THE GEOMETRY OF LEAST SQUARES

Suppose

- $\mathbf{y} \in \mathbb{R}^3$
- $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2$ linearly independent
- Linear subspace $\mathcal{X} \subset \mathbb{R}^3$ defined by

$$\mathcal{X} \equiv \{\mathbf{x} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2, \beta_1, \beta_2 \in \mathbb{R}\}.$$

- $\mathbf{X}$ a 3×2 matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ 0 & 0 \end{bmatrix}$$

- $\widehat{\mathbf{y}} \in \mathcal{X}$ is the projection of $\mathbf{y}$ onto $\mathcal{X}$, and satisfies

$$\widehat{\mathbf{y}} = \arg \min_{\mathbf{y}^* \in \mathcal{X}} \|\mathbf{y} - \mathbf{y}^*\|^2.$$

  We have that $\mathbf{y}^* = \mathbf{X}\beta$ for some β and hence $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\beta}$ where

$$\mathbf{X}^\top (\mathbf{y} - \mathbf{X}\widehat{\beta}) = \mathbf{0}.$$

- It follows that

$$\widehat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \qquad \widehat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \mathbf{H}\mathbf{y}.$$

- Can repeat for linear subspaces $\mathcal{X}_1, \mathcal{X}_2 \subset \mathbb{R}^3$

$$\mathcal{X}_1 \equiv \{\mathbf{x} : \mathbf{x} = \beta_1 \mathbf{x}_1, \beta_1 \in \mathbb{R}\} \qquad \mathcal{X}_2 \equiv \{\mathbf{x} : \mathbf{x} = \beta_2 \mathbf{x}_2, \beta_2 \in \mathbb{R}\}$$

  with matrices

$$\mathbf{X}_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ 0 \end{bmatrix} \qquad \mathbf{X}_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ 0 \end{bmatrix}$$

