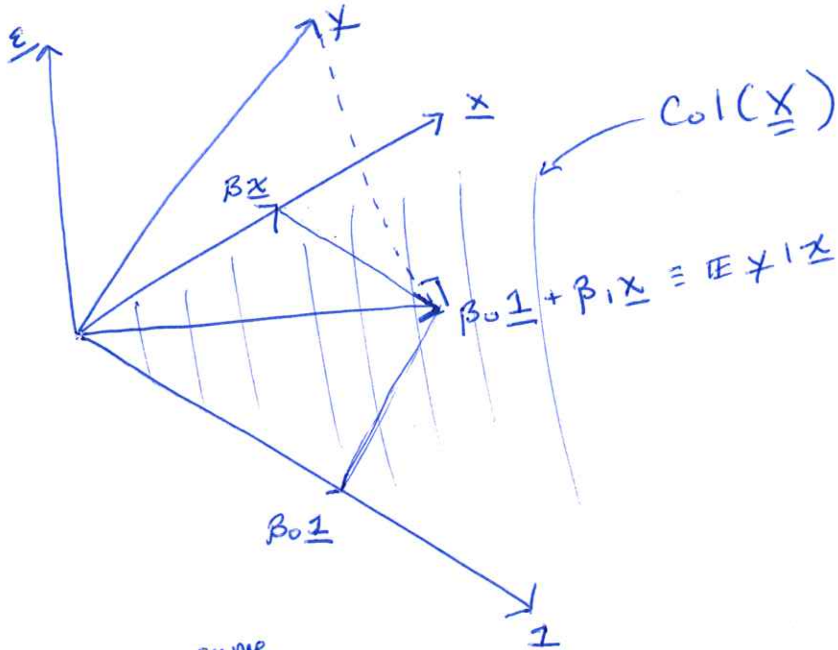
Simple linear regression <u>Geometry</u> intuition:

Picture $\mathbb{R}^3$:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \beta_0 + \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \beta_1 + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

$$\underline{y} = \underline{1}\, \beta_0 + \underline{x}\, \beta_1 + \underline{\varepsilon}$$



The dashed line is the same size as the error vector.

View 1: To minimize the size of $\underline{\varepsilon}$ is to min $\sqrt{\Sigma \varepsilon_i^2}$

i.e. min $\sqrt{\varepsilon^T \varepsilon}$

i.e. min $\varepsilon^T \varepsilon$ since $\sqrt{\ }$ is monotonic.

View 2:

we want $\hat{\underline{\varepsilon}}$ to be orthogonal to $col(\underline{\underline{X}})$

i.e. we want $\underline{\underline{X}}^T \hat{\underline{\varepsilon}} = \underline{0}$

i.e. $\underline{\underline{X}}^T (y - \hat{y}) = \underline{0}$

$\Rightarrow \underline{\underline{X}}^T (y - \underline{\underline{X}} \hat{\underline{\beta}}) = \underline{0}$

$\Rightarrow \underline{\underline{X}}^T \underline{y} = \underline{\underline{X}}^T \underline{\underline{X}} \hat{\underline{\beta}}$

$\Rightarrow \boxed{\hat{\underline{\beta}} = (\underline{\underline{X}}^T \underline{\underline{X}})^{-1} \underline{\underline{X}}^T \underline{y}}$