# Polynomial Regression & Model Flexibility

Supervised Learning

Daniel E. Acuna
Associate Professor, University of Colorado Boulder

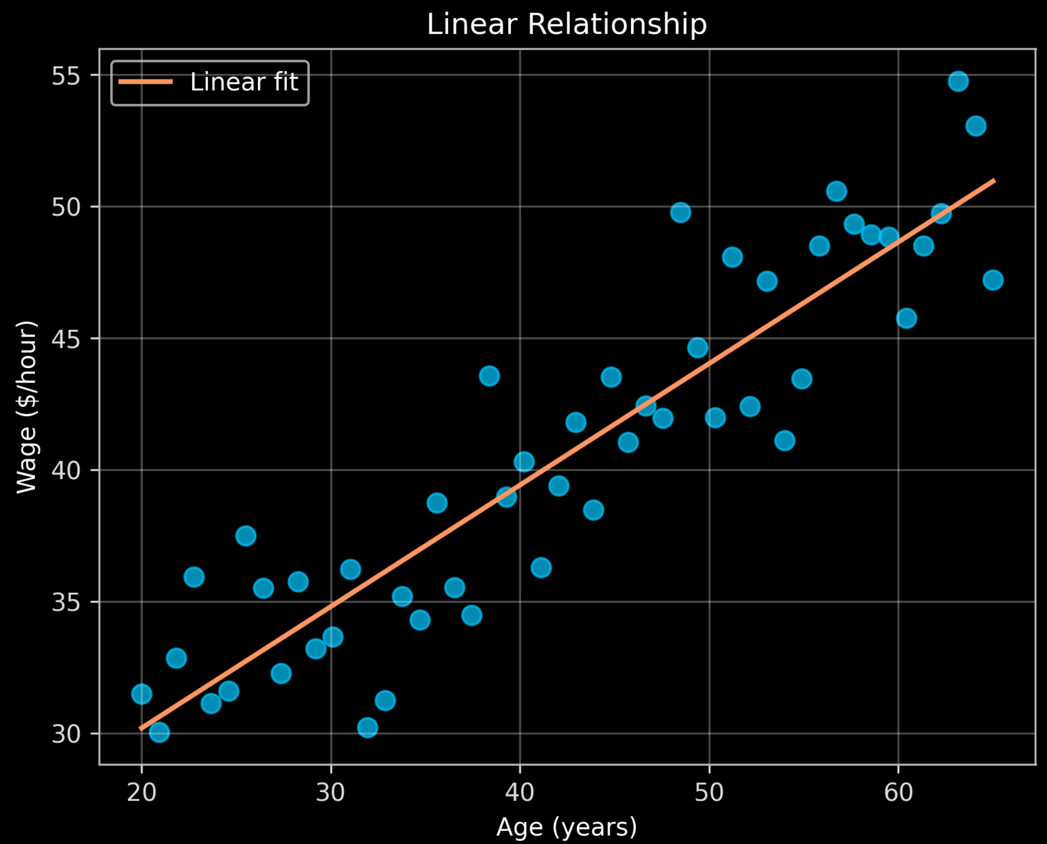
# Contents of This Video

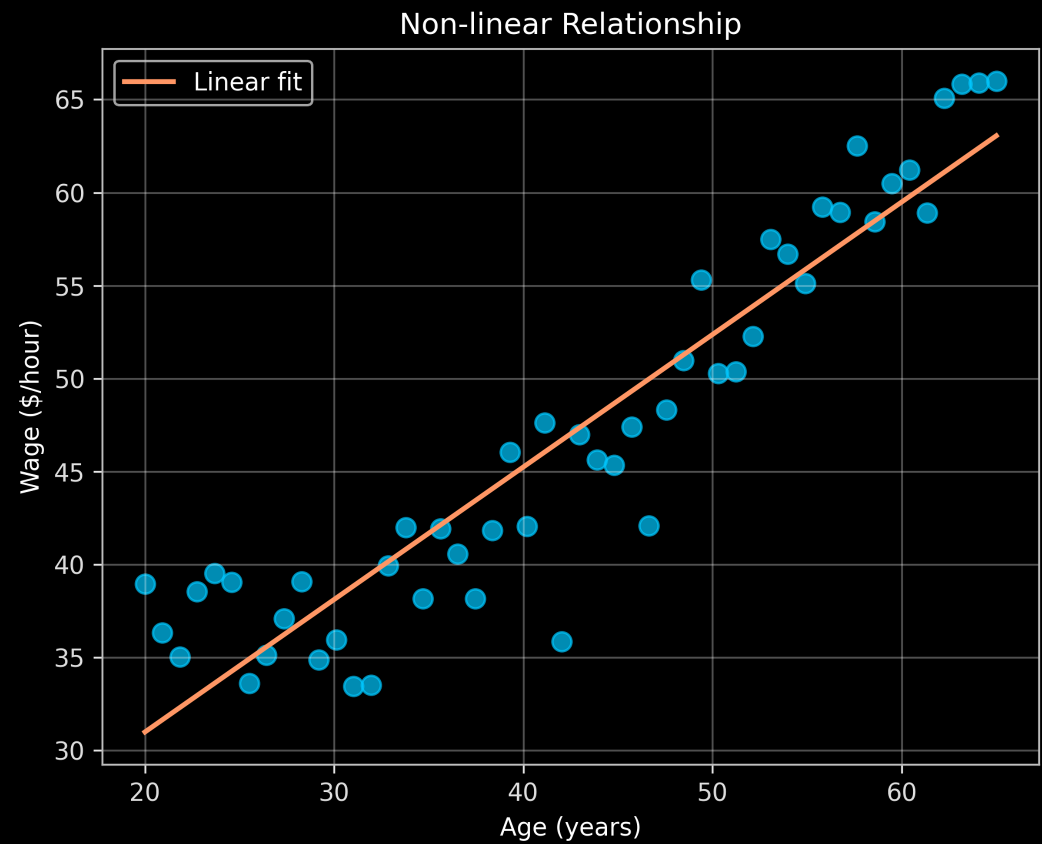
In this video, we will cover:

- Limitations of linear models for non-linear relationships
- Introduction to polynomial regression
- Mathematical formulation of polynomial models
- Example: Wage vs Age with polynomial terms
- Choosing the right polynomial degree
- Trade-off between flexibility and overfitting
- Model selection considerations

