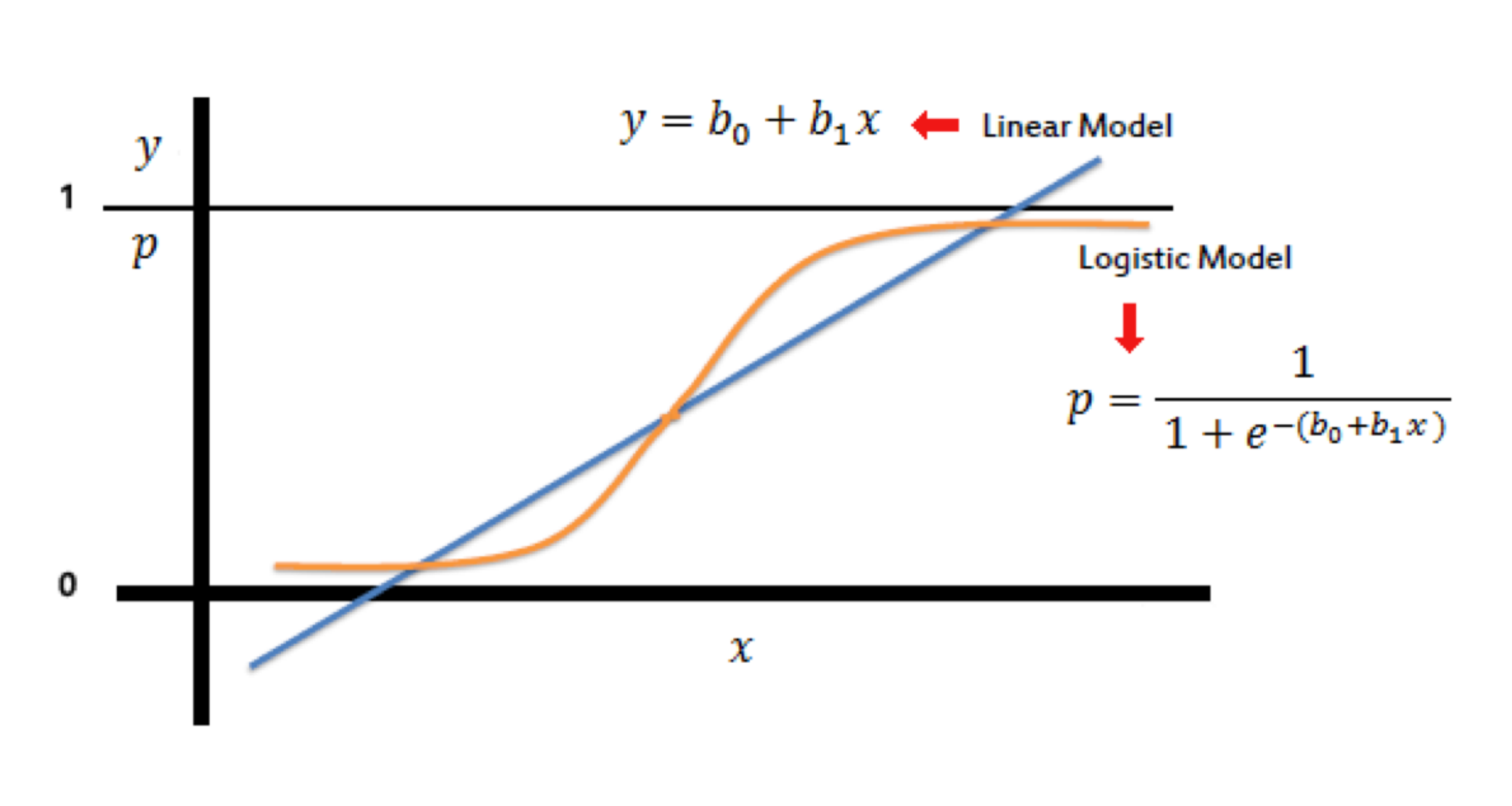
Interactive Example: http://setosa.io/ev/

LOGIT FUNCTION

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$



LOGISTIC REGRESSION



Logistic Relationship

Find the β parameters that best fit:

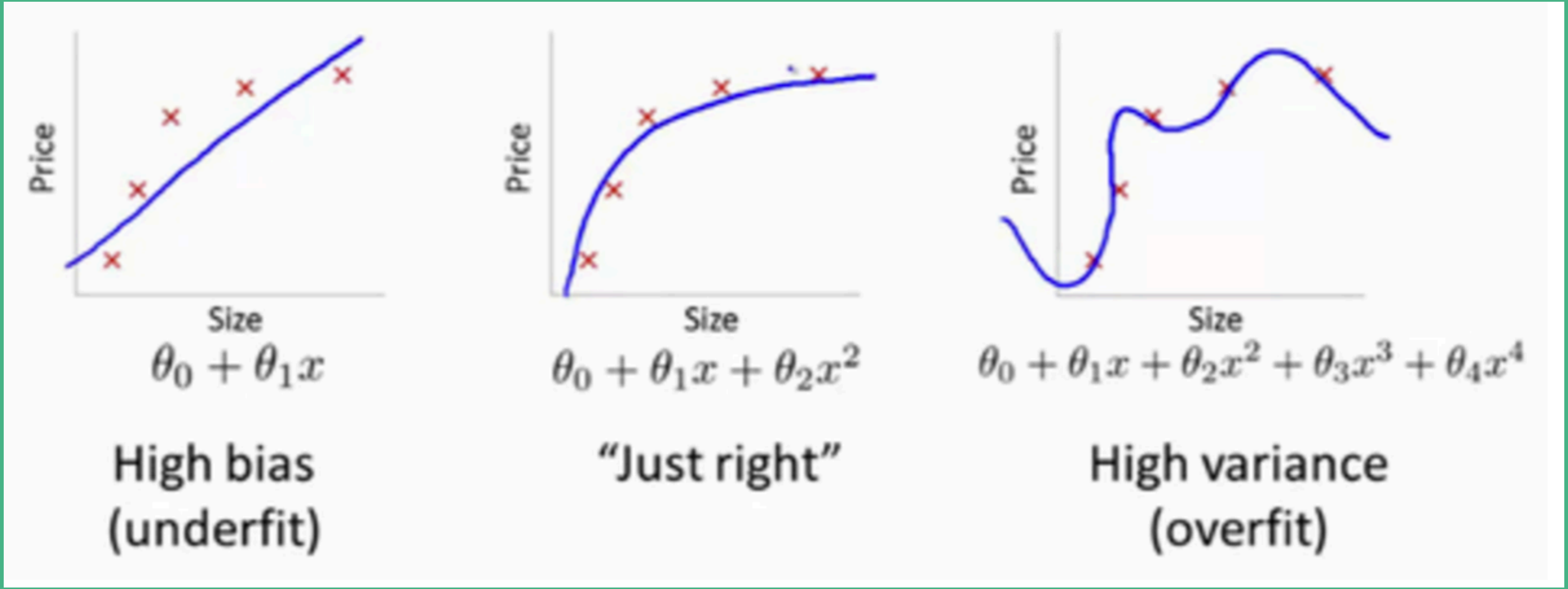
$y = 1$ if $\beta_0 + \beta_1 x + \epsilon > 0$

