Model Accuracy and Bias-Variance Tradeoff

Supervised Learning

Daniel E. Acuna
Associate Professor, University of Colorado Boulder

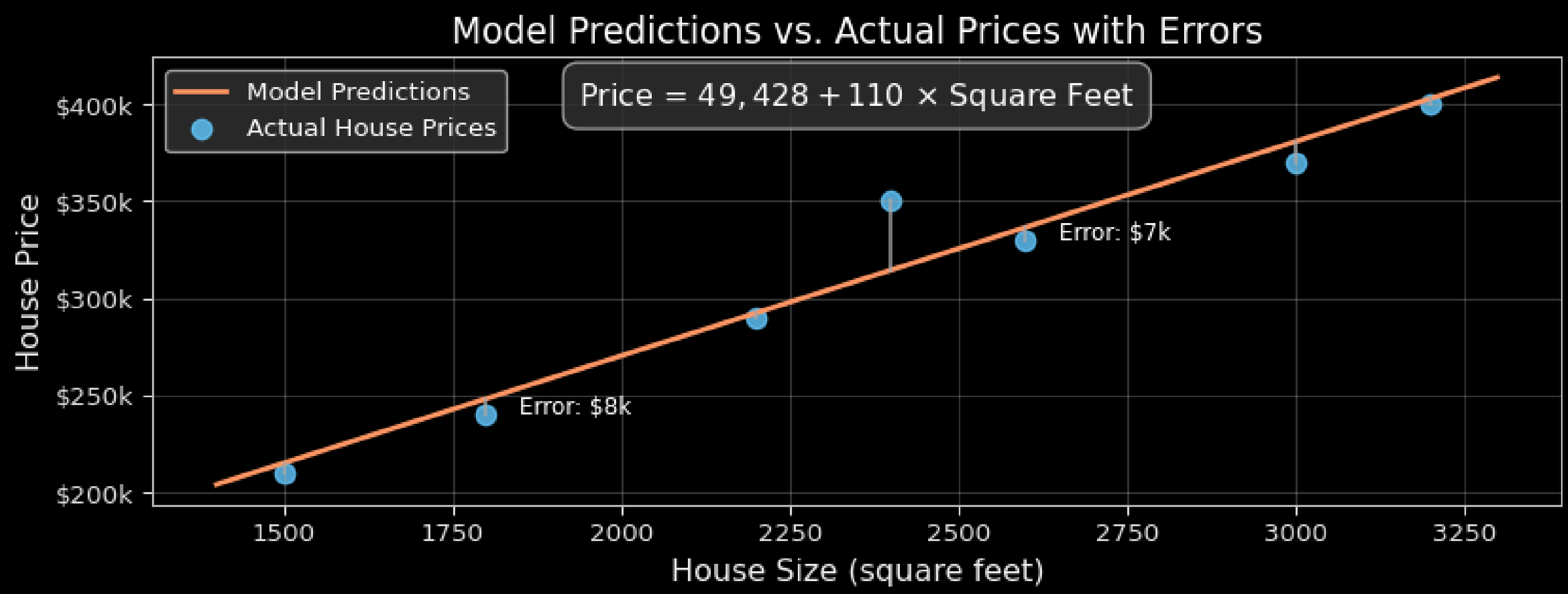
Contents of This Video

In this video, we will cover:

- Understanding model error and accuracy metrics
- The concept of bias and variance in machine learning
- How the bias-variance tradeoff affects model performance
- Examples of underfitting (high bias) and overfitting (high variance)
- Visual explanations using the bullseye analogy
- Practical strategies for finding the optimal model complexity

Understanding Model Error

$$\text{Price} = \beta_0 + \beta_1 \times \text{Square Feet}$$



Square Feet	Actual Price	Predicted	Error	Squared Error
1500	$210k	$215k	$-5.1k	$25.58M²
1800	$240k	$248k	$-8.2k	$66.96M²
2200	$290k	$292k	$-2.4k	$5.53M²
2400	$350k	$314k	$35.6k	$1264.90M²
2600	$330k	$337k	$-6.5k	$42.49M²
3000	$370k	$381k	$-10.7k	$114.19M²
3200	$400k	$403k	$-2.8k	$7.67M²

Metric	Formula	Value
MAE	$\frac{1}{n}\sum\lvert y_i - \hat{y}_i\rvert$	$10.2k
MSE	$\frac{1}{n}\sum(y_i - \hat{y}_i)^2$	$218.19M²
RMSE	$\sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2}$	$14.8k

Interactive Linear Regression Model

What Makes a Good Model?

