The OLS Estimator (3)

STA 211: The Mathematics of Regression

Yue Jiang

January 31, 2023

Review: span and column space

A set of vectors is **linearly independent** if no linear combination (besides all zeroes) of the vectors equals the zero vector; that is, if none of the vectors can be written as a **linear combination** of the others (and non are the zero vector).

Review: span and column space

A set of vectors is **linearly independent** if no linear combination (besides all zeroes) of the vectors equals the zero vector; that is, if none of the vectors can be written as a **linear combination** of the others (and non are the zero vector).

The **span** of a set of vectors is the set of all possible linear combinations of them (you may also recall that a linearly independent set of vectors that spans a subspace forms a **basis** for that subspace, but this is less relevant for today).

The **column space** of $\mathbf{X}$ is the span of the columns of $\mathbf{X}$.

- What does the column space of $\mathbf{X}$ represent in plain English?

A miscellaneous follow-up item...

Suppose we are trying to predict the amount of sleep a Duke student gets based on whether they are in Pratt (vs. non-Pratt; these are the only two options). Consider the following model:

$$Sleep_i = \beta_0 + \beta_1 1(Pratt_i == "Yes") + \beta_2 1(Pratt_i == "No")$$

In-class assignment (I originally intended for this to be homework, but figured it'd be enlightening to go through in class!):

- Write out the design matrix for this hypothesized linear model.
- Demonstrate that the design matrix is not of full column rank (that is, affirmatively provide one of the columns in terms of the others).
- Use this intuition to explain why when we include categorical predictors, we cannot include both indicators for every level of the variable *and* an intercept.

A geometric interpretation (see board)

The span of a single vector in $\mathbb{R}^2$

A geometric interpretation (see board)

The span of two vectors in $\mathbb{R}^3$

A geometric interpretation (see board)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$$

($\mathbf{X}$ as a function being applied to $\boldsymbol{\beta}$)

A geometric interpretation (see board)

$$\mathbf{y} = \mathbf{X}\beta$$

($\mathbf{X}$ as a salad *(thanks Zi Chong Kao for the analogy)*)

A geometric interpretation (see board)

$$\mathbf{y} = \mathbf{X}\beta$$

($\mathbf{X}$ as "the space of all its possible outputs")

A geometric interpretation (see board)

Let's live in $\mathbb{R}^3$ for now (just for visualization purposes):

$$\mathbf{y} = \beta_0 \mathbf{x}_0 + \beta_1 \mathbf{x}_1, \text{ for instance with } n = 3:$$

$$\underbrace{\begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}}_{\mathbf{X}} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

- What might the column space of $\mathbf{X}$ look like? Where do the vectors $\mathbf{x}_0$ (the vector of 1s) and $\mathbf{x}_1$ fit in?
- Does $\mathbf{y}$ live in the column space of $\mathbf{X}$?

A geometric interpretation (see board)

- Uh oh, it looks like $\mathbf{y}$ doesn't live in $\mathcal{C}(\mathbf{X})$.
- Can we find another vector $\mathbf{z}$ that's in $\mathcal{C}(\mathbf{X})$, but is also "as close as possible" to $\mathbf{y}$?

A geometric interpretation (see board)

- Uh oh, it looks like $\mathbf{y}$ doesn't live in $\mathcal{C}(\mathbf{X})$.
- Can we find another vector $\mathbf{z}$ that's in $\mathcal{C}(\mathbf{X})$, but is also "as close as possible" to $\mathbf{y}$?
- Why would this vector be expressible as $\mathbf{z} = \mathbf{X}\mathbf{w}$ for some $\mathbf{w}$?
- What's the "difference" between $\mathbf{z}$ and $\mathbf{y}$ (let's call it $\mathbf{e} = \mathbf{y} - \mathbf{z}$) (and how would we make this "as close as possible," which is to say, to minimize its length)? Have we seen this thing before?

A geometric interpretation (see board)

- Uh oh, it looks like $\mathbf{y}$ doesn't live in $\mathcal{C}(\mathbf{X})$.
- Can we find another vector $\mathbf{z}$ that's in $\mathcal{C}(\mathbf{X})$, but is also "as close as possible" to $\mathbf{y}$?
- Why would this vector be expressible as $\mathbf{z} = \mathbf{X}\mathbf{w}$ for some $\mathbf{w}$?
- What's the "difference" between $\mathbf{z}$ and $\mathbf{y}$ (let's call it $\mathbf{e} = \mathbf{y} - \mathbf{z}$) (and how would we make this "as close as possible," which is to say, to minimize its length)? Have we seen this thing before?
- This is called the **projection** of $\mathbf{y}$ onto $\mathcal{C}(\mathbf{X})$. What would this vector $\mathbf{z}$ look like, geometrically?
- How would we choose a $\mathbf{z}$ that minimizes the distance between $\mathbf{y}$ and something that lives in $\mathcal{C}(\mathbf{X})$?

A geometric interpretation (see board)

Note that the vector $\mathbf{e}$ is orthogonal to the plane $\mathcal{C}(\mathbf{X})$ (that is, the plane spanned by the variables in $\mathbf{X}$). This means that for any vector in $\mathcal{C}(\mathbf{x})$, the inner product between this vector and $\mathbf{e}$ is 0.

A geometric interpretation (see board)

Note that the vector $\mathbf{e}$ is orthogonal to the plane $\mathcal{C}(\mathbf{X})$ (that is, the plane spanned by the variables in $\mathbf{X}$). This means that for any vector in $\mathcal{C}(\mathbf{x})$, the inner product between this vector and $\mathbf{e}$ is 0.

We just established $\mathbf{X}^T\mathbf{e} = \mathbf{0}$. Also notice that $\mathbf{e} = \mathbf{y} - \mathbf{z}$, and $\mathbf{z} = \mathbf{X}\mathbf{w}$ for some vector $\mathbf{w}$:

$$\mathbf{X}^T(\mathbf{y} - \mathbf{X}\mathbf{w}) = \mathbf{0}$$

- Solve this equation for $\mathbf{w}$. What is the solution? What assumption did you have to make?

Of course, we can generalize this to any n-dimensional inner product space (it's just easier to visualize things in three dimensions)

Homework 3: Due Feb. 7

1. Recall the **QR** factorization of a full column rank $n \times p$ matrix $\mathbf{X}$ into the product of an $n \times p$ matrix $\mathbf{Q}$ with orthonormal columns and an invertible upper triangular $p \times p$ matrix $\mathbf{R}$. Express the least squares solution $\widehat{\beta}$ in terms of $\mathbf{Q}$ and $\mathbf{R}$. Compare this solution to $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$. Why might someone want to use the **QR** decomposition instead (hint: can you think of a reason after going through Exercise 2)?
2. Consider the estimation problem encountered on Slide 9. Use the **QR** decomposition to solve for the least squares solution (hint: use the Gram-Schmidt process to do this).
3. Explain why the residuals $\mathbf{y} - \widehat{\mathbf{y}}$ live in the orthogonal complement of the space spanned by $\mathbf{X}$. What is the dimension of this space?