$y = 0$, otherwise

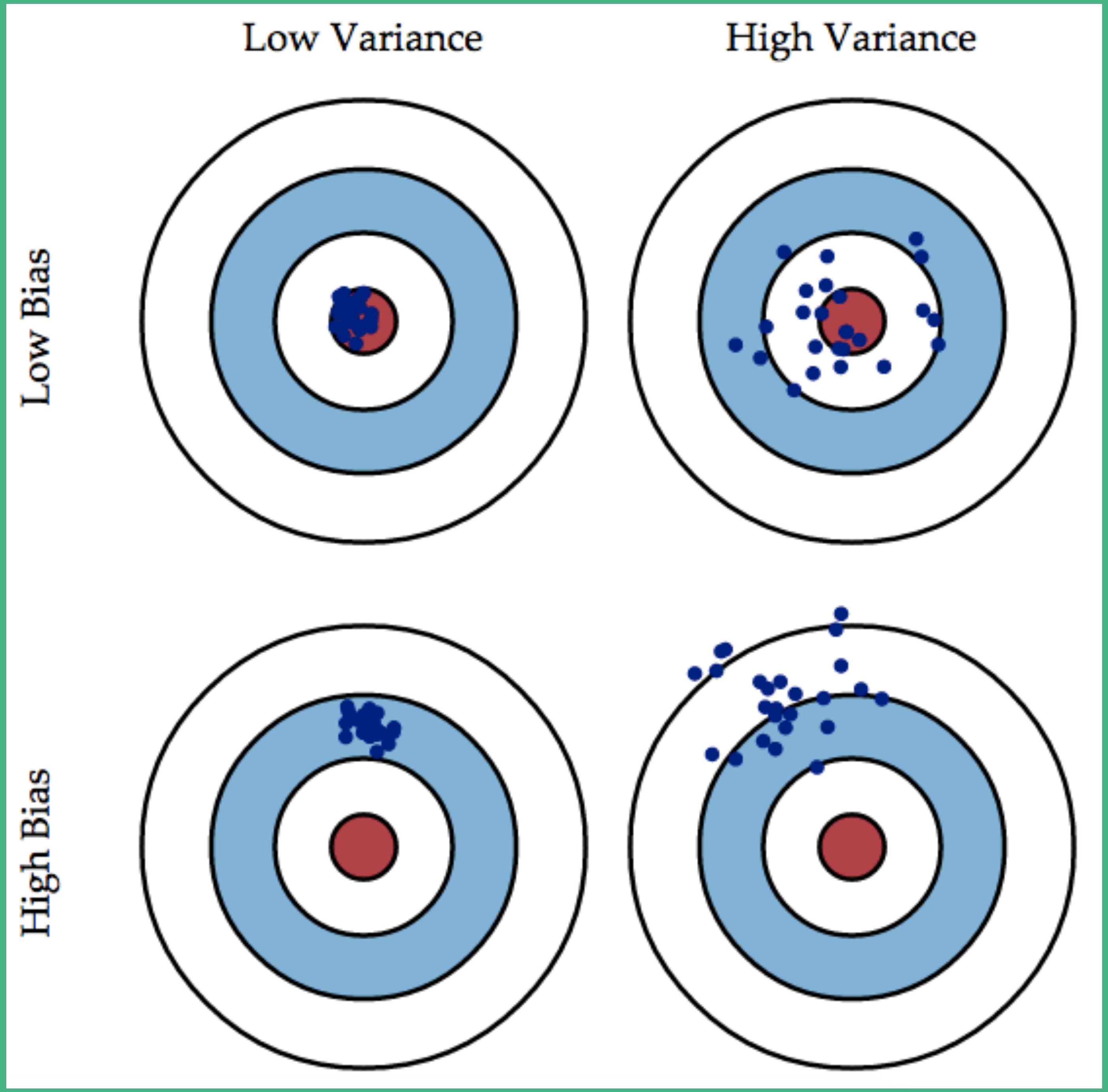
Follows:

