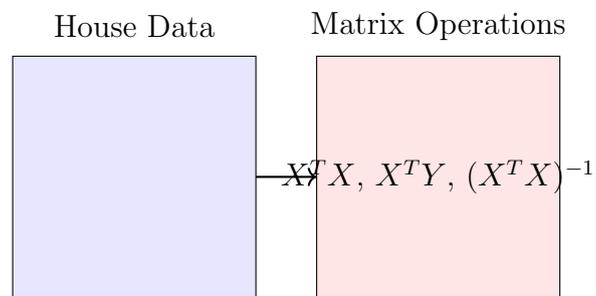


# Real-World Example: House Price Prediction

Understanding  $X$ ,  $X^T X$ ,  $X^T Y$ , and  $(X^T X)^{-1}$



# 1 Real-World Example: Predicting House Prices

Let's consider a real-world example of predicting house prices based on two features: size (square feet) and number of bedrooms.

## 1.1 The Dataset

We have data for 5 houses:

House	Size (sq. ft.)	Bedrooms	Price (\$1000)
House 1	1500	3	300
House 2	2000	4	400
House 3	2500	3	500
House 4	3000	5	600
House 5	3500	4	700

## 1.2 Visualizing the Data

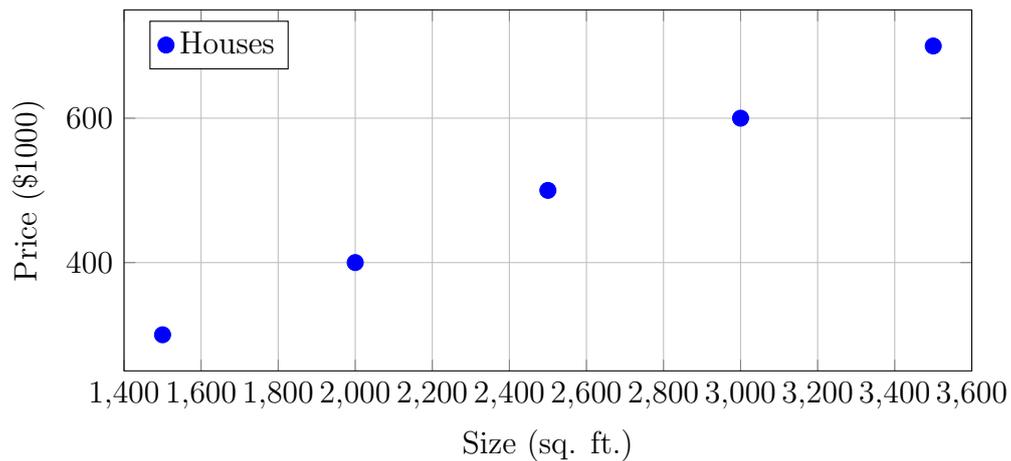


Figure 1: Relationship between house size and price

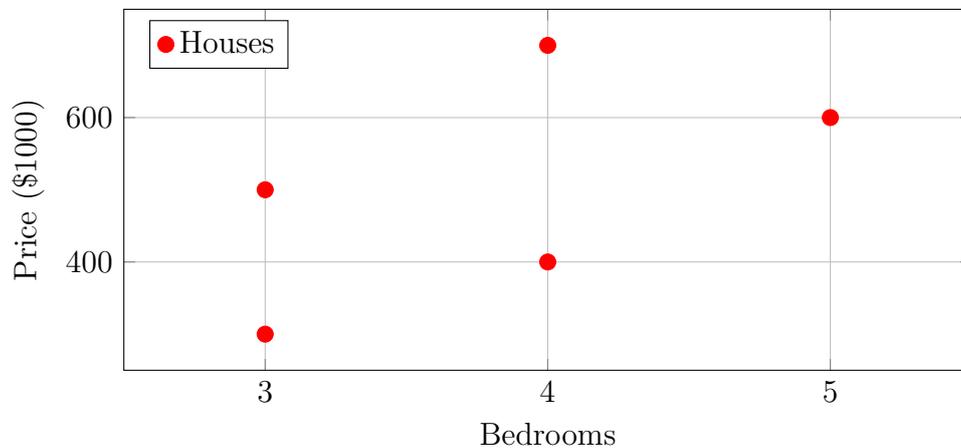


Figure 2: Relationship between bedrooms and price

### 1.3 The Design Matrix $X$ and Target Vector $y$

We create the design matrix  $X$  with a column of 1s for the intercept term:

