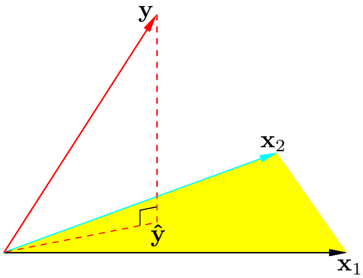
The Geometric Interpretation of LS

Recall that the LS optimization

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2,$$

which is equivalent to finding a vector $\mathbf{v}$ from the subspace $C(\mathbf{X})$ that minimizes $\|\mathbf{y} - \mathbf{v}\|^2$.



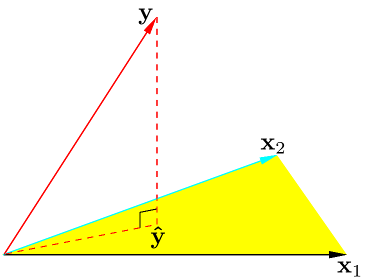
The Geometric Interpretation of LS

Recall that the LS optimization

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2,$$

which is equivalent to finding a vector $\mathbf{v}$ from the subspace $C(\mathbf{X})$ that minimizes $\|\mathbf{y} - \mathbf{v}\|^2$.

Intuitively we know what the optimal $\mathbf{v}$ is: it's the projection of $\mathbf{y}$ onto the space $C(\mathbf{X})$.



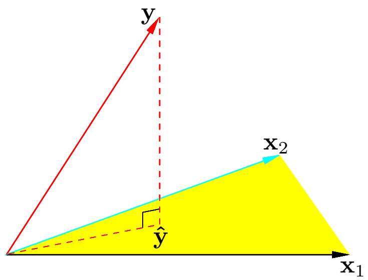
The Geometric Interpretation of LS

Recall that the LS optimization

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2,$$

which is equivalent to finding a vector $\mathbf{v}$ from the subspace $C(\mathbf{X})$ that minimizes $\|\mathbf{y} - \mathbf{v}\|^2$.

Intuitively we know what the optimal $\mathbf{v}$ is: it's the projection of $\mathbf{y}$ onto the space $C(\mathbf{X})$.



The essence of LS: decompose the data vector $\mathbf{y}$ into two orthogonal components,

$$\mathbf{y}_{n\times 1} = \hat{\mathbf{y}}_{n\times 1} + \mathbf{r}_{n\times 1}.$$