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). The solution is

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

and as described in the module above, so long as the null space of B is trivial (only solution of $B\mathbf{z} = \mathbf{0}$, where $\mathbf{z}$ is the vector of variables we are solving for, is $\mathbf{z} = \mathbf{0}$).

Problem Show that for any matrix B, $null(B) = null(B^TB)$.

Now the matrix B will have a trivial null space in most practical situations where B is a tall matrix, unless some of the features are a linear combination of others (in which case, such linear combinations are redundant from a linear-model point of view and can be discarded). When we have no redundant features in B, B^TB will be invertible, as you saw in the module from EE345 above.

We briefly summarize the basic insight behind the geometry of linear regression from the module above, 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 $\mathbf{x}$. Recall that the column space of B is the set of all $B\mathbf{x}$ (a $n-$coordinate vector) 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

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

  and that $col(B)$ is a linear space (*i.e.,* closed under linear combinations). Therefore, 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 $Y - B\mathbf{x}_{OLS} \in (B^T)$ (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 (why?), yields the expression for $\mathbf{x}_{OLS}$.