Yiyang Xu

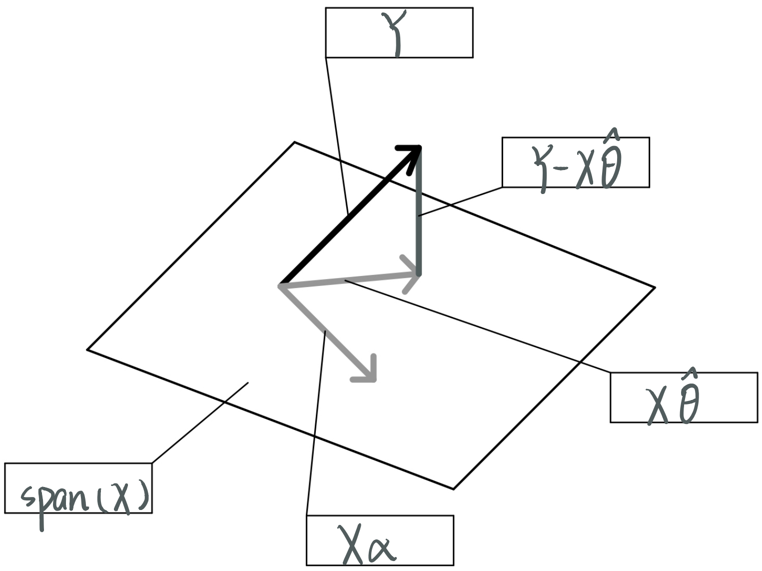
Assignment 3

ECE 4710J

Due 11/17/2021

Geometry of Least Squares

1. Suppose we have a dataset represented with the design matrix span($\mathbb{X}$) and response vector $\mathbb{Y}$. We use linear regression to solve for this and obtain optimal weights as $\hat{\theta}$. Draw the geometric interpretation of the column space of the design matrix span($\mathbb{X}$), the response vector $\mathbb{Y}$, the residuals $\mathbb{Y}-\mathbb{X}\hat{\theta}$, and the predictions $\mathbb{X}\hat{\theta}$ (using optimal parameters) and $\mathbb{X}\alpha$ (using an arbitrary vector α).



   (a) What is always true about the residuals in least squares regression? Select all that apply.

   - ☑ A. They are orthogonal to the column space of the design matrix.
   - ☑ B. They represent the errors of the predictions.
   - ☐ C. Their sum is equal to the mean squared error.
   - ☐ D. Their sum is equal to zero.
   - ☐ E. None of the above.

(b) Which are true about the predictions made by OLS? Select all that apply.

- ☑ A. They are projections of the observations onto the column space of the design matrix.
- ☑ B. They are linear combinations of the features.
- ☑ C. They are orthogonal to the residuals.
- ☐ D. They are orthogonal to the column space of the features.
- ☐ E. None of the above.

(c) We fit a simple linear regression to our data $(x_i, y_i), i = 1, 2, 3$, where x_i is the independent variable and y_i is the dependent variable. Our regression line is of the form $\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x$. Suppose we plot the relationship between the residuals of the model and the $\hat{y}s$, and find that there is a curve. What does this tell us about our model?

- ☐ A. The relationship between our dependent and independent variables is well represented by a line.
- ☐ B. The accuracy of the regression line varies with the size of the dependent variable.
- ☑ C. The variables need to be transformed, or additional independent variables are needed.

Understanding Dimensions

2. In this exercise, we will examine many of the terms that we have been working with in regression (e.g. $\hat{\theta}$) and connect them to their dimensions and to concepts that they represent. First, we define some notation. The $n \times p$ design matrix $\mathbb{X}$ corresponds to n observations on p features. (In lecture, we stated that we sometimes say $\mathbb{X}$ has $p+1$ features, where the addition feature is a column of all 1s for the intercept term, but strictly speaking that column doesn't need to exist. In this problem, one of the p columns *may* be a column of all 1s.) $\mathbb{Y}$ is the response variable. It is a vector, containing the true response for all observations. We assume in this problem that we use $\mathbb{X}$ and $\mathbb{Y}$ to compute optimal parameters $\hat{\theta}$ for a linear model, and that this linear model generates predictions using $\hat{\mathbb{Y}} = \mathbb{X}\hat{\theta}$ as we saw in lecture and in Question 1 of this discussion. Each of the n rows in our design matrix $\mathbb{X}$ contains all features for a single observation. Each of the p columns in our design matrix $\mathbb{X}$ contains a single feature, for all observations. We denote the rows and columns of $\mathbb{X}$ as follows:

$$\begin{array}{ll} \mathbb{X}_{:,j} & j^{th} \text{ column vector in } \mathbb{X}, j = 1, \ldots, p \\ \mathbb{X}_{i,:} & i^{th} \text{ row vector in } \mathbb{X}, i = 1, \ldots, n \end{array}$$

Below, on the left, we have several expressions, labelled a through h, and on the right we have several terms, labelled 1 through 10. **For each expression, determine its shape (e.g., $n \times p$), and match it to one of the given terms.** Terms may be used more than once or not at all. If a specific expression is nonsensical because the dimensions don't line up for a matrix multiplication, write "N/A" for both.

(a) $\mathbb{X}$
(b) $\hat{\theta}$
(c) $\mathbb{X}_{:,j}$
(d) $\mathbb{X}_{1,:} \cdot \hat{\theta}$
(e) $\mathbb{X}_{:,1} \cdot \hat{\theta}$
(f) $\mathbb{X}\hat{\theta}$
(g) $(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{Y}$
(h) $(I - \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T)\mathbb{Y}$

1. the residuals
2. 0
3. 1st response, y_1
4. 1st predicted value, $\hat{y_1}$
5. 1st residual, e_1
6. the estimated coefficients
7. the predicted values
8. the features for a single observation
9. the value of a specific feature for all observations
10. the design matrix

As an example, for 2a, you would write: "2a. **Dimension:** $n \times p$, **Term:** 10".

2a	Dimension: $n \times p$	Term: 10
2b	$p \times 1$	6
2c	$n \times 1$	9
2d	1×1	4
2e	N/A	N/A
2f	$n \times 1$	7
2g	$p \times 1$	6
2h	$n \times 1$	1