Properties of Simple Linear Regression

1. (7 points) In Lecture 12, we spent a great deal of time talking about simple linear regression, which you also saw in Data 8. To briefly summarize, the simple linear regression model assumes that given a single observation x, our predicted response for this observation is $\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x$.

   In Lecture 12, we saw that the $\hat{\theta}_0$ and $\hat{\theta}_1$ that minimize the average L_2 loss for the simple linear regression model are:

$$\hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x}$$
$$\hat{\theta}_1 = r\frac{\sigma_y}{\sigma_x}$$

   Or, rearranging terms, our predictions $\hat{y}$ are:

$$\hat{y} = \bar{y} + r\sigma_y \frac{x - \bar{x}}{\sigma_x}$$

   (a) (3 points) As we saw in lecture, a residual e_i is defined to be the difference between a true y_i and predicted $\hat{y}_i$. Specifically, $e_i = y_i - \hat{y}_i$.
   Prove, using the equation for $\hat{y}$ above, that $\sum_{i=1}^{n} e_i = 0$.

$$\sum e_i = \sum (y_i - \hat{y}_i)$$
$$= \sum y_i - \sum \hat{y}_i$$
$$= \sum y_i - \sum \left(\bar{y} + r\sigma_y \frac{x_i - \bar{x}}{\sigma_x}\right)$$
$$= \left(\sum y_i - n\bar{y}\right) + \frac{r\sigma_y}{\sigma_x}\left(\sum x_i - n\bar{x}\right)$$
$$= \left(\sum y_i - \sum y_i\right) + \frac{r\sigma_y}{\sigma_x}\left(\sum x_i - \sum x_i\right) \quad \text{since } \bar{y} = \frac{1}{n}\sum y_i, \ \bar{x} = \frac{1}{n}\sum x_i$$
$$= 0 + 0$$
$$= 0.$$

(b) (2 points) Using your result from part a, prove that $\bar{y} = \bar{\hat{y}}$.

$$\bar{\hat{y}} = \frac{1}{n}\sum \hat{y}$$

$$= \frac{1}{n}\sum(\hat{\theta}_0 + \hat{\theta}_1 x_i)$$

$$= \hat{\theta}_0 + \hat{\theta}_1 \bar{x}$$

$$= \bar{y} - \hat{\theta}_1 \bar{x} + \hat{\theta}_1 \bar{x}$$

$$= \bar{y}$$

(c) (2 points) Prove that $(\bar{x}, \bar{y})$ is on the simple linear regression line.

Since $\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x$, $\hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x}$

Then $\hat{y} = \bar{y} - \hat{\theta}_1 \bar{x} + \hat{\theta}_1 x$.

When $x = \bar{x}$:

$\hat{y} = \bar{y} - \hat{\theta}_1 \bar{x} + \hat{\theta}_1 \bar{x} = \bar{y}$

Thus, the line passes thru $(\bar{x}, \bar{y})$.

Geometric Perspective of Least Squares

2. (7 points) In Lecture 13, we viewed both the simple linear regression model and the multiple linear regression model through the lens of linear algebra. The key geometric insight was that if we train a model on some design matrix $\mathbb{X}$ and true response vector $\mathbb{Y}$, our predicted response $\hat{\mathbb{Y}} = \mathbb{X}\hat{\theta}$ is the vector in span($\mathbb{X}$) that is closest to $\mathbb{Y}$.

   In the simple linear regression case, our optimal vector θ is $\hat{\theta} = [\hat{\theta}_0, \hat{\theta}_1]^T$, and our design matrix is

$$\mathbb{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} | & | \\ \mathbb{1} & \vec{x} \\ | & | \end{bmatrix}$$

   This means we can write our predicted response vector as $\hat{\mathbb{Y}} = \mathbb{X}\begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{bmatrix}$, and also as $\hat{\mathbb{Y}} = \hat{\theta}_0\mathbb{1} + \hat{\theta}_1\vec{x}$.

   Note, in this problem, $\vec{x}$ refers to the n-length vector $[x_1, x_2, ..., x_n]^T$. In other words, it is a feature, not an observation.

   For this problem, assume we are working with the simple linear regression model, though the properties we establish here hold for any linear regression model that contains an intercept term.

   (a) (3 points) Using the geometric properties from lecture, prove that $\sum_{i=1}^{n} e_i = 0$. *Hint: Recall, we define the residual vector as $e = \mathbb{Y} - \hat{\mathbb{Y}}$, and $e = [e_1, e_2, ..., e_n]^T$.*

From the lecture, we know that if θ is the optimal, then $\vec{e}$ is orthogonal to span(X).