# Linear vs Non-linear Relationships

## Some relationships in real data are curved, not straight




**Wage typically rises early in career, peaks mid-career, then may decline**

# Introducing Polynomial Regression

**Extend linear regression by adding polynomial terms**

**Linear model:**

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

**Quadratic model (degree 2):**

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

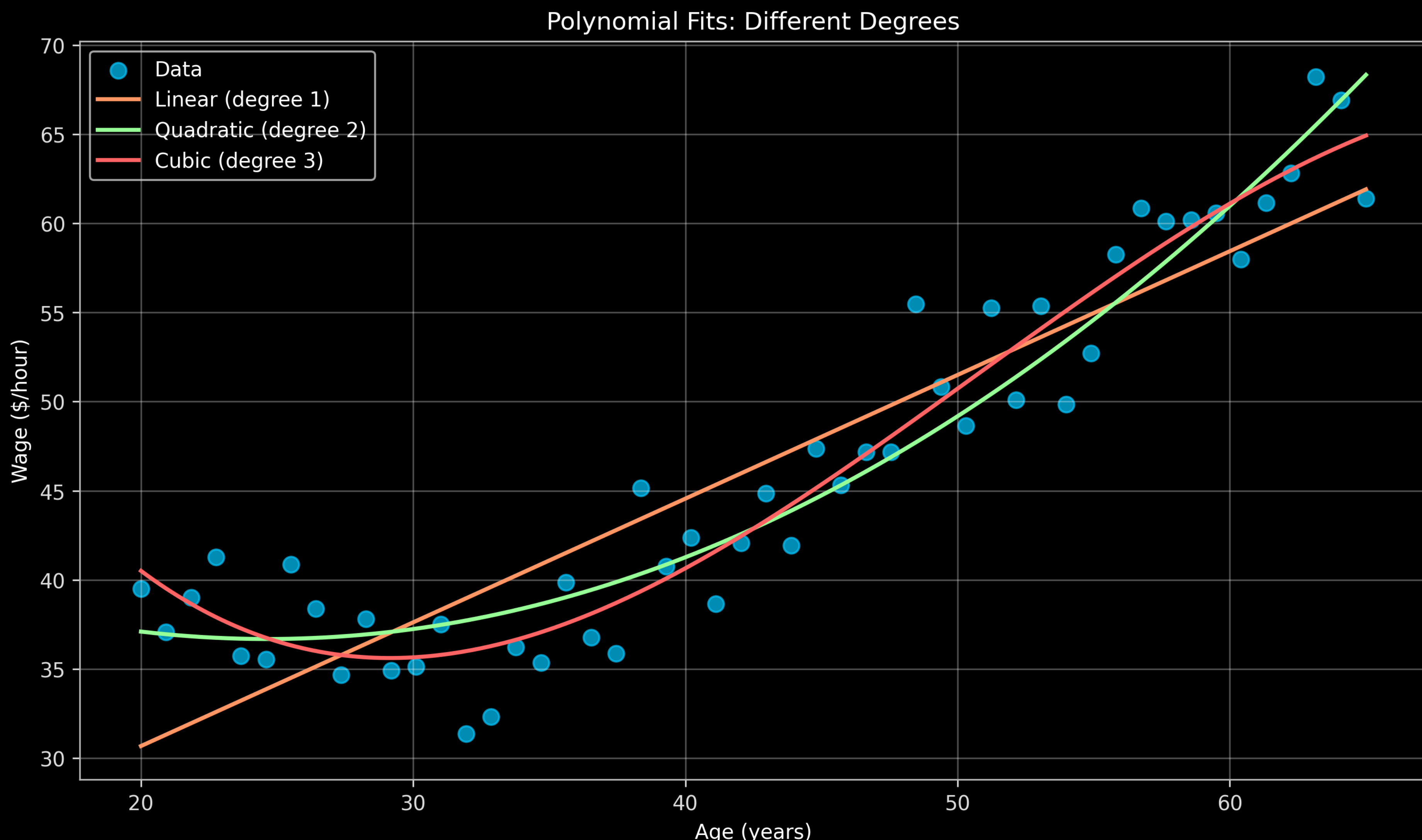
**Cubic model (degree 3):**

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon$$

**General polynomial of degree d:**

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_d X^d + \varepsilon$$

# Example: Wage vs Age with Polynomial Terms



# Interpreting Polynomial Models

**Individual coefficients are hard to interpret directly**

For cubic model: $\text{Wage} = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \beta_3 \text{Age}^3$

