DSC 40A

Theoretical Foundations of Data Science I

Least Squares Solutions

- The **least squares solutions** for the slope w_1 and intercept w_0 are:

$$w_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad w_0 = \bar{y} - w_1\bar{x}$$

where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

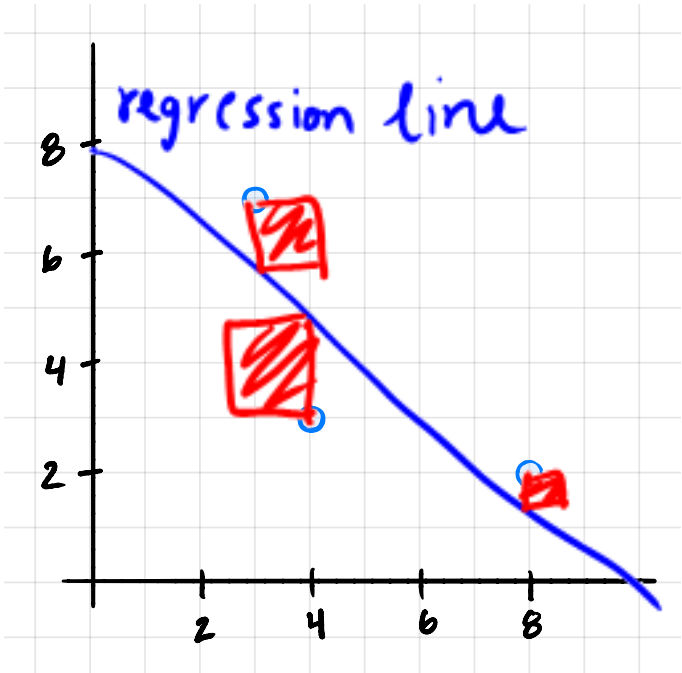
In This Video

We'll do an example and interpret the least squares solutions.

Recommended Reading

Course Notes: Chapter 2, Section 1

Example



$\bar{x} = 5$

$\bar{y} = 4$

$$w_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = -11/14$$

$$w_0 = \bar{y} - w_1\bar{x} = 4 - \left(-11/14\right) \cdot 5 = 4 + 55/14 \approx 8$$

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
3	7	-2	3	-6	4
4	3	-1	-1	1	1
8	2	3	-2	-6	9
				sum -11	sum 14

Interpretation of Intercept

$$w_0 = \bar{y} - w_1 \bar{x}$$

- What is $H(\bar{x})$?

$$H(x) = w_1 x + w_0$$

$$H(\bar{x}) = w_1 \bar{x} + w_0$$

$$= w_1 \bar{x} + \bar{y} - w_1 \bar{x}$$

$$= \bar{y}$$

intercept

Question

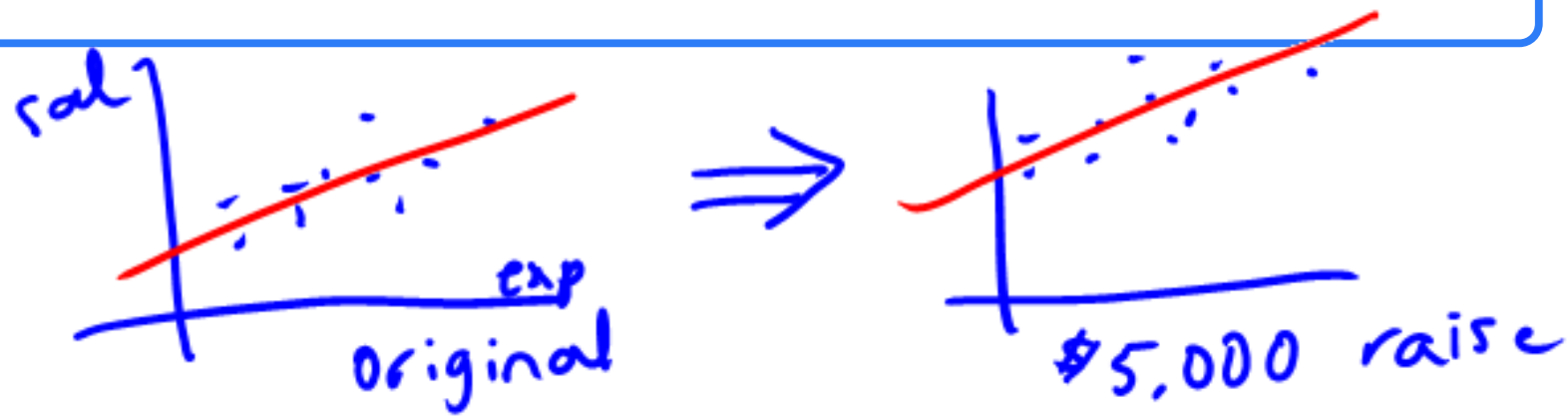
We fit a linear prediction rule for salary given years of experience. Then everyone gets a $5,000 raise. Which of these happens?