$$X = \begin{bmatrix} 1 & 1500 & 3 \\ 1 & 2000 & 4 \\ 1 & 2500 & 3 \\ 1 & 3000 & 5 \\ 1 & 3500 & 4 \end{bmatrix} \quad y = \begin{bmatrix} 300 \\ 400 \\ 500 \\ 600 \\ 700 \end{bmatrix}$$

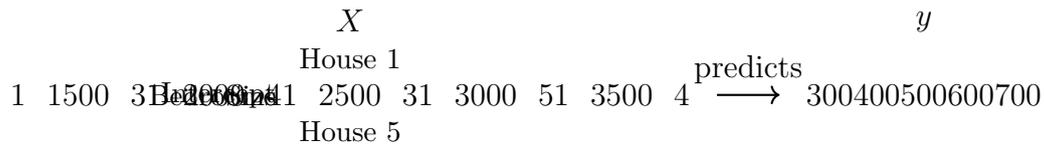


Figure 3: Design matrix  $X$  and target vector  $y$

### 1.4 Centering the Data

To make the interpretation easier, let's center the data by subtracting the mean from each feature:

Mean size = 2500, Mean bedrooms = 3.8, Mean price = 500

$$X_{\text{centered}} = \begin{bmatrix} 1 & -1000 & -0.8 \\ 1 & -500 & 0.2 \\ 1 & 0 & -0.8 \\ 1 & 500 & 1.2 \\ 1 & 1000 & 0.2 \end{bmatrix} \quad y_{\text{centered}} = \begin{bmatrix} -200 \\ -100 \\ 0 \\ 100 \\ 200 \end{bmatrix}$$

## 2 Calculating $X^T X$

Let's compute  $X^T X$  for our centered data:

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1000 & -500 & 0 & 500 & 1000 \\ -0.8 & 0.2 & -0.8 & 1.2 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & -1000 & -0.8 \\ 1 & -500 & 0.2 \\ 1 & 0 & -0.8 \\ 1 & 500 & 1.2 \\ 1 & 1000 & 0.2 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2,500,000 & 600 \\ 0 & 600 & 3.2 \end{bmatrix}$$

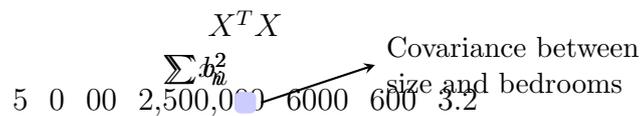


Figure 4:  $X^T X$  matrix with annotations

### 2.1 Interpreting $X^T X$

- The diagonal elements represent the variances of the features:
  - 5: This is the number of observations (the intercept column is all 1s)
  - 2,500,000: This is the sum of squared deviations for size (variance  $\times$  n)
  - 3.2: This is the sum of squared deviations for bedrooms (variance  $\times$  n)
- The off-diagonal elements represent the covariances between features:
  - 0: The intercept column is orthogonal to the centered features
  - 600: This is the covariance between size and bedrooms (covariance  $\times$  n)
- The near-zero off-diagonals with the intercept show that our centering worked correctly