In particular, $(1 \cdots 1)^T$ is in span(X).

$\Rightarrow \quad \vec{e} \cdot \vec{1} = (1 \quad 1 \quad \cdots \quad 1)\begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \sum e_i = 0$

Thus, $\sum e_i = 0$.

(b) (2 points) Explain why the vector $\vec{x}$ (as defined in the problem) and the residual vector e are orthogonal. *Hint: Two vectors are orthogonal if their dot product is 0.*

Since $\vec{e}$ is orthogonal to span $(\vec{1}, \vec{x})$ and $\vec{x} \in$ span $(\vec{1}, \vec{x})$.

$\vec{e}$ is orthogonal to $\vec{x}$.

(c) (2 points) Explain why the predicted response vector $\hat{\mathbb{Y}}$ and the residual vector e are orthogonal.

Since $\hat{Y} = X\hat{\theta} = \begin{bmatrix} 1 & \vec{x} \end{bmatrix} \begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{bmatrix} = \hat{\theta}_0 \cdot \vec{1} + \hat{\theta}_1 \vec{x}$

$\Rightarrow \hat{Y} \in$ span $(\vec{1}, \vec{x})$

Since $\vec{e}$ is orthogonal to span $(\vec{1}, \vec{x})$

$\Rightarrow \vec{e}$ is also orthogonal to $\hat{Y}$.

Properties of a Linear Model With No Constant Term

Suppose that we don't include an intercept term in our model. That is, our model is now simply $\hat{y} = \gamma x$, where γ is the single parameter for our model that we need to optimize. (In this equation, x is a scalar, corresponding to a single observation.)

As usual, we are looking to find the value $\hat{\gamma}$ that minimizes the average squared loss ("empirical risk") across our observed data $\{(x_i, y_i)\}, i = 1, \ldots, n$.

$$R(\gamma) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \gamma x_i)^2$$

The normal equations derived in lecture no longer hold. In this problem, we'll derive a solution to this simpler model. We'll see that the least squares estimate of the slope in this model differs from the simple linear regression model, and will also explore whether or not our properties from the previous problem still hold.

3. (4 points) Use calculus to find the minimizing $\hat{\gamma}$. That is, prove that

$$\hat{\gamma} = \frac{\sum x_i y_i}{\sum x_i^2}$$

Note: This is the slope of our regression line, analogous to $\hat{\theta}_1$ from our simple linear regression model.

$$\frac{\partial R}{\partial \gamma} = -\frac{2}{n} \sum x_i (y_i - \gamma x_i) = 0$$

$$\Rightarrow \sum x_i y_i - \gamma \sum x_i^2 = 0$$

$$\Rightarrow \hat{\gamma} = \frac{\sum x_i y_i}{\sum x_i^2}$$

4. (8 points) For our new simplified model, our design matrix is

$$\mathbb{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} | \\ \vec{x} \\ | \end{bmatrix}$$

And so our predicted response vector $\hat{\mathbb{Y}}$ can be expressed as $\hat{\mathbb{Y}} = \hat{\gamma}\vec{x}$. ($\vec{x}$ here is defined the same way it was in Question 2.)

Earlier in this homework, we established several properties that held true for the simple linear regression model that contained an intercept term. For each of the following four properties, state whether or not they still hold true even when there isn't an intercept term. Be sure to justify your answer.

(a) (2 points) $\sum_{i=1}^{n} e_i = 0$.

False. We require $\vec{1}$ to be a column of the feature matrix. Since it has no intercept term, the feature matrix may not have $\vec{1}$ as a column, thus $\vec{1} \cdot \vec{e}$ is not necessarily 0.

(b) (2 points) The column vector $\vec{x}$ and the residual vector e are orthogonal.

Since $\vec{e}$ is orthogonal to span$(\vec{x})$ and $\vec{x} \in$ span$(\vec{x})$.

$\vec{e}$ is orthogonal to $\vec{x}$.

(c) (2 points) The predicted response vector $\hat{\mathbb{Y}}$ and the residual vector e are orthogonal.

$\hat{Y} = \hat{\gamma}\vec{x} \Rightarrow \hat{Y} \in$ span$(\vec{x})$

since $\vec{e}$ is orthogonal to span$(\vec{x})$

$\hat{Y}$ and e are also orthogonal.

(d) (2 points) $(\bar{x}, \bar{y})$ is on the regression line.

False. Since we don't have an intercept term, $\hat{\theta}_0$ does not exist anymore. But our proof involve $\hat{\theta}_0$, thus now we cannot conclude that $(\bar{x}, \bar{y})$ is still on the line.