

Exercise 5

Social Data Science

1 Linear Regression

- a) Suppose we are fitting a linear regression model to predict an individual's weight w in kilograms using their height h in centimeters as input. Specifically, we fit a model:

$$w = \beta_0 + \beta_1 \cdot h .$$

The fitted model estimates $\hat{\beta}_0 = -20$ and $\hat{\beta}_1 = .6$. Interpret these coefficients by drawing the model predicted relationship between height and weight. According to the model, how much heavier can we expect an individual who is 200cm to be than one who is 150 cm?

- b) Suppose we add gender as a term in our model. Specifically, the feature m is 1 if the individual is male, 0 if female. Thus, our model from a) extends to:

$$w = \beta_0 + \beta_1 \cdot h + \beta_2 \cdot m .$$

The fitted model estimates $\hat{\beta}_0 = -34$ and $\hat{\beta}_1 = .5$ and $\hat{\beta}_2 = -5$. Interpret these coefficients by drawing the model predicted relationship between height and weight for men and women (Hint: draw two separate lines). Interpret in words the estimated $\hat{\beta}_2$ coefficient.

2 Fitted Values vs Response Values

Consider the fitted values that result from performing linear regression without an intercept. In this setting, the i -th fitted value takes the form

$$\hat{y}_i = x_i \hat{\beta},$$

where

$$\hat{\beta} = \frac{(\sum_{i=1}^n x_i y_i)}{\sum_{i=1}^n x_i^2}.$$

Show that we can write

$$\hat{y}_i = \sum_{j=1}^n a_j y_j,$$

and give an expression for a_i .

3 The R^2 Statistic

Again consider the case of a simple linear regression of Y onto X . Show that in this case, the R^2 statistic is equal to the square of the correlation between X and Y . You may assume that $\bar{x} = \bar{y} = 0$.