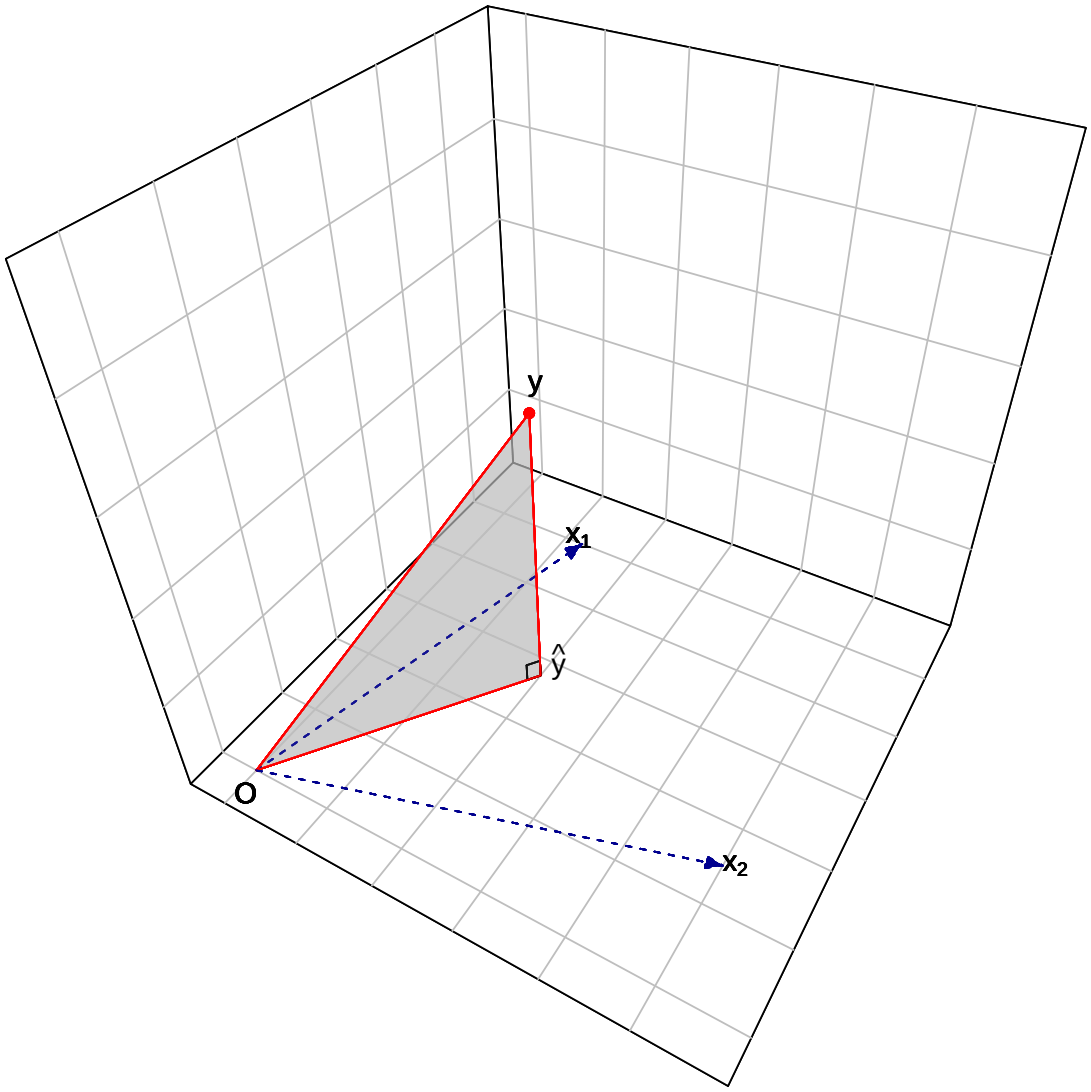
  and estimates and fitted values

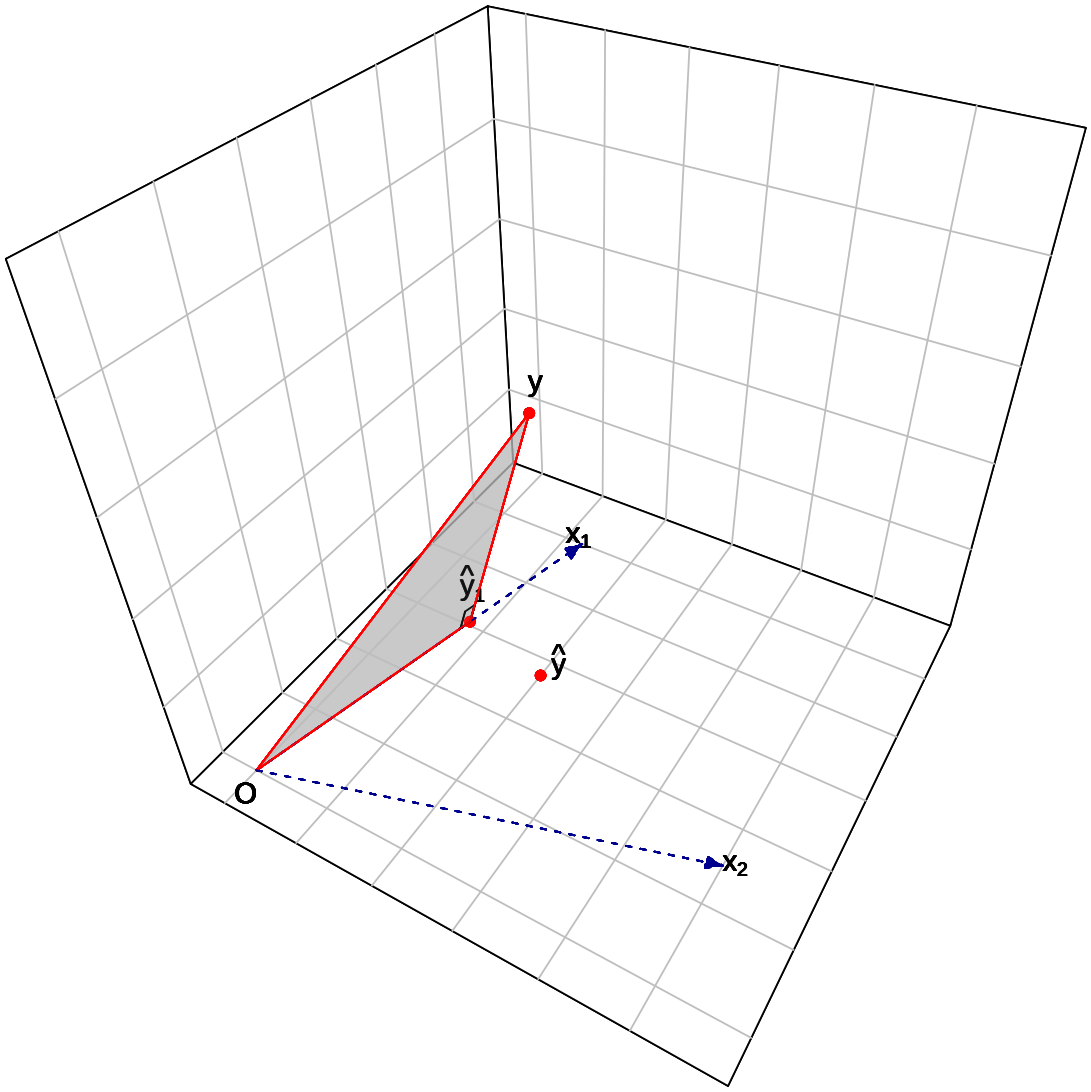
$$\widehat{\beta}_1 = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{y} \qquad \widehat{\mathbf{y}}_1 = \mathbf{X}_1(\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{y} = \mathbf{H}_1 \mathbf{y}.$$