a) slope increases, intercept increases

b) slope decreases, intercept increases

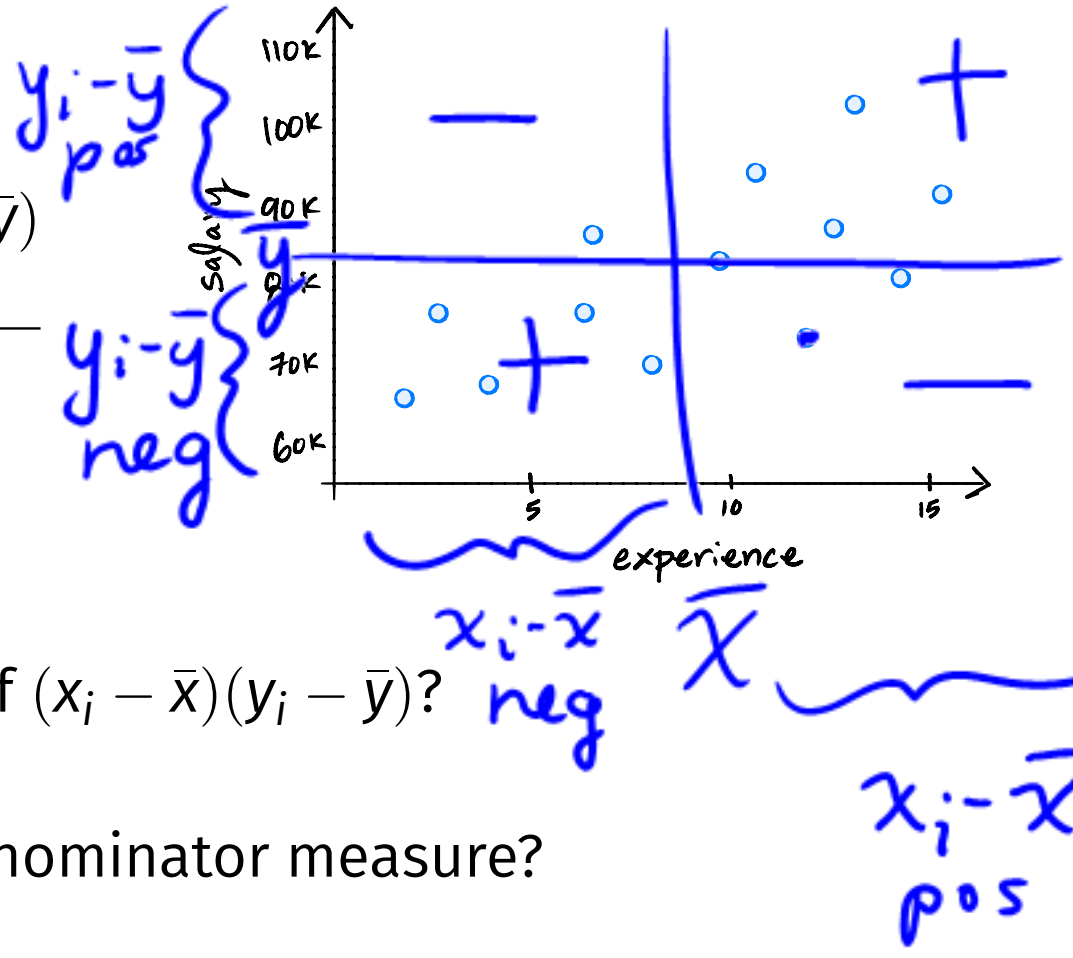
c) slope stays same, intercept increases

d) slope stays same, intercept stays same



Interpretation of Slope

$$w_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$



- What is the sign of $(x_i - \bar{x})(y_i - \bar{y})$?
- What does the denominator measure?

What's next?

- Using linear regression formulas to fit certain special nonlinear functions to data.
- Generalizing to arbitrary polynomials.
- Generalizing to multiple predictor variables.