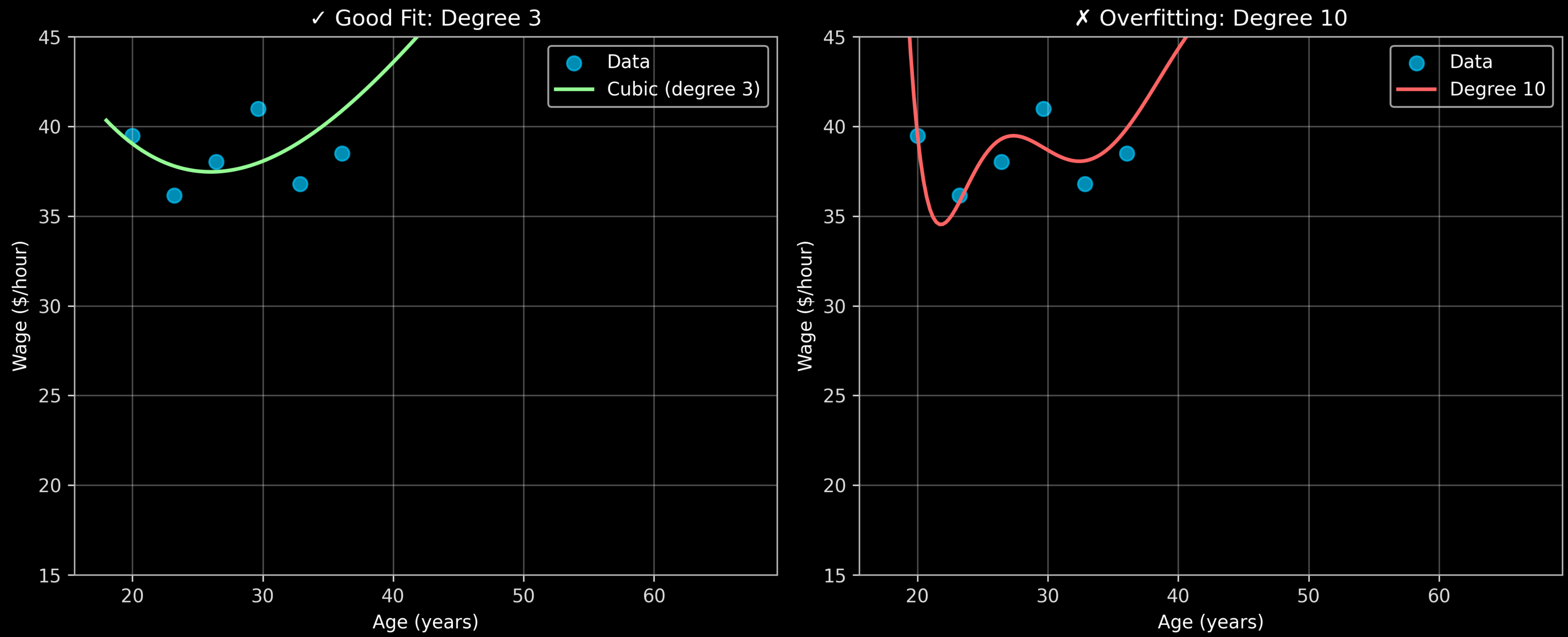
- $\beta_2$ doesn't simply mean "effect of $\text{Age}^2$"
- Coefficients work together to create the overall shape

**Better interpretation: Focus on overall pattern**

- "Wage increases with age up to around 50 years, then decreases"
- Look at the curve's shape rather than individual coefficients
- Use the model for prediction, not coefficient interpretation