### 3 Calculating $X^T Y$

Let's compute  $X^T Y$  for our centered data:

$$X^T Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1000 & -500 & 0 & 500 & 1000 \\ -0.8 & 0.2 & -0.8 & 1.2 & 0.2 \end{bmatrix} \begin{bmatrix} -200 \\ -100 \\ 0 \\ 100 \\ 200 \end{bmatrix}$$

$$X^T Y = \begin{bmatrix} 0 \\ 500,000 \\ 320 \end{bmatrix}$$

$X^T$					$Y$					$X^T Y$							
1	1	1	1	1	-1000	-500	0	500	1000	-0.8	0.2	-0.8	1.2	0.2	0	500,000	320

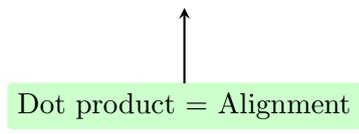


Figure 5: Visualization of  $X^T Y$  calculation

#### 3.1 Interpreting $X^T Y$

- The first element (0): The sum of the centered target values (should be zero by design)
- The second element (500,000): The covariance between size and price (covariance  $\times n$ )
- The third element (320): The covariance between bedrooms and price (covariance  $\times n$ )
- These values represent how much each feature "pulls" in the direction of the target

## 4 Calculating $(X^T X)^{-1}$

Let's compute  $(X^T X)^{-1}$  for our centered data:

$$X^T X = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2,500,000 & 600 \\ 0 & 600 & 3.2 \end{bmatrix}$$

Where  $D = (2,500,000)(3.2) - (600)^2 = 8,000,000 - 360,000 = 7,640,000$

$$(X^T X)^{-1} = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & \frac{3.2}{D} & \frac{-600}{D} \\ 0 & \frac{-600}{D} & \frac{2,500,000}{D} \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.000000418 & -0.0000785 \\ 0 & -0.0000785 & 0.327 \end{bmatrix}$$

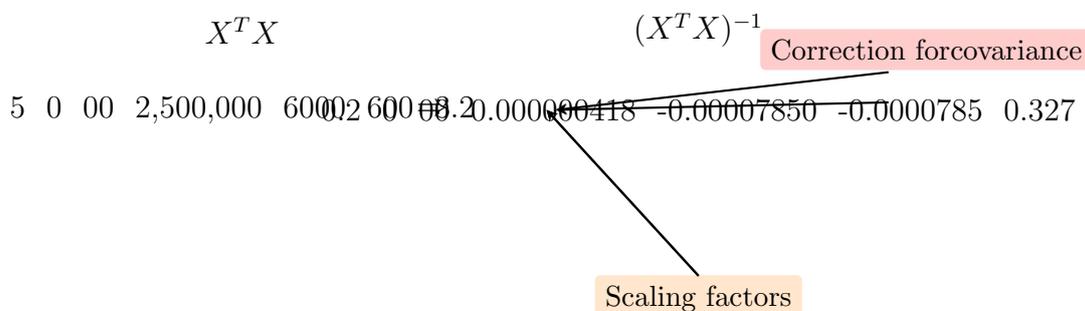


Figure 6: Visualization of  $(X^T X)^{-1}$  as a scaling and correction matrix

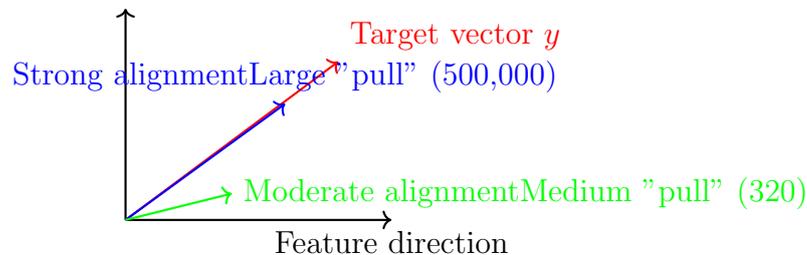
### 4.1 Interpreting $(X^T X)^{-1}$

- The diagonal elements represent the scaling factors for each feature:
  - 0.2: Scaling for the intercept ( $1/5$ )
  - 0.000000418: Scaling for size ( $1/2,500,000$  with covariance adjustment)
  - 0.327: Scaling for bedrooms ( $1/3.2$  with covariance adjustment)
- The off-diagonal elements correct for the covariance between features:
  - -0.0000785: Correction factor for the size-bedrooms relationship
- The small value for size scaling reflects its large variance
- The larger value for bedrooms scaling reflects its smaller variance

## 5 Detailed Explanation of Key Concepts

### 5.1 What Does "Pull" Mean?

The term "pull" refers to how much a feature aligns with the target direction. It's measured by the covariance between the feature and target, with larger values indicating stronger relationships.



