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

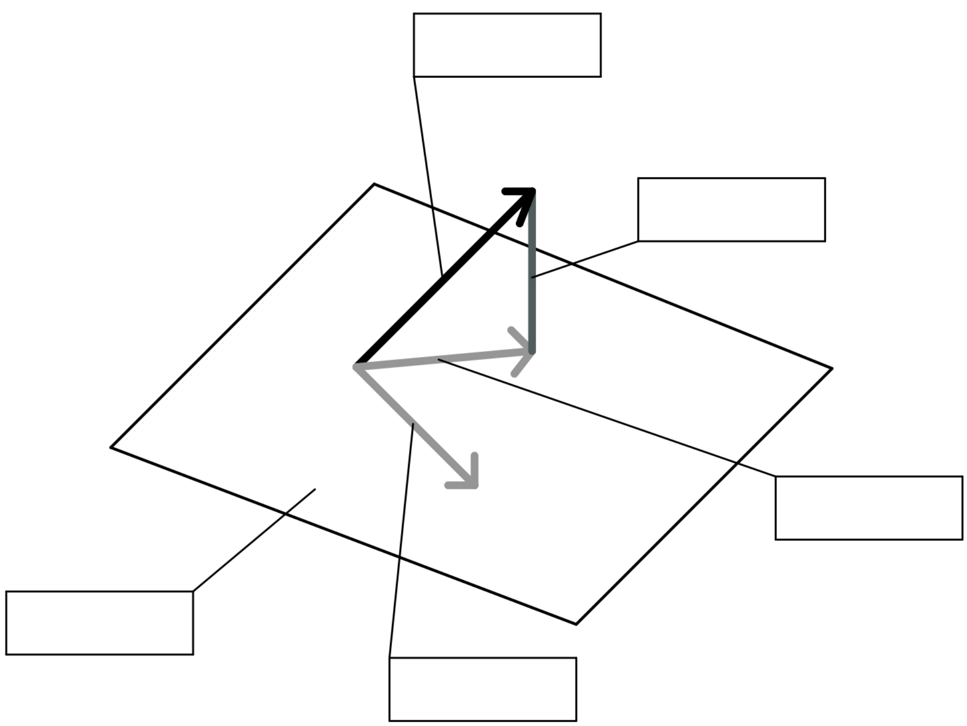
Discussion #9 Solutions

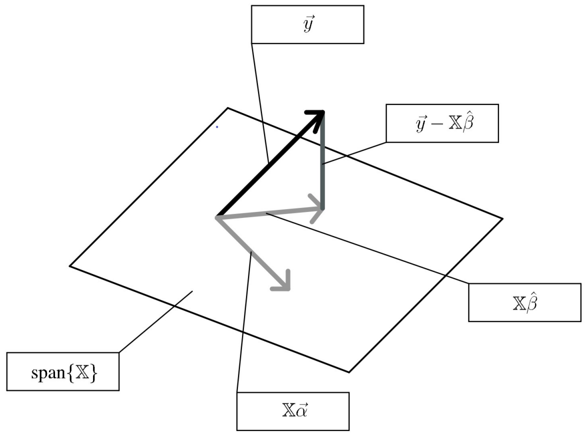
Name:

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.)

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





Solution:

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

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

Solution: Since $\vec{\hat{y}}$ is the projection of $\vec{y}$ onto the column space of $\mathbb{X}$, by definition of projection, it is closest to $\vec{y}$, i.e.,

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

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.

Solution: We can say that the residuals are orthogonal to the column space of $\mathbb{X}$. The residuals are:

$$\begin{aligned}\vec{e} &:= \vec{y} - \vec{\hat{y}} \\ &:= \vec{y} - \mathbb{X}\vec{\hat{\beta}}\end{aligned}$$

The projection of $\vec{y}$ onto the column space of $\mathbb{X}$ implies that $\vec{e} \cdot \vec{\hat{y}} = 0$. The notion of Pythagorean's theorem is

$$\|\vec{y}\|^2 = \|\vec{\hat{y}}\|^2 + \|\vec{e}\|^2.$$

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 $\mathbb{X}^t\vec{y} = \mathbb{X}^t\mathbb{X}\vec{\hat{\beta}}$.

