5.2 Geometry of linear regression

In Ordinary Least Squares (OLS), we choose the vector $\mathbf{x}_{OLS}$ satisfying

$$\mathbf{x}_{OLS} = \arg\min_{\mathbf{x}\in\mathbb{R}^k} ||Y - B\mathbf{x}||^2.$$

It is quite easy to figure out what $\mathbf{x}_{OLS}$ should be, see the module on the Geometry of Linear Regression (from EE345). This document summarizes the background for your convenience. This document does not however attempt a full explanation of all relevant prior concepts from EE345. There are a number of places where you are asked (why? or definition?). Make sure you know the relevant concepts marked.

The solution will be

$$\mathbf{x}_{OLS} = (B^T B)^{-1} B^T Y,$$

as long as the null space (definition?) of B is trivial (meaning that the only solution of $B\mathbf{z} = \mathbf{0}$ is $\mathbf{z} = \mathbf{0}$, where $\mathbf{z}$ is the vector of variables we are solving for)—or equivalently, when the columns of B are linearly independent (why?)

Problem Show that for any matrix B, $null(B) = null(B^T B)$. Therefore prove that if B has a trivial null space, $B^T B$ will be invertible

And in most practical situations where B is a tall matrix, the matrix B will have a trivial null space, unless some of the features are introduced as a linear combination of others. But such linear combinations are redundant from a linear-model point of view and can be discarded from computation.

We briefly summarize the basic insight behind the geometry of linear regression, building on three elementary observations.

- The first elementary observation is that for any vector $\mathbf{x}$, $B\mathbf{x}$ is a linear combination of the columns of $\mathbf{x}$ (why?). Recall that the column space of B is a linear space (definition?) and is the set of all possible linear combinations $B\mathbf{x}$ that can be obtained when we allow $\mathbf{x}$ to range through every possible $k-$coordinate vector, *i.e.,* $\mathbf{x} \in \mathrm{R}^k$, namely

  $$col(B) = \left\{ B\mathbf{x} : \mathbf{x} \in \mathbb{R}^k \right\}.$$

  No matter how we choose $\mathbf{x}_{OLS}$, $B\mathbf{x}_{OLS}$ is a point in the linear space $col(B)$.

- The second elementary observation is that $Y - B\mathbf{x}$ is a vector connecting the tip of Y to a point in the linear space $col(B)$.

- The third observation, combining the observations above, is that finding $\mathbf{x}_{OLS}$ is equivalent to finding the point $\mathbf{w}_{OLS}$ in $col(B)$ closest to Y, namely,

$$\mathbf{w}_{OLS} = \min_{\mathbf{w} \in col(B)} ||Y - \mathbf{w}||^2,$$

  and we will have $\mathbf{w}_{OLS} = B\mathbf{x}_{OLS}$. The point $\mathbf{w}_{OLS}$ minimizing the equation above is simply the projection of Y onto the $col(B)$, namely that $Y - \mathbf{w}_{OLS} = Y - B\mathbf{x}_{OLS}$ is orthogonal to *every* vector in $col(B)$, or that $Y - B\mathbf{x}_{OLS} \in null(B^T)$ (why?)

From the last observation above, we have

$$B^T(Y - B\mathbf{x}_{OLS}) = \mathbf{0},$$

which after rearranging, and noting that $B^T B$ is invertible if B has a trivial null space, yields the expression for $\mathbf{x}_{OLS}$.