Simple Linear Regression Concepts

Supervised Learning

Daniel E. Acuna
Associate Professor, University of Colorado Boulder

Contents of This Video

In this video, we will cover:

- Definition and purpose of simple linear regression
- The linear equation and its components
- Interpretation of slope and intercept coefficients
- Visual representation of the regression line
- The meaning of errors/residuals
- Using regression for prediction and inference
- Real-world applications and limitations

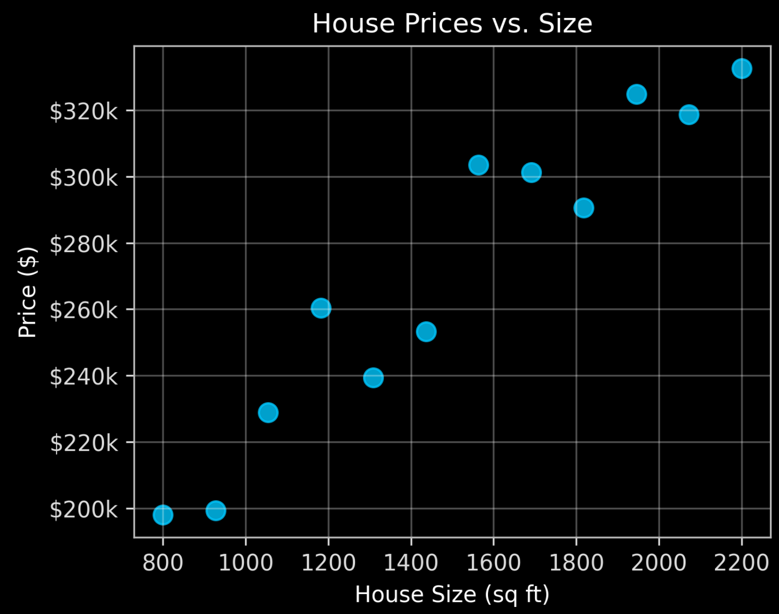
What is Linear Regression?

- A supervised learning method for predicting continuous outcomes
- Models the relationship between:
  - Input variable(s) (features/predictors)
  - Output variable (response/target)
- **Simple** linear regression: 1 input → 1 output
- **Multiple** linear regression: multiple inputs → 1 output

A Motivating Example: House Prices

Can we predict house price from size?

- Input (X): House size (square feet)
- Output (Y): House price (dollars)
- Each data point: One house sale
- Relationship: Generally, bigger houses → higher prices



The Simple Linear Regression Model

The mathematical equation:

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

Where:

- Y is the target variable (e.g., house price)
- X is the input feature (e.g., house size)
- β_0 is the intercept (Y-value when X = 0)
- β_1 is the slope (change in Y for 1-unit increase in X)
- ε (epsilon) is the error term

Understanding the Line: Visualizing β_0 and β_1

Interpreting the Coefficients

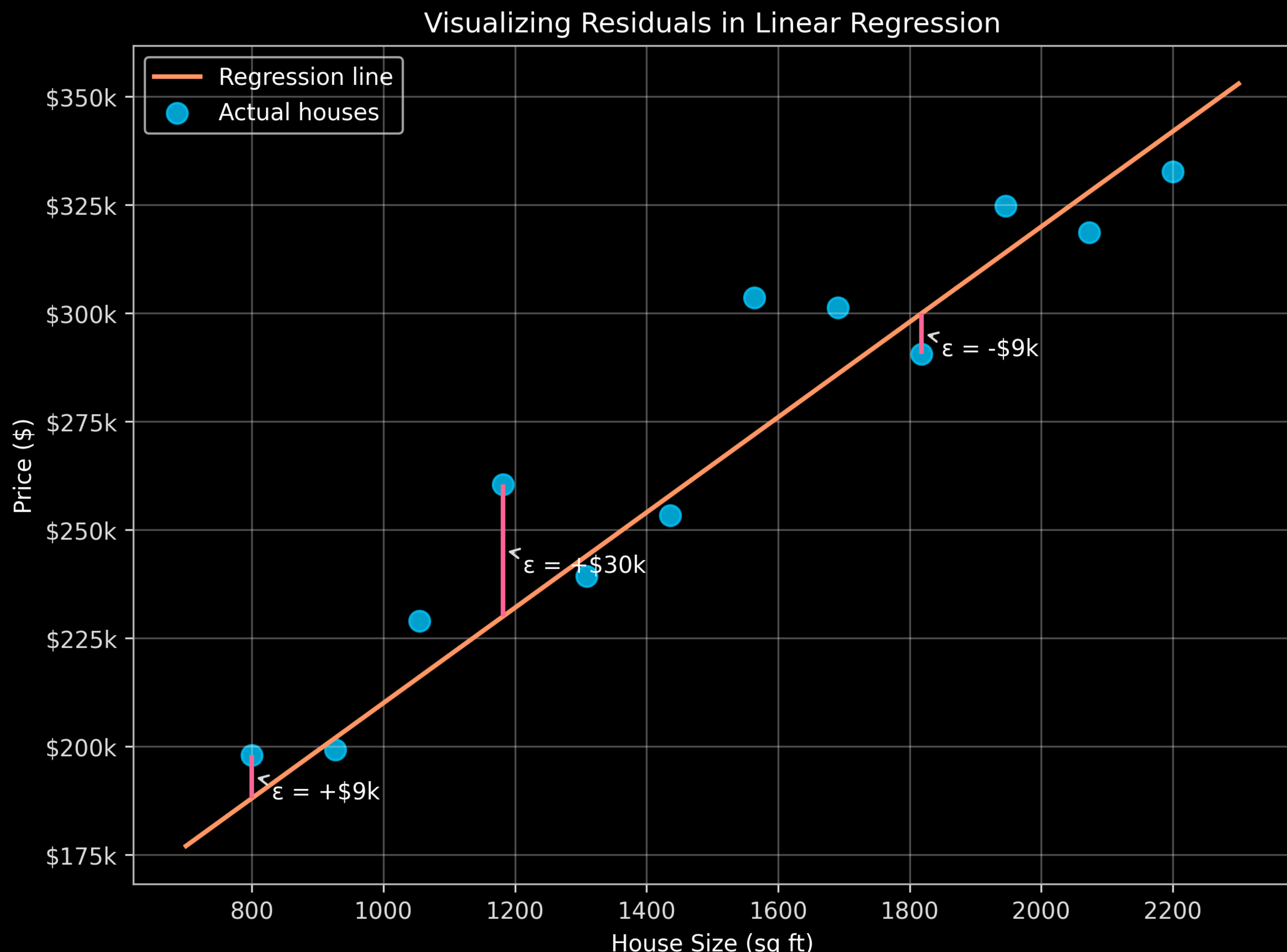
Intercept (β_0)