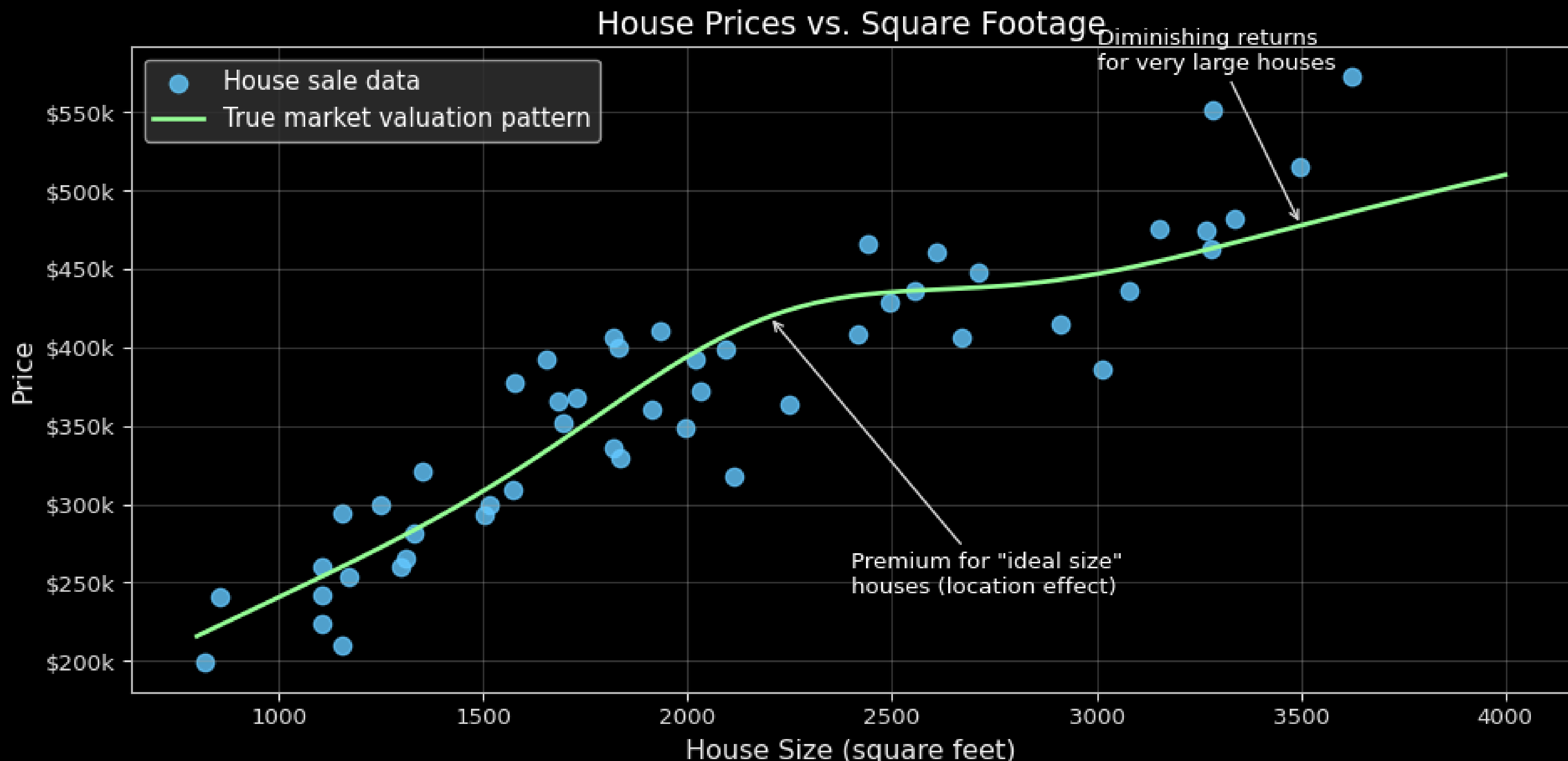
- **Good performance on training data**
- **Good performance on unseen data** (generalization)
- Balance between:
  - Capturing patterns in data
  - Avoiding noise/random fluctuations

