Math 342W/642/742W

Recitation – Day #8 (2.27.25)

I. OLS Basics

(i) Define $\boldsymbol{b}$ in terms of the solution to the *normal equations* to the *least squares* problem?

$$\boldsymbol{b} = \left(X^T X\right)^{-1} X\boldsymbol{y}$$

(ii) What does finding $\boldsymbol{b}$ accomplish in the context of the overall machine learning problem of *regression*?

Finding $\boldsymbol{b}$, the vector of weights that minimizes SSE and returns the optimal g in the hypothesis set $\mathcal{H} = \{\boldsymbol{w}^T\boldsymbol{x} \mid \boldsymbol{w} \in \mathbb{R}^{p+1}\}$, accomplishes the learning/estimation of the parameters of the linear regression model that can be used for prediction.

(iii) Define the "*hat*" matrix, denoted by H.

The $n \times n$ hat matrix H is $H = X(X^TX)^{-1}X^T$ since $\hat{\boldsymbol{y}} = X\boldsymbol{b} = X\left(X^TX\right)^{-1}X^T\boldsymbol{y} = H\boldsymbol{y}$

(iv) What are the two properties H? What do we call such matrices?

H is an **orthogonal projection matrix** which has two properties:

- Symmetric: $H = H^T$
- Idempotent: $H^2 = H$

(v) What does H do to the given vector of responses/labels $\boldsymbol{y}$?

H projects the given vector of responses/labels $\boldsymbol{y}$ onto the $(p+1)-$dimensional subspace of the column space of X producing the vector $\hat{\boldsymbol{y}}$.

II. Preliminaries for Geometric Interpretation of OLS

We have $X \in \mathbb{R}^{n\times(p+1)}$.

(i) Define colsp$[X]$.

colsp$[X]$ = the set of all linear combinations of the columns of X

(ii) What is the rank of X? How so? What does this mean for the rank of H?

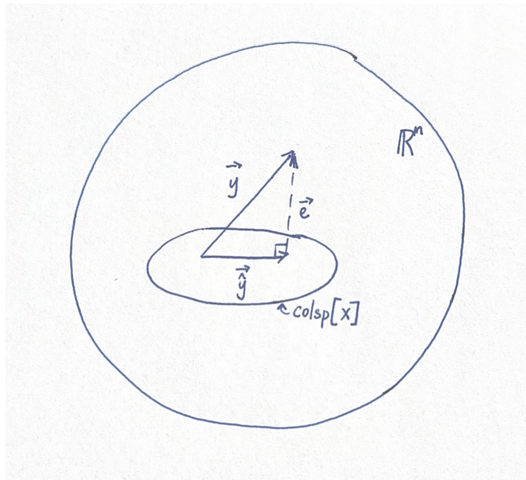
The rank of X is $p+1$ because X by construction has $p+1$ linearly independent column vectors. By the definition of the matrix H, it follows that the rank of H also equals $p+1$.

(iii) Unpack mathematically what $\hat{\boldsymbol{y}} \in \text{colsp}[X]$ means.

$$\hat{\boldsymbol{y}} \in \text{colsp}[X] \iff \exists\, \boldsymbol{b} \text{ such that } X\boldsymbol{b} = \hat{\boldsymbol{y}}$$

III. Geometric Interpretation of OLS

(i) Give an illustration of OLS with $\boldsymbol{y}, \hat{\boldsymbol{y}}, \boldsymbol{e}, \mathbb{R}^n$ and colsp$[X]$.



(ii) In what space does the *residual vector*, $\boldsymbol{e}$ reside in? What is its dimension?

The residual vector $\boldsymbol{e}$ resides in the residual space also known as colsp$[X_\perp]$ and the dimension of the residual space equals $n-(p+1)$.

(iii) What is the matrix that projects $\boldsymbol{y}$ onto the *residual space*? What is its rank?

The matrix $(I_n - H)$ projects $\boldsymbol{y}$ onto the residual space and has rank$[I_n - H] = n-(p+1)$.

(iv) In what way can the "*full space*", $\mathbb{R}^n$, be decomposed in?

$$\mathbb{R}^n = \text{colsp}[X] \oplus \text{colsp}[X_\perp]$$

IV. QR-Decomposition

(i) What is the goal behind the QR-decomposition/factorization of a matrix? What is accomplished?

The $QR-$decomposition of an $n \times k$ matrix A of full rank produces an $n \times k$ orthogonal matrix Q and a $k \times k$ upper triangular matrix R such that $A = QR$.

(ii) What is the process that creates the QR-decomposition of a matrix called?

The Gram-Schmidt process is the process by which any given basis of a vector space is made into an orthogonal and even orthonormal basis for the same given vector space.

(iii) How will we now define the orthogonal projection matrix H?

$$H = QQ^T$$