$$P(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

FITTING A MODEL

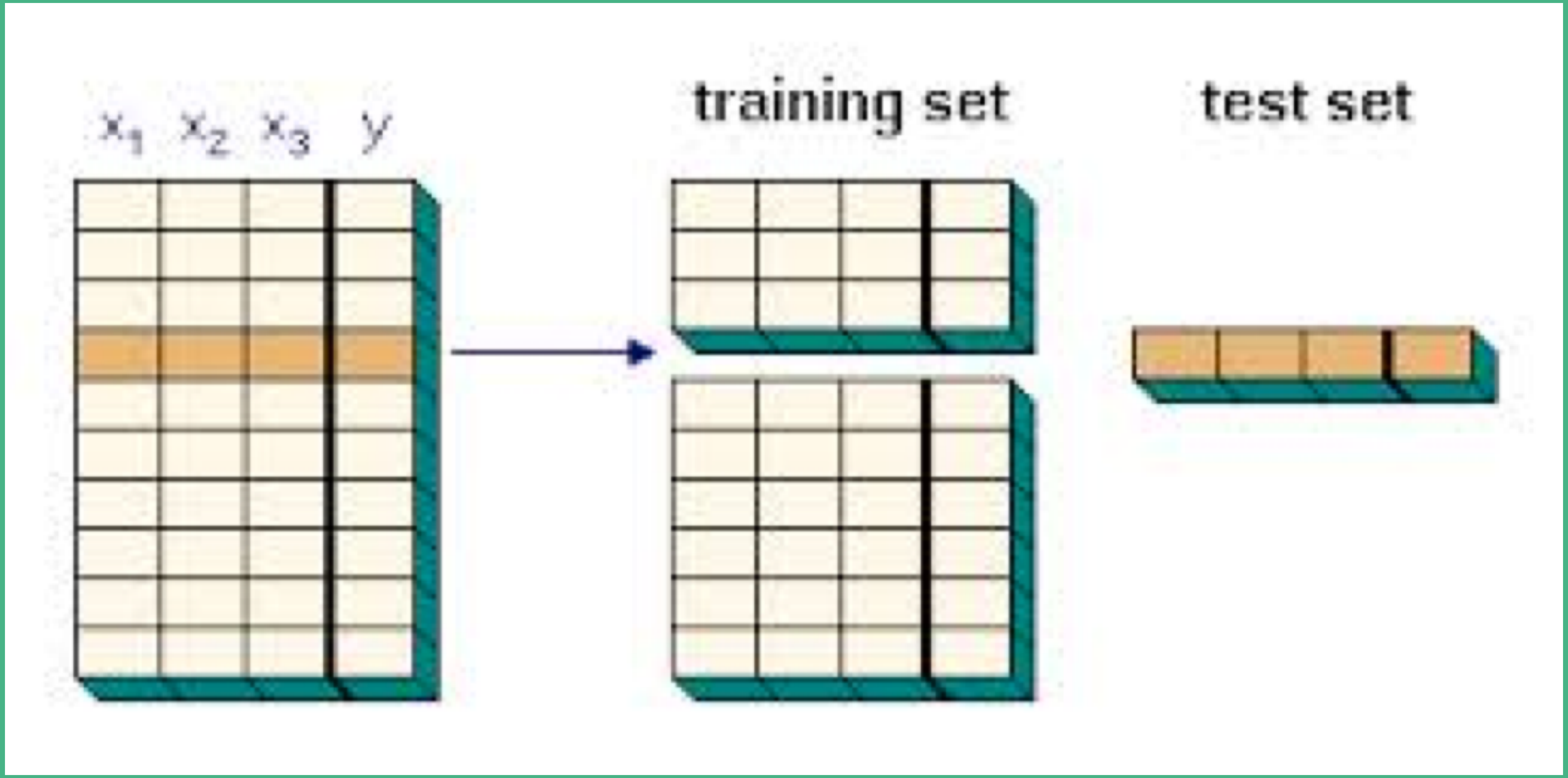


Bias-Variance Tradeoff



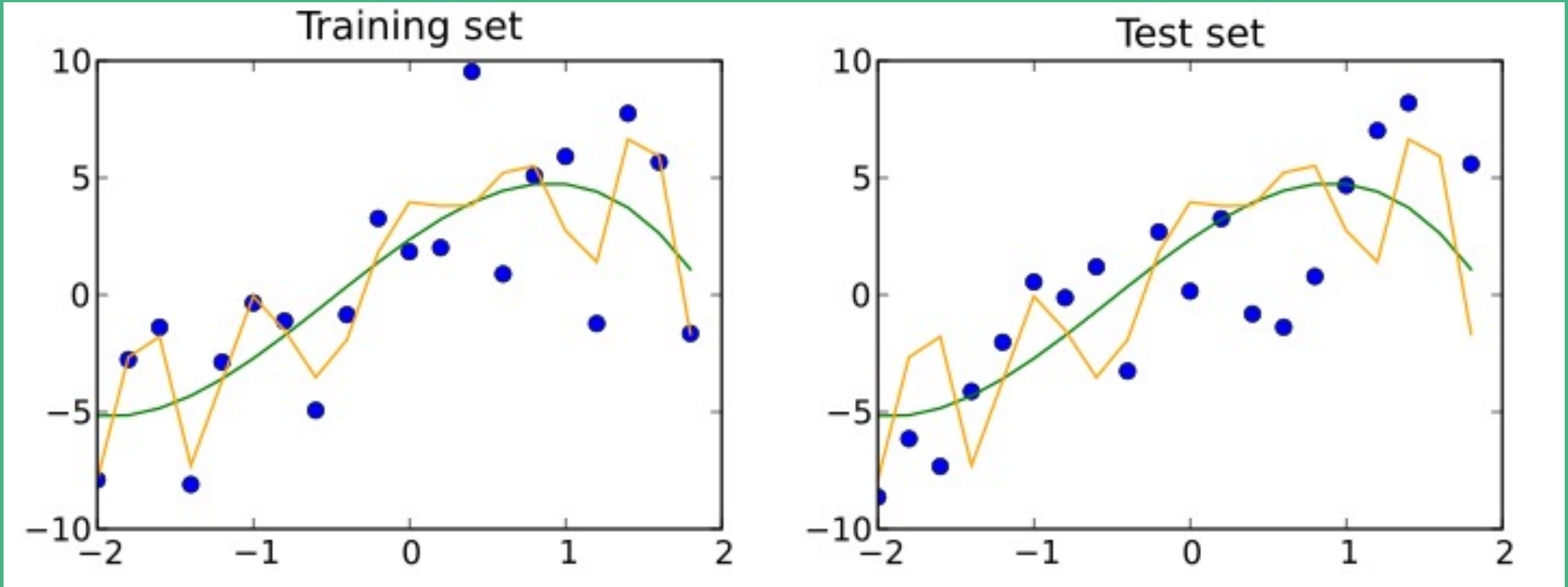
TRAIN AND TEST DATASETS