The Bias-Variance Tradeoff

- **Bias**: Error from incorrect assumptions
- **Variance**: Error from sensitivity to fluctuations
- **Tradeoff**: Reducing one often increases the other
- Goal: Find the sweet spot with minimal total error

Target with bullseye

A Real-World Example: House Prices and Square Footage



Understanding Bias

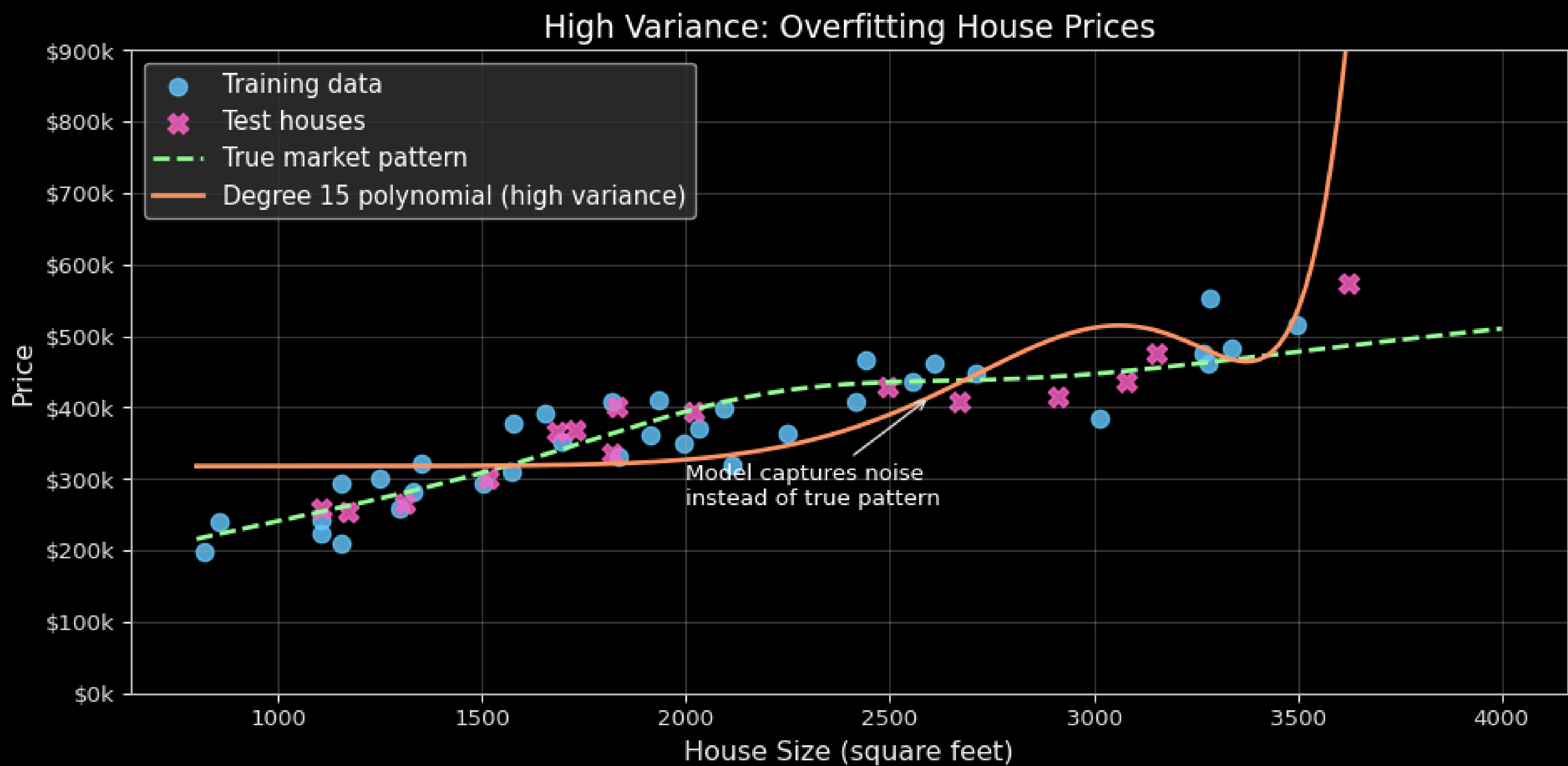
- **High Bias**: Model makes strong assumptions
- Leads to **underfitting**
- Example: Linear model for non-linear data
- Symptoms:
  - Poor performance on training data
  - Poor performance on test data

