

Linear Regression Example: Apple Price Prediction

Aymen Negadi

Python Implementation

Listing 1: Linear Regression with scikit-learn

```
1 import numpy as np
2 from sklearn.linear_model import LinearRegression
3
4 # Data: Weight (grams) and Price (dollars)
5 X = np.array([100, 150, 200, 250, 300]).reshape(-1, 1)
6 y = np.array([2, 3, 4, 5, 6])
7
8 # Create and train the model
9 model = LinearRegression()
10 model.fit(X, y)
11
12 # Predict price for 180 grams
13 prediction = model.predict([[180]])
14
15 print("Predicted price:", prediction[0])
```

Mathematical Explanation

1. Data

$$X = [100, 150, 200, 250, 300] \quad (\text{weight in grams})$$

$$y = [2, 3, 4, 5, 6] \quad (\text{price in dollars})$$

We want the equation of a straight line:

$$y = aX + b$$

where a is the slope and b is the intercept.

2. Calculate the Slope a

The slope in simple linear regression:

$$a = \frac{\text{Cov}(X, y)}{\text{Var}(X)}$$

First, calculate the means:

$$\bar{X} = \frac{100 + 150 + 200 + 250 + 300}{5} = 200$$

$$\bar{y} = \frac{2 + 3 + 4 + 5 + 6}{5} = 4$$

Covariance Calculation

$$\text{Cov}(X, y) = \frac{1}{n} \sum (X_i - \bar{X})(y_i - \bar{y})$$

X_i	y_i	$X_i - \bar{X}$	$y_i - \bar{y}$	Product
100	2	-100	-2	200
150	3	-50	-1	50
200	4	0	0	0
250	5	50	1	50
300	6	100	2	200
Sum:				500

$$\text{Cov}(X, y) = \frac{500}{5} = 100$$

Variance Calculation

$$\text{Var}(X) = \frac{1}{n} \sum (X_i - \bar{X})^2$$

$X_i - \bar{X}$	Square
-100	10000
-50	2500
0	0
50	2500
100	10000
Sum:	25000

$$\text{Var}(X) = \frac{25000}{5} = 5000$$

Slope

$$a = \frac{100}{5000} = 0.02$$

3. Calculate the Intercept b

$$b = \bar{y} - a\bar{X} = 4 - (0.02 \times 200) = 4 - 4 = 0$$

4. Regression Equation

$$y = 0.02X + 0 \quad \text{or simply} \quad y = 0.02X$$

Prediction

To predict the price of an apple weighing 180 grams:

$$y = 0.02 \times 180 = 3.6$$

So the predicted price is \$3.60.

Verification with Python Output

The Python code outputs: Predicted price: 3.6, which matches our mathematical calculation.