# Choosing the Polynomial Degree

## Key question: What degree should we use?




**Too low:** Misses important patterns • **Too high:** Overfits to noise

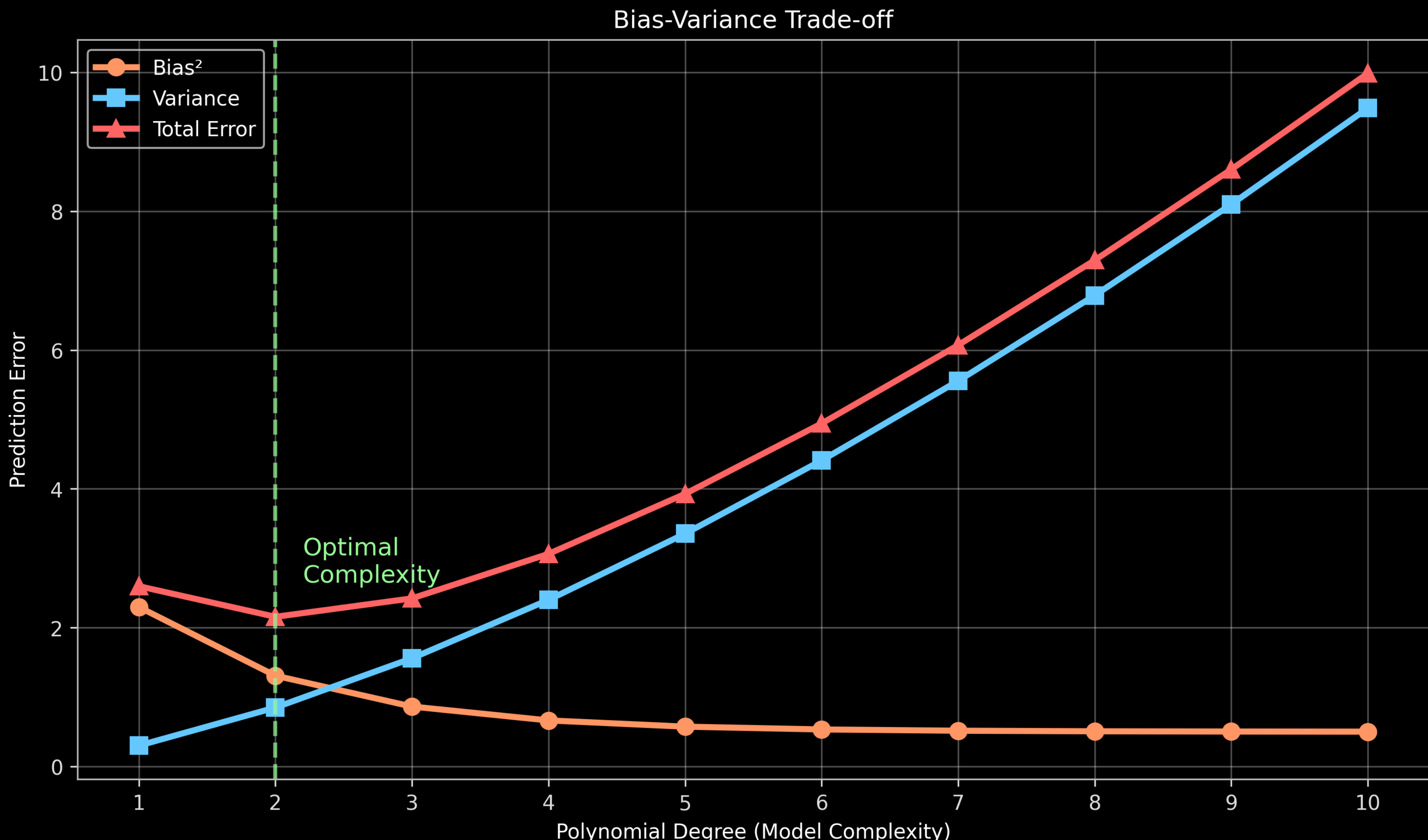
# The Overfitting Problem

## High-degree polynomials can fit noise instead of signal


## High $R^2$ on training data ≠ Good predictions on new data

# Flexibility vs Simplicity Trade-off

## The classic bias-variance trade-off in action




**Goal: Find the sweet spot that minimizes total prediction error**

# Model Selection Strategies

**How to choose the right polynomial degree?**

**Cross-validation approach:**

- Split data into training and validation sets
- Try different degrees (1, 2, 3, 4, …)
- Choose degree with best validation performance
- Test final model on separate test set

**Practical guidelines:**

- Start simple: try linear, then quadratic, then cubic
- Look for diminishing returns in performance improvement
- Consider domain knowledge about expected relationships
- Balance performance with interpretability needs

# Other Ways to Add Flexibility

**Polynomial terms are just one approach**

**Alternative methods for model flexibility:**

- **Interaction terms:** Allow effect of one variable to depend on another
- Example: Age × Education interaction in wage prediction
- **Splines:** Piecewise polynomials for local flexibility
- **Decision Trees:** Non-parametric, rule-based models
- **Neural Networks:** Highly flexible, complex relationships

**Key principle:** More flexibility always comes with overfitting risk

# Summary

**Key takeaways from polynomial regression:**

- **Extends linear regression** to capture non-linear relationships
- **Still uses least squares fitting** - just treats $X^2$, $X^3$, etc. as features
- **Interpretation changes** - focus on curve shape, not individual coefficients
- **Flexibility vs overfitting trade-off** is central to model selection
- **Cross-validation helps** choose appropriate polynomial degree
- **Balance complexity** with interpretability and generalization

**Polynomial regression provides a simple way to capture curved relationships while maintaining the linear regression framework**