Understanding Bias: Linear Model for House Prices

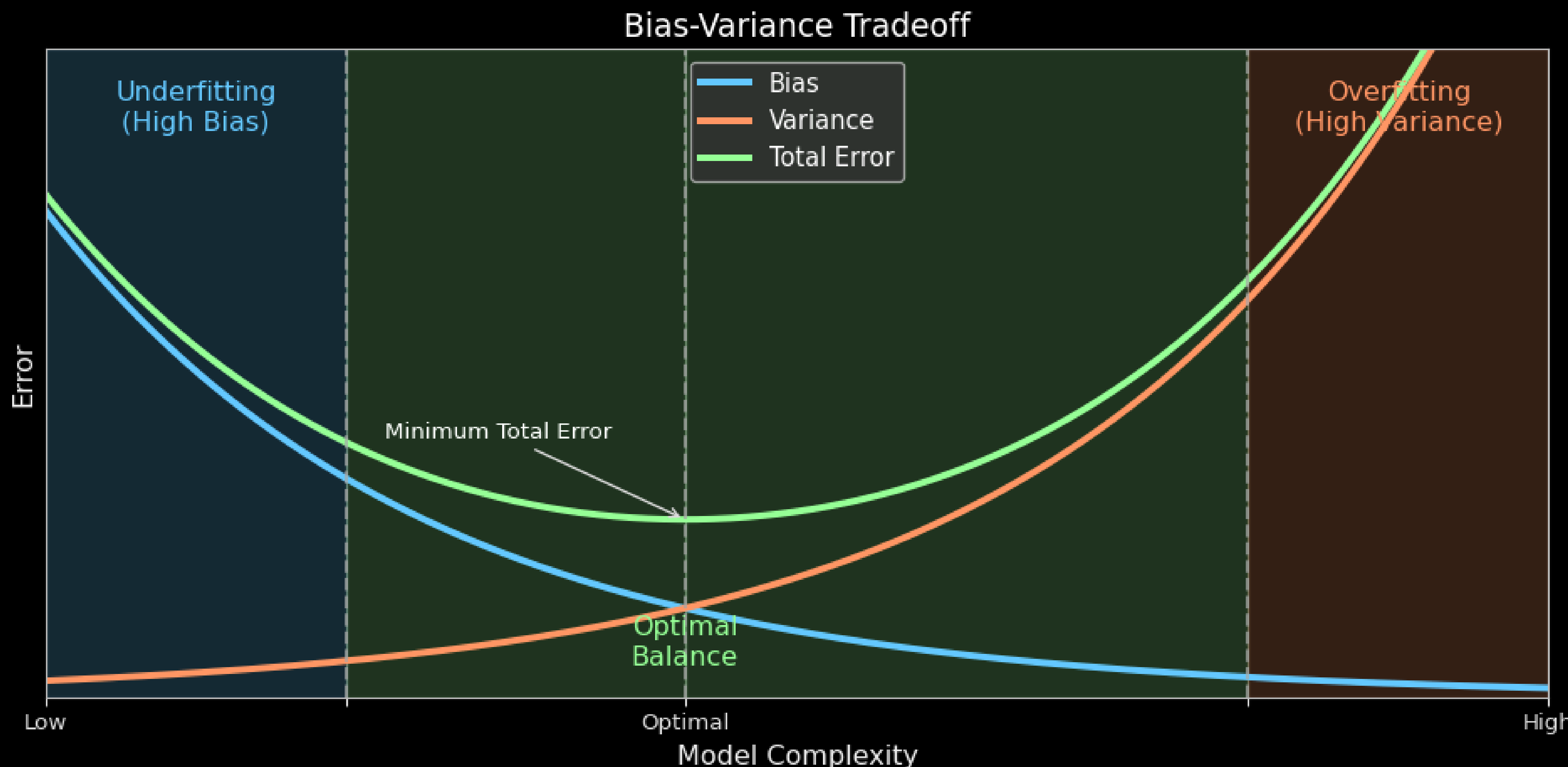
Understanding Variance

- **High Variance**: Model is too sensitive to training data
- Leads to **overfitting**
- Example: Complex polynomial fitting random noise
- Symptoms:
  - Excellent performance on training data
  - Poor performance on test data