> **Solution:** From above, every vector $\vec{x_i}, i = 1, 2, \ldots, p$ is orthogonal to the residuals, i.e. their dot product is 0.
>
> Mathematically,
>
> $$\begin{aligned} \mathbb{X}^t\vec{e} &= 0, \text{ or equivalently,} \\ \mathbb{X}^t(\vec{y} - \mathbb{X}\hat{\beta}) &= 0. \end{aligned}$$

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?

> **Solution:** The design matrix must be invertible; hence, no linearly dependent features.
>
> In order for the columns to be linearly independent, the rank of $\mathbb{X}$ must match the number of columns in $\mathbb{X}$.
>
> In addition, it must be the case that the number of columns of the design matrix is less than or equal to the number of rows.

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).

Solution: The argument is the same for any pair of $\vec{x}$s so we show the orthogonality for one pair, $\vec{x}_A\dot{\vec{x}}_B$.

$$\begin{aligned} \vec{x}_A\dot{\vec{x}}_B &= \sum_{i=1}^{1} 0 x_{A,i} x_{B,i} \\ &= \sum_{i=1}^{4}(1 \times 0) + \sum_{i=5}^{7}(0 \times 1) + \sum_{i=8}^{10}(0 \times 0) \\ &= 0 \end{aligned}$$

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.

> **Solution:** Here, we note that
> $$\mathbb{X}^t = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$
> We also note that
> $$\mathbb{X}^t\mathbb{X} = \begin{bmatrix} \vec{x}_A^t\vec{x}_A & \vec{x}_A^t\vec{x}_B & \vec{x}_A^t\vec{x}_C \\ \vec{x}_B^t\vec{x}_A & \vec{x}_B^t\vec{x}_B & \vec{x}_B^t\vec{x}_C \\ \vec{x}_C^t\vec{x}_A & \vec{x}_C^t\vec{x}_B & \vec{x}_C^t\vec{x}_C \end{bmatrix}$$
> Since we earlier established the orthogonality of the vectors in $\mathbb{X}$, we find $\mathbb{X}^t\mathbb{X}$ to be the diagonal matrix:
> $$\mathbb{X}^t\mathbb{X} = \begin{bmatrix} n_A & 0 & 0 \\ 0 & n_B & 0 \\ 0 & 0 & n_C \end{bmatrix}$$

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

> **Solution:** Note in the previous solution we found $\mathbb{X}^t$. The solution follows from recognizing that for a row in $\mathbb{X}^t$, e.g., the first row, we have
> $$\sum_{i=1}^{n} x_{A,i} \times y_i = \sum_{i=1}^{4} y_i = \sum_{i\in \text{ group A}} y_i$$

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

Solution: By inspection, we can find

$$[\mathbb{X}^t\mathbb{X}]^{-1} = \begin{bmatrix} \frac{1}{n_A} & 0 & 0 \\ 0 & \frac{1}{n_B} & 0 \\ 0 & 0 & \frac{1}{n_C} \end{bmatrix}$$

When we pre-multiply $\mathbb{X}^t\vec{y}$ by this matrix, we get

$$\begin{bmatrix} \frac{1}{n_A} & 0 & 0 \\ 0 & \frac{1}{n_B} & 0 \\ 0 & 0 & \frac{1}{n_C} \end{bmatrix} \begin{bmatrix} \sum_{i\in A} y_i \\ \sum_{i\in B} y_i \\ \sum_{i\in C} y_i \end{bmatrix} = \begin{bmatrix} \bar{y}_A \\ \bar{y}_B \\ \bar{y}_C \end{bmatrix}$$