Split the Data - 80% / 20%

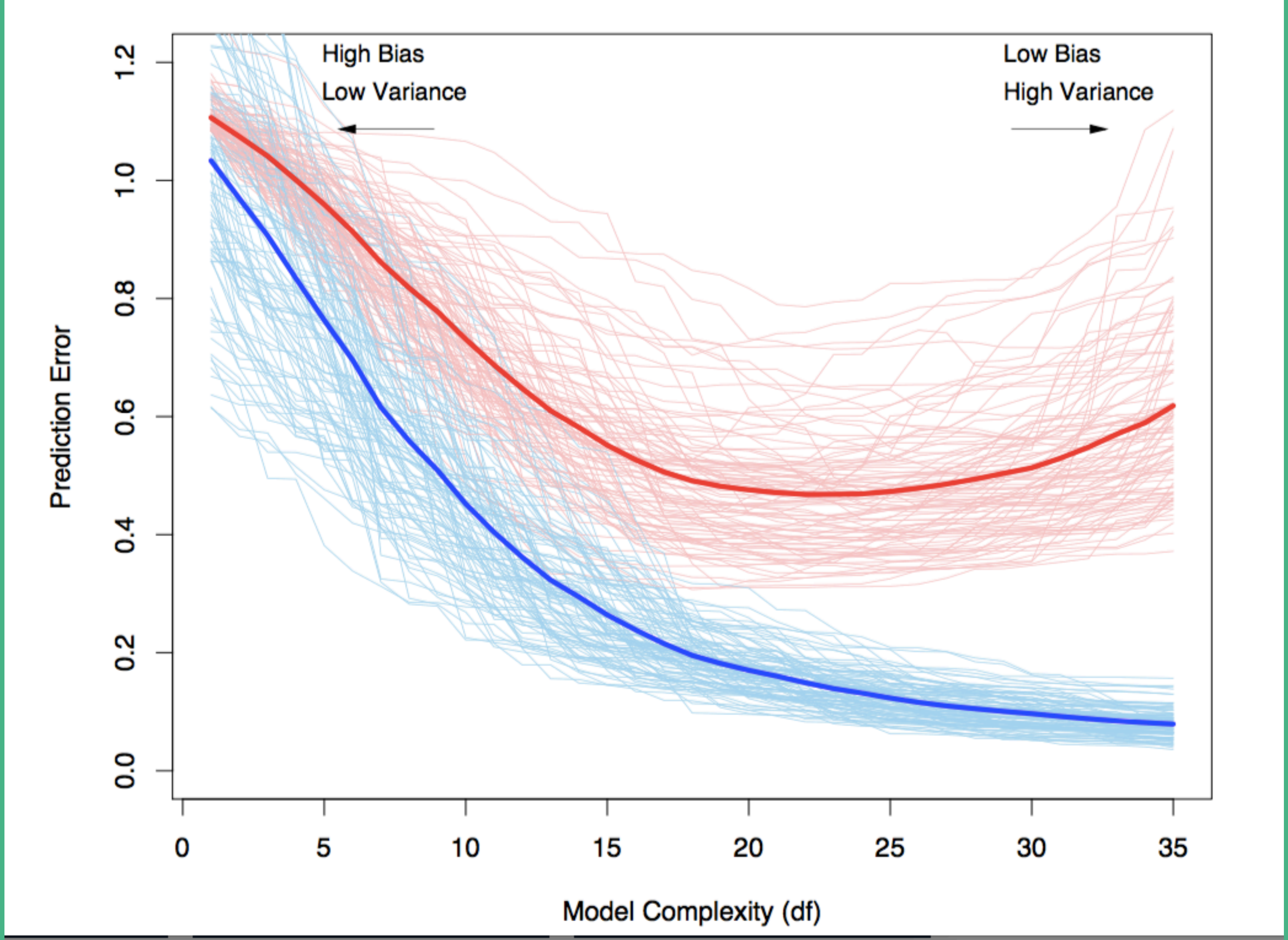


TRAIN AND TEST DATASETS

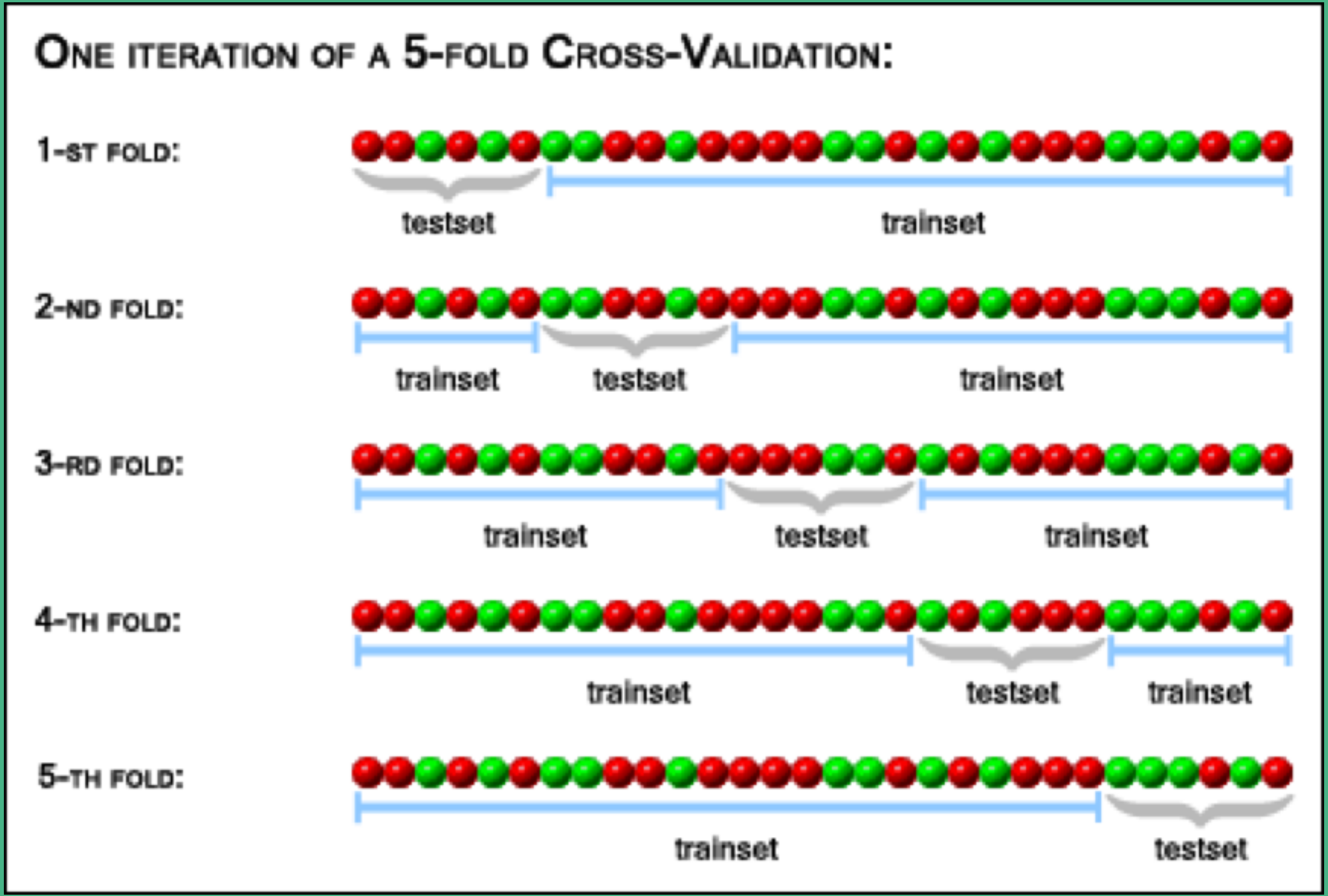
Measure the error on Test data



MODEL COMPLEXITY

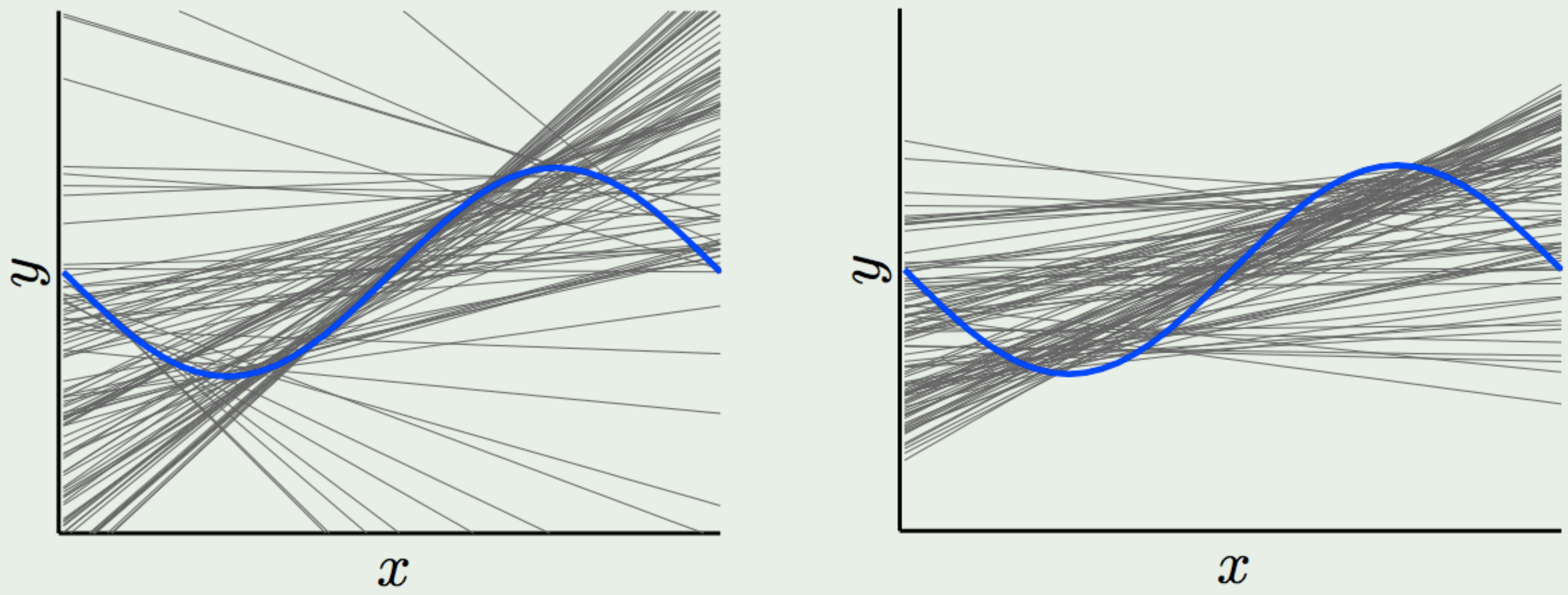