- Value of Y when X = 0
- In our example:
  - β_0 = \$100,000
  - Baseline house price when size = 0
  - Often not literally meaningful
  - Sets the vertical position of the line

Slope (β_1)

- Change in Y for a 1-unit increase in X
- In our example:
  - β_1 = \$110 per square foot
  - Each extra sq ft adds \$110 to price
  - Positive slope: X↑→Y↑
  - Negative slope: X↑→Y↓
  - Zero slope: X has no effect on Y

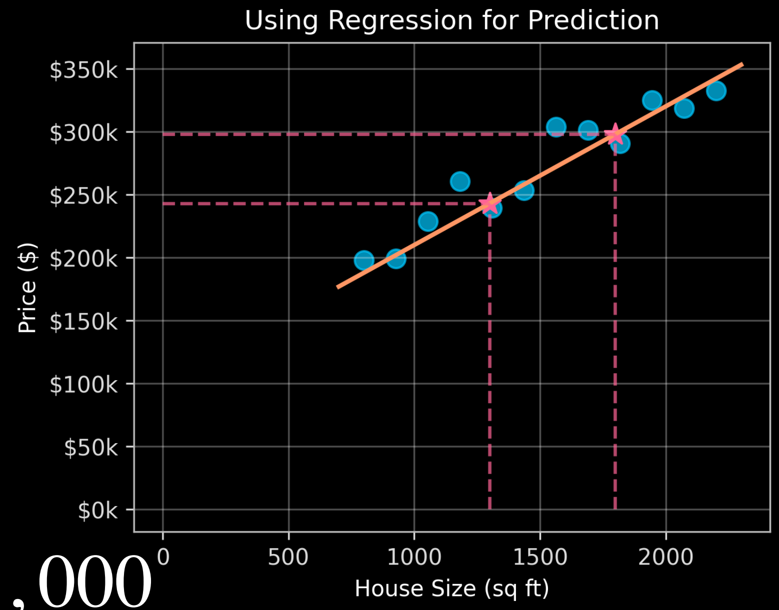
Understanding the Error Term (ε)



Using the Model for Prediction

If our fitted model is:

$$\text{Price} = \$100,000 + \$110 \times \text{Size}$$



We can predict the price of a new house:

For a 1,300 sq ft house:

$$\text{Price} = \$100,000 + \$110 \times 1,300 = \$243,000$$

For a 1,800 sq ft house:

$$\text{Price} = \$100,000 + \$110 \times 1,800 = \$298,000$$

Using the Model for Inference

Linear regression provides insights about relationships:

- **Direction**: Is the relationship positive or negative?
  - $\beta_1 > 0$: X and Y move in same direction
  - $\beta_1 < 0$: X and Y move in opposite directions
- **Magnitude**: How strong is the effect?
  - $|\beta_1|$ large: X has a strong effect on Y
  - $|\beta_1|$ small: X has a weak effect on Y
- **Practical meaning**: What does β_1 tell us about our domain?
  - In our example: Each square foot adds about $110 to home value

Applications Beyond House Prices

Linear regression is used for many prediction tasks:

- **Finance**: Predicting stock returns based on economic indicators
- **Healthcare**: Estimating patient recovery time based on treatment dosage
- **Education**: Predicting student test scores from study hours
- **Marketing**: Forecasting sales based on advertising spend
- **Environmental science**: Modeling temperature changes over time

All follow the same principle: find the best-fitting line to describe the relationship between X and Y.

Limitations of Simple Linear Regression

Assumes a linear relationship

- Real data may have nonlinear patterns
- Can miss complex relationships
- May oversimplify reality

Only uses one predictor

- Most real-world outcomes depend on multiple factors
- House prices depend on more than just size

Limited predictive power

- Single feature → more unexplained variation

Won’t capture interactions

- When the effect of one variable depends on another

Assumes constant variance

- Error may vary across the range of X

What We've Covered

In this video, we've learned:

- The basic concept and equation of simple linear regression
- How to interpret the intercept (β_0) and slope (β_1) coefficients
- The meaning of the error term (ε)
- How to visualize a regression line and its components
- Using regression for both prediction and inference
- Limitations of the simple linear model