Dot product measures alignment between feature and target  
Larger values indicate stronger relationships

Figure 7: Visualizing the "pull" concept through vector alignment

In our house price example: - Size has a strong pull (500,000) because it aligns well with price - Bedrooms have a moderate pull (320) because they align somewhat with price - The intercept has no pull (0) because it's orthogonal to price changes

### 5.2 What Is the "Covariance Adjustment"?

The covariance adjustment accounts for correlations between features. It ensures each coefficient represents only the unique contribution of that feature, not shared effects with other features.

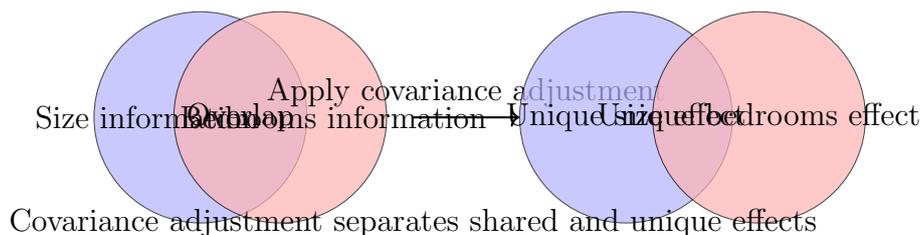


Figure 8: Covariance adjustment separates shared information between features

In our example: - Size and bedrooms are correlated (covariance = 600) - Without adjustment, we'd double-count their effects on price - The covariance adjustment ensures each feature gets credit only for its unique contribution

### 5.3 What Does the Correction Factor -0.0000785 Do?

The correction factor -0.0000785 specifically corrects for the correlation between size and bedrooms, ensuring each feature gets credit only for its unique contribution to predicting price.

Correction factors for  
size-bedrooms correlation

These values ensure each feature gets credit only for its unique effect

Figure 9: The role of the correction factor in the coefficient calculation

The correction factor works by: 1. Measuring how much size and bedrooms overlap in their information 2. Subtracting the shared effect from each feature's coefficient 3. Ensuring the final coefficients represent only unique contributions

## 6 Calculating the Coefficients $\beta$

Now let's compute the coefficients using the normal equation:

$$\beta = (X^T X)^{-1} X^T Y$$

$$\beta = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.000000418 & -0.0000785 \\ 0 & -0.0000785 & 0.327 \end{bmatrix} \begin{bmatrix} 0 \\ 500,000 \\ 320 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 0 \\ (0.000000418)(500,000) + (-0.0000785)(320) \\ (-0.0000785)(500,000) + (0.327)(320) \end{bmatrix}$$

$$\beta = \begin{bmatrix} 0 \\ 0.209 - 0.02512 \\ -39.25 + 104.64 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.18388 \\ 65.39 \end{bmatrix}$$

$(X^T X)^{-1}$	$X^T Y$	$=$	$\beta$
0.2 0 00	0.000000418 -0.00007850 -0.00007850 0.327		00.1838865.39

Final coefficients

Figure 10: Calculation of coefficients  $\beta$

### 6.1 Interpreting the Coefficients

- $\beta_0 = 0$ : The intercept is zero because we centered the data
- $\beta_1 = 0.18388$ : For each additional square foot, the price increases by \$183.88
- $\beta_2 = 65.39$ : For each additional bedroom, the price increases by \$65,390
- These coefficients make sense intuitively: size has a bigger impact on price than bedrooms

### 6.2 The Original Scale

To get the coefficients on the original scale, we need to account for the centering:

$$\begin{aligned} \text{Original intercept} &= \beta_0 - \beta_1 \cdot \text{mean size} - \beta_2 \cdot \text{mean bedrooms} \\ &= 0 - 0.18388 \cdot 2500 - 65.39 \cdot 3.8 = -459.7 - 248.48 = -708.18 \end{aligned}$$

So the prediction equation is:

$$\text{Price} = -708.18 + 0.18388 \cdot \text{Size} + 65.39 \cdot \text{Bedrooms}$$