CROSS VALIDATION



REGULARIZATION

Attempts to impose Occam's razor on the solution



without regularization

with regularization

Model Evaluation

Mean Squared Error

$$MSE = 1/n \sum_{k=1}^{n} (y_i - \hat{y}_i)^2$$

MODEL EVALUATION

Confusion Matrix

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

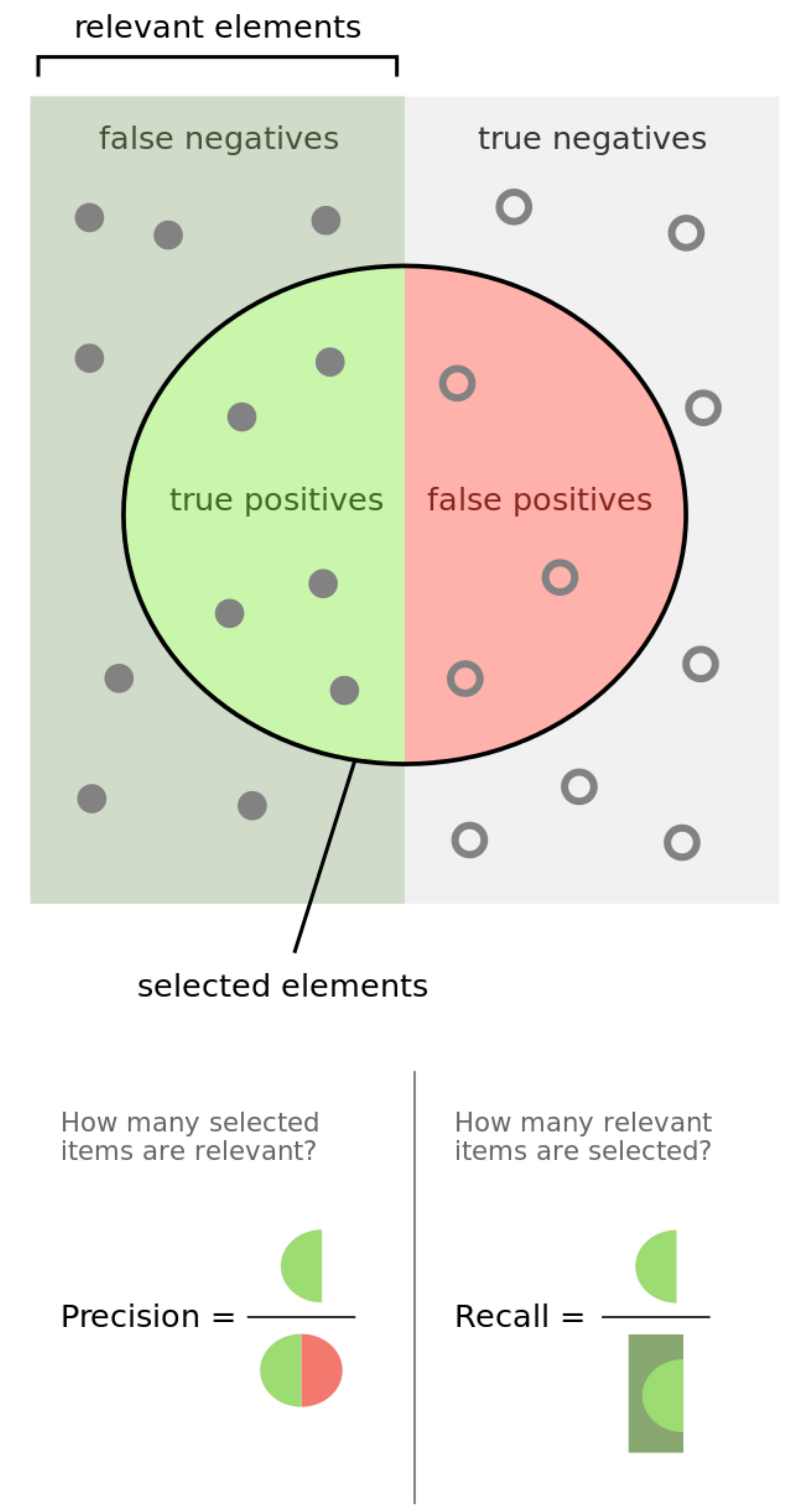
MODEL EVALUATION

Classification Metrics

Recall (TPR) = TP / (TP + FN)

Precision = TP / (TP + FP)

