



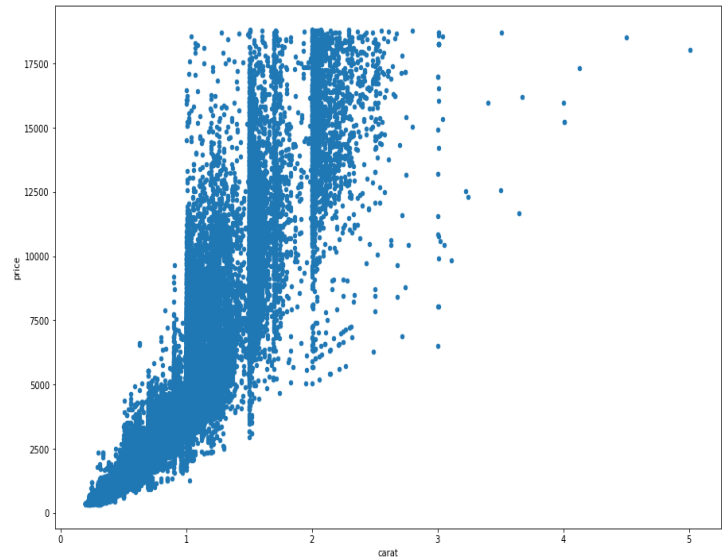
M6-03: Linear Regression

Part of the "Towards Machine Learning" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m6-03/>

REGRESSION AND THE EQUATION OF THE REGRESSION LINE

Linear regression: The idea of trying to fit a line as closely as possible to as many points as possible is known as linear regression. The most common technique is to try to fit a line that minimizes the squared distance to each of those points. This is called OLS or Ordinary Least Squares Regression:



Regression

One way to make predictions is by using the *regression equation*.

Since our regression estimates form a straight line, we can describe them using an equation in slope-intercept form:

$$\hat{y} = b_0 + b_1 * x_1$$

$$\text{slope of the regression line} = r \cdot \frac{\text{SD of } y}{\text{SD of } x}$$

intercept of the regression line = average of y – slope • average of x

Interpretations

SLOPE= The average increase in Y associated with a 1-unit increase in X .

Y-INTERCEPT= The predicted value of Y when X is equal to 0.



M6-03: Linear Regression

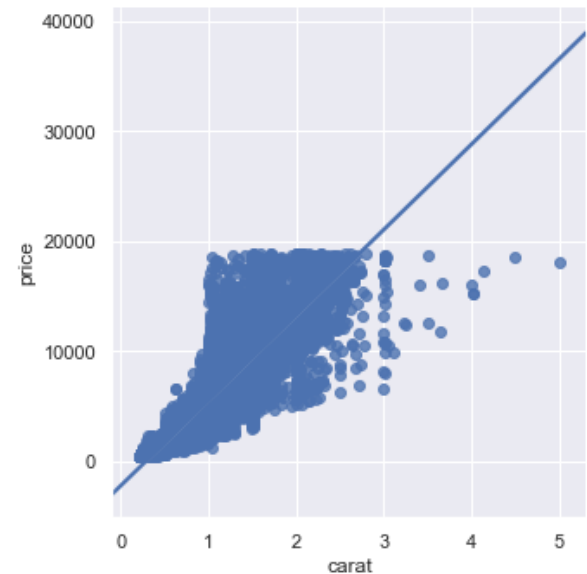
Part of the "Towards Machine Learning" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m6-03/>

Puzzle #1:

Let's go ahead and find the slope and y-intercept of the regression equation using these 5 summary statistics (rounded):

r= 0.92	Average	SD
Carat	0.80	0.50
Price	3933	3989
Slope (b1):		
y-intercept (b0):		
Regression Equation:		



```
df = pd.read_csv('diamonds.csv')  
sns.lmplot(x='carat', y='price', data=df)
```

Use the regression equation on the previous page to predict the price of a diamond that is 2.5 carats.

How do you interpret the slope and the y-intercept in the above equation?

Warning About Regression:

- Predicting y at values of x beyond the range of x in the data is called **extrapolation**.
- This is risky because we have no evidence to believe that the association between x and y remains linear for unseen values of x.



M6-03: Linear Regression

Part of the "Towards Machine Learning" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m6-03/>

- Extrapolated predictions can be absolutely wrong.