Understanding Variance: Overfitting House Prices



Visualizing Underfitting vs. Overfitting

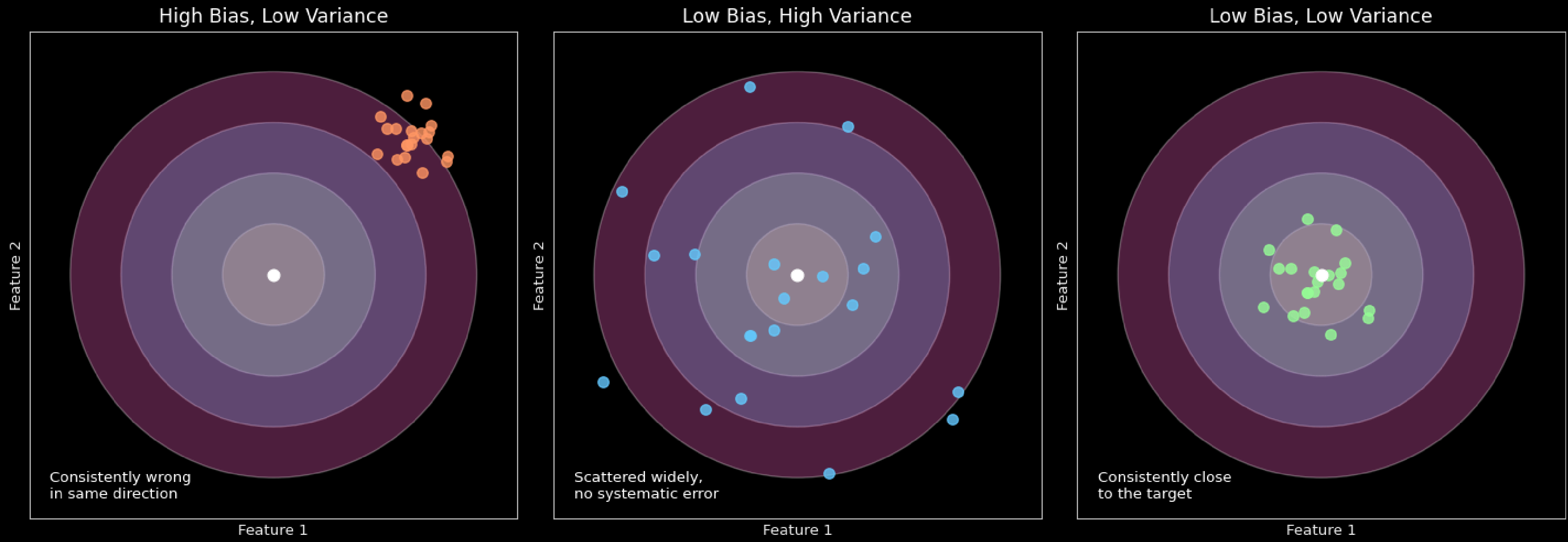


The Bullseye Analogy

- **High Bias, Low Variance**: Consistent but inaccurate shots (systematic error)
- **Low Bias, High Variance**: Scattered shots with no systematic error
- **Low Bias, Low Variance**: Accurate and precise - bullseye!

Target Bullseye

Bias-Variance Tradeoff Visualization: The Bullseye Analogy



Balancing the Tradeoff

- **Model complexity** directly affects the tradeoff
- Simple models → high bias, low variance
- Complex models → low bias, high variance
- Ways to find balance:
  - Cross-validation
  - Regularization techniques
  - Ensemble methods

Total Error Decomposition

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

- **Bias**: How far predictions are from true values
- **Variance**: How much predictions vary
- **Irreducible Error**: Noise in the problem itself

Quiz

What happens to bias and variance as model complexity increases?

A. Bias increases, variance decreases
B. Bias decreases, variance increases
C. Both bias and variance increase
D. Both bias and variance decrease

B. As model complexity increases, bias typically decreases (the model becomes more flexible and can better fit patterns) while variance increases (the model becomes more sensitive to training data fluctuations).

What We’ve Covered

In this video, we’ve discussed:

- How to measure model accuracy using various error metrics
- The fundamental bias-variance tradeoff in machine learning
- Understanding underfitting as a high bias problem
- Understanding overfitting as a high variance problem
- The bullseye analogy for visualizing bias and variance
- Strategies for finding the optimal balance in model complexity