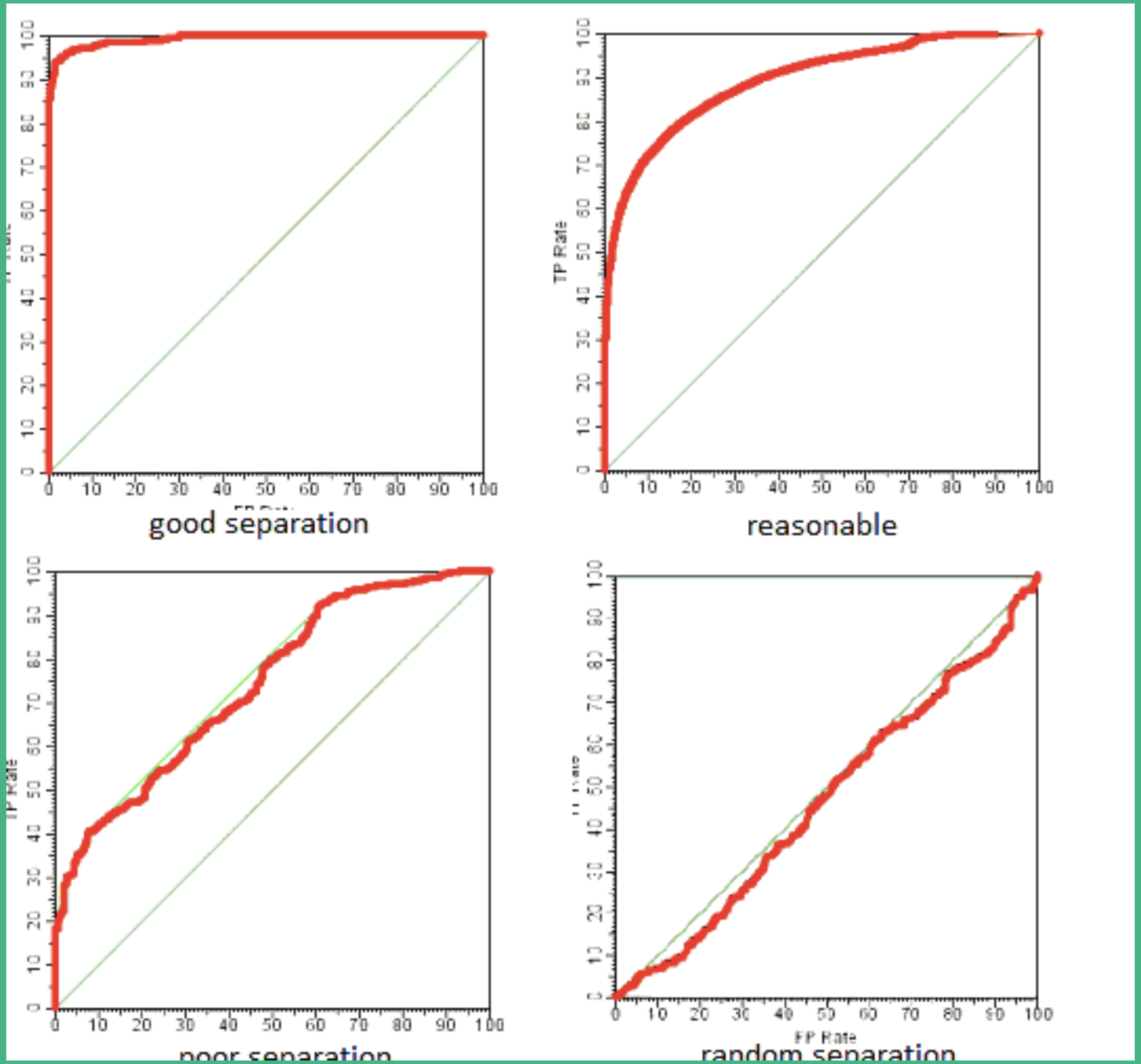
Specificity (TNR) = TN / (TN + FP)



