DS 100/200: Principles and Techniques of Data Science	**Date: October 23, 2019**

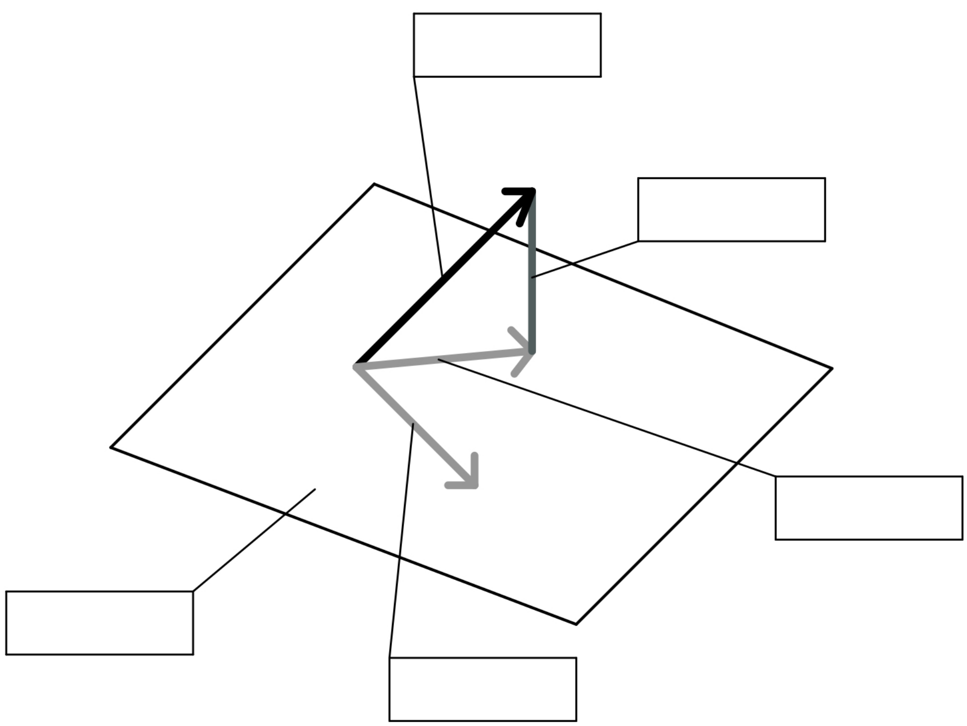
Discussion #9

Name:

Geometry of Least Squares

1. Consider the following diagram for the geometry of least squares. Fill in the blanks on the diagram with one of the following: (Note that $\hat{\beta}$ is the optimal β, and α is an arbitrary vector.)

   - span$\{\mathbb{X}\}$
   - $\vec{y}$
   - $\mathbb{X}\vec{\alpha}$
   - $\mathbb{X}\hat{\beta}$
   - $\vec{y} - \mathbb{X}\hat{\beta}$



2. Use the figure above, to explain why, for all $\alpha \in \mathbb{R}^p$,

$$\|\vec{y} - \mathbb{X}\alpha\|^2 \geq \|\vec{y} - \mathbb{X}\hat{\beta}\|^2$$

3. From the figure above, what can we say about the residuals and the column space of X? Explain your statement using linear algebra ideas.

4. Derive the normal equations from the fact above. That is, starting from the orthogonality of the residuals and column space of $\mathbb{X}$, derive $\mathbb{X}^t\vec{y} = \mathbb{X}^t\mathbb{X}\vec{\hat{\beta}}$.

5. What must be be true about $\mathbb{X}$ for the normal equation to be solvable, i.e., to get a solution for $\vec{\hat{\beta}}$? What does this imply about the rank of $\mathbb{X}$ and the features that it represents?

Dummy Variables/One-hot Encoding

In order to include a qualitative variable in a model, we convert it into a collection of dummy variables. These dummy variables take on only the values 0 and 1. For example, suppose we have a qualitative variable with 3 levels, call them A, B, and C, respectively. For concreteness, we use a specific example with 10 observations:

$$[A, A, A, A, B, B, B, C, C, C]$$

In linear modeling, we represent this variable with 3 dummy variables, $\vec{x}_A$, $\vec{x}_B$, and $\vec{x}_C$ arranged left to right in the following design matrix. This representation is also called one-hot encoding.

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

We will show that the fitted coefficients for $\vec{x}_A$, $\vec{x}_B$, and $\vec{x}_C$ are $\bar{y}_A$, $\bar{y}_B$, and $\bar{y}_C$, the average of the y_i values for each of the groups, respectively.

6. Show that the columns of $\mathbb{X}$ are orthogonal, (i.e., the dot product between any pair of column vectors is 0).

7. Show that
$$\mathbb{X}^t\mathbb{X} = \begin{bmatrix} n_A & 0 & 0 \\ 0 & n_B & 0 \\ 0 & 0 & n_C \end{bmatrix}$$
Here, n_A, n_B, n_C are the number of observations in each of the three groups defined by the levels of the qualitative variable.

8. Show that
$$\mathbb{X}^t\vec{y} = \begin{bmatrix} \sum_{i\in A} y_i \\ \sum_{i\in B} y_i \\ \sum_{i\in C} y_i \end{bmatrix}$$

9. Use the results from the previous questions to solve the normal equations for $\hat{\beta}$, i.e.,
$$\begin{aligned} \hat{\beta} &= [\mathbb{X}^t\mathbb{X}]^{-1}\mathbb{X}^t\vec{y} \\ &= \begin{bmatrix} \bar{y}_A \\ \bar{y}_B \\ \bar{y}_C \end{bmatrix} \end{aligned}$$