MODEL EVALUATION

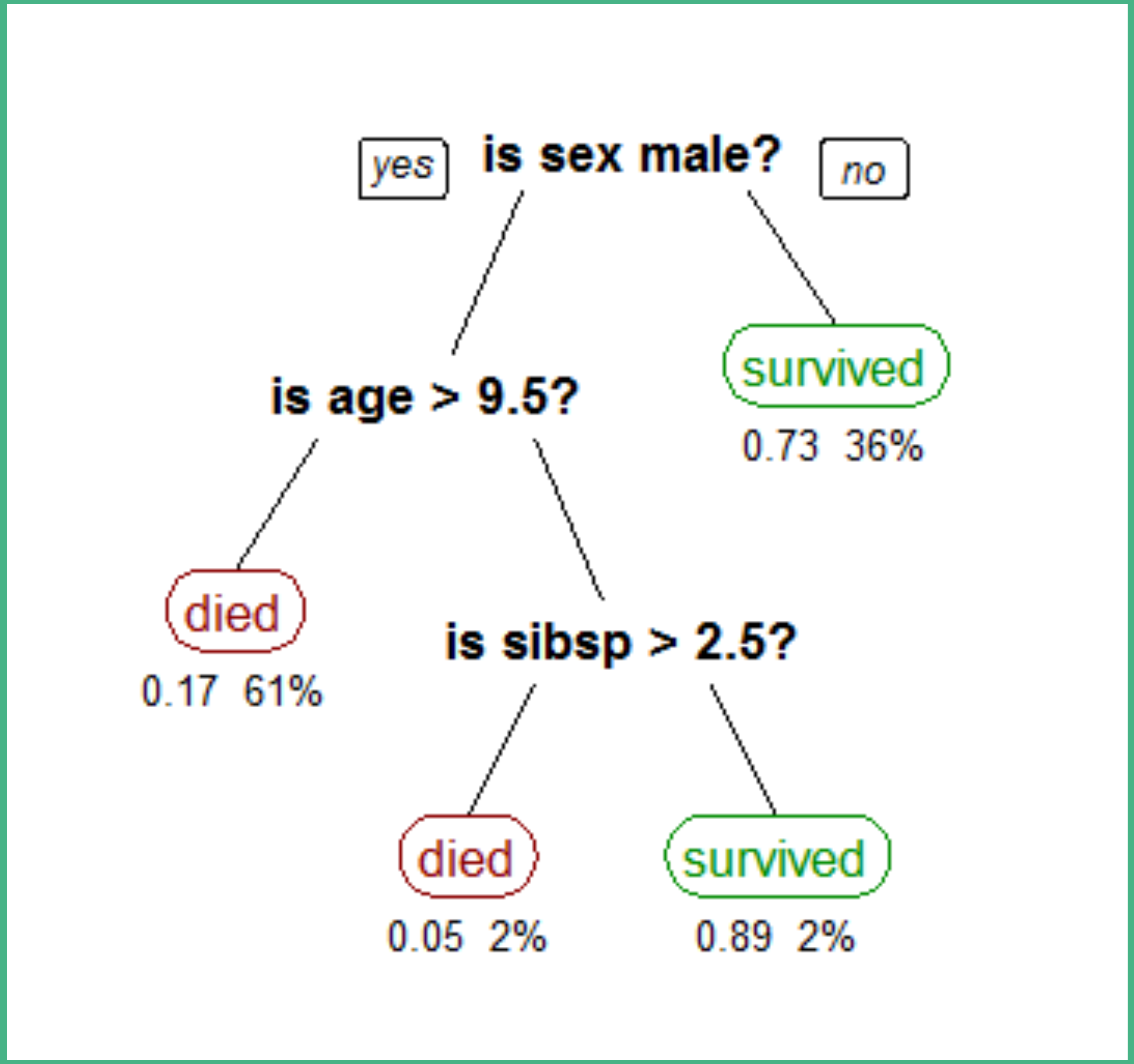
Receiver Operating Characteristic Curve

Plot of TPR vs FPR at different discrimination threshold



DECISION TREE

Example: Survivor on Titanic



Decision Tree

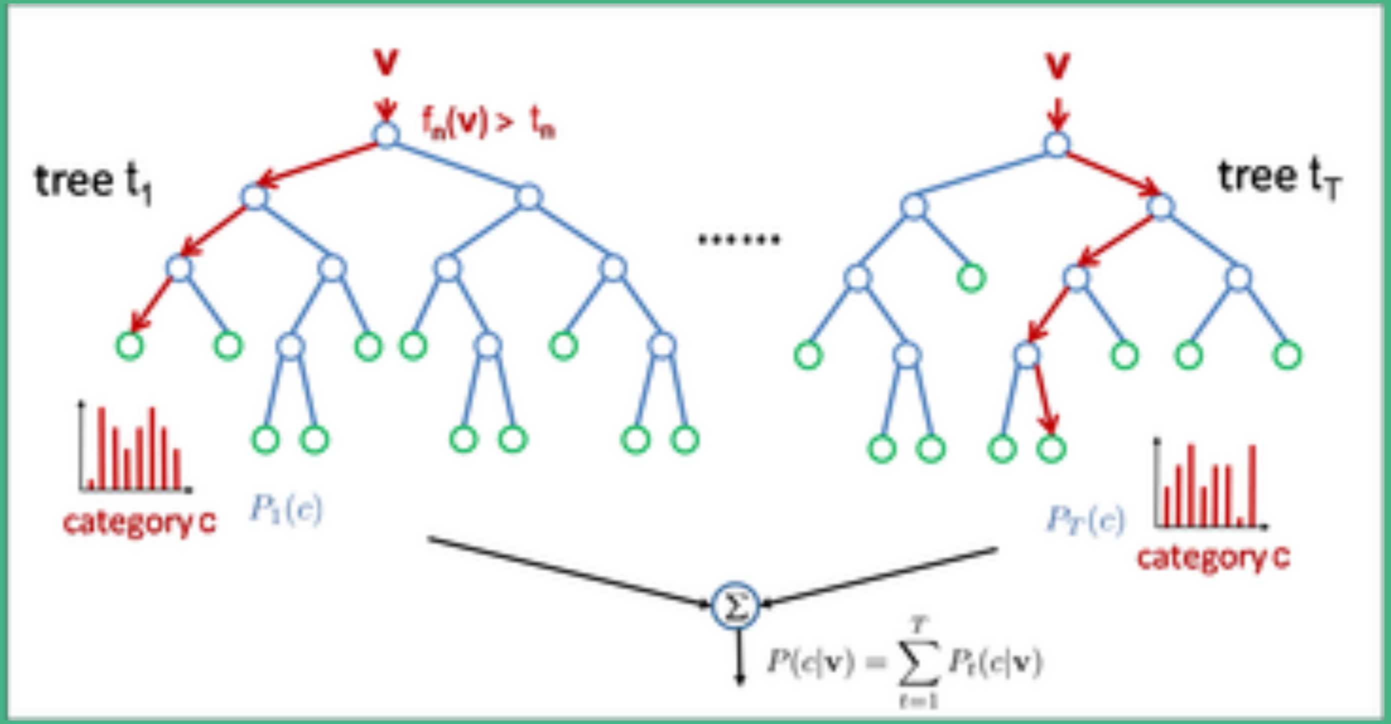
- Easy to interpret
- Little data preparation
- Scales well with data
- White-box model
- Instability - changing variables, altering sequence
- Overfitting

Bagging

- Also called bootstrap aggregation, reduces variance
- Uses decision trees and uses a model averaging approach

Random Forest

- Combines bagging idea and random selection of features.
- Similar to decision trees are constructed – but at each split, a random subset of features is used.



"IF YOU TORTURE THE DATA ENOUGH, IT WILL CONFESS."

Ronald Case

Challenges

- Data Snooping
- Selection Bias
- Survivor Bias
- Omitted Variable Bias
- Black-box model Vs White-Box model
- Adherence to regulations

Reflections

» Steep learning curve

» Different level

» Balance speed and coverage

» Be considerate