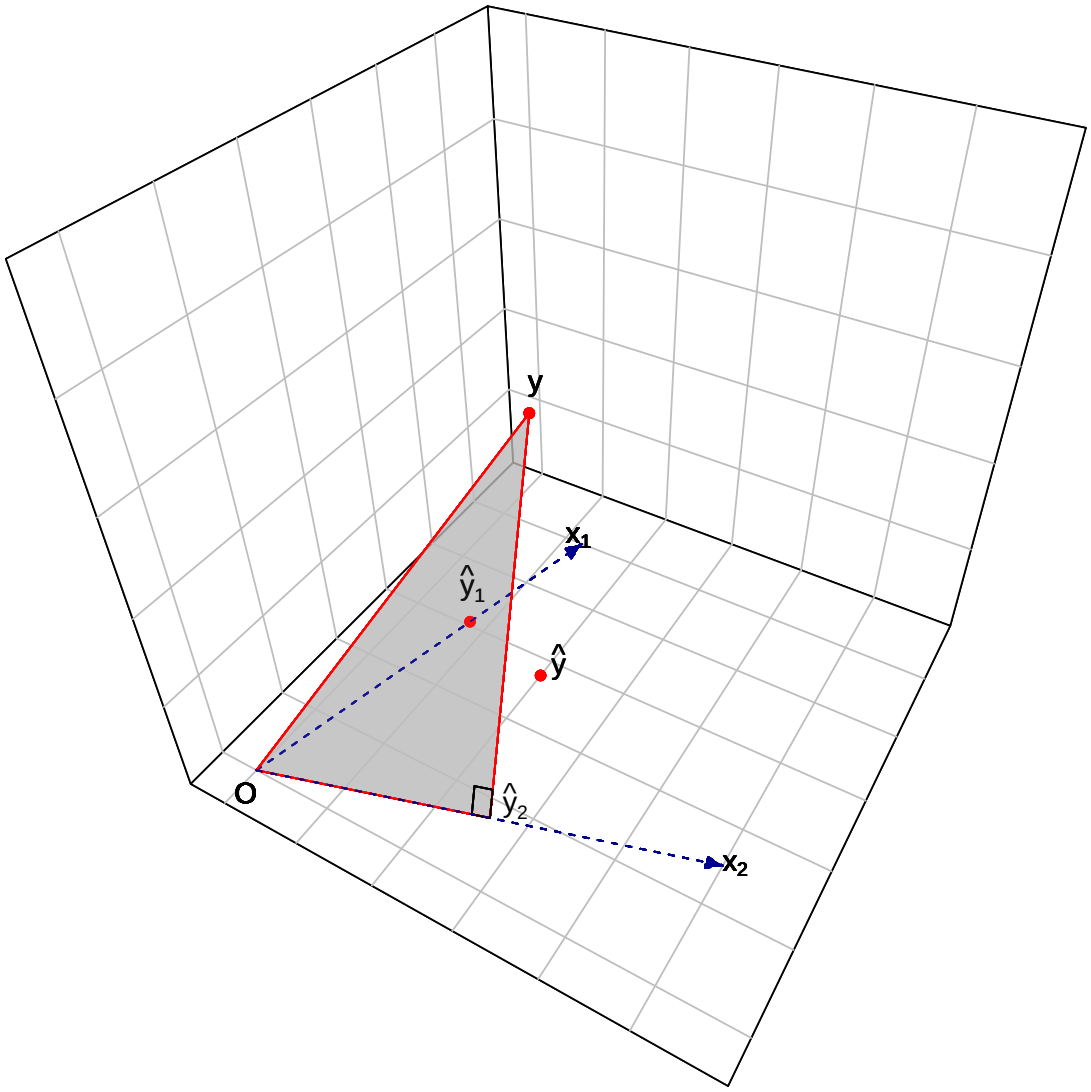
  and

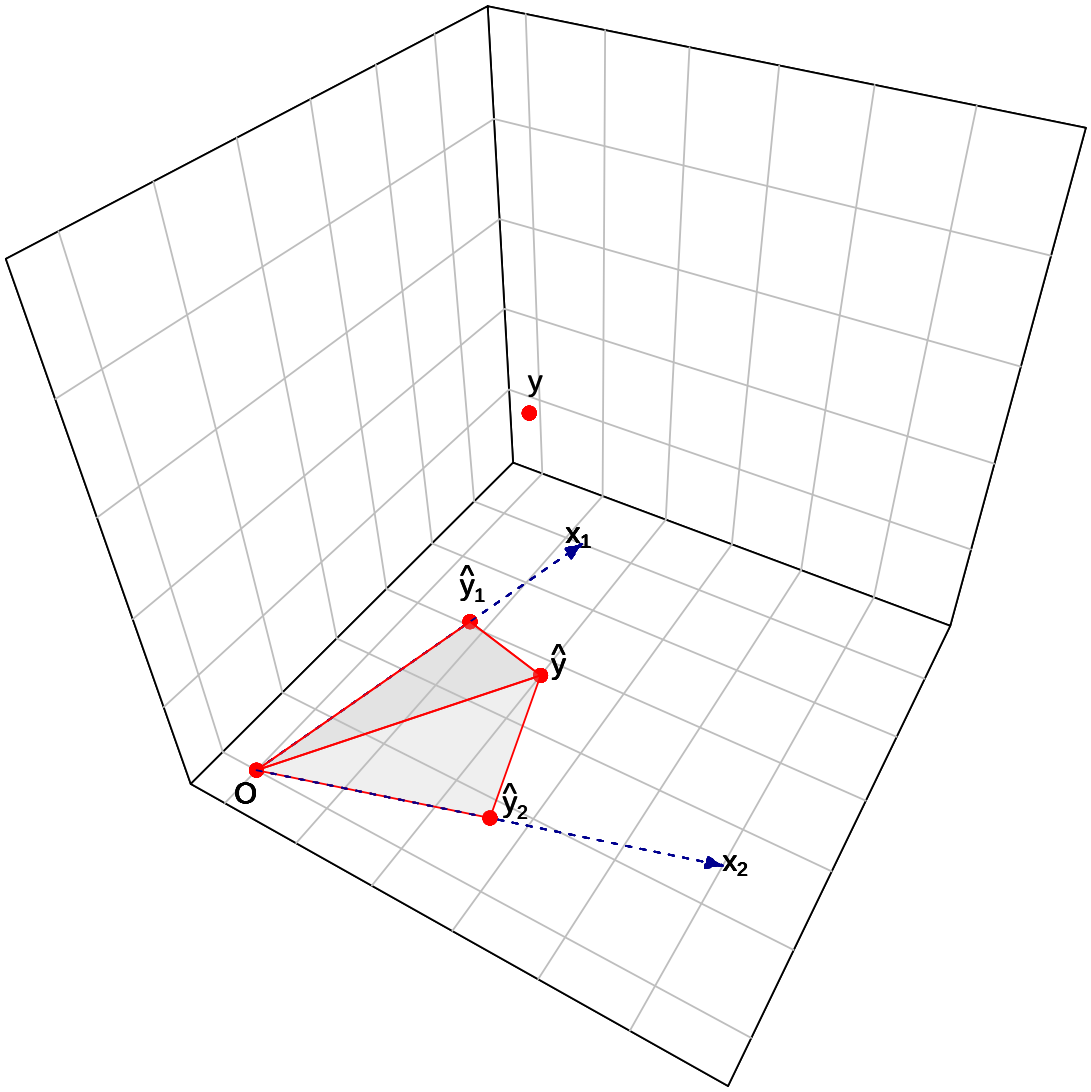
$$\widehat{\beta}_2 = (\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{y} \qquad \widehat{\mathbf{y}}_2 = \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{y} = \mathbf{H}_2 \mathbf{y}.$$

  respectively.


y
x1
ŷ
O
x2


y
x1
ŷ1
ŷ
O
x2


y
x1
ŷ1
ŷ
O
ŷ2
x2




```
t(Yhat) %*% (Y-Yhat)

+      [,1]
+ [1,]    0

t(Y1hat) %*% (Y-Y1hat)

+                [,1]
+ [1,] -6.661338e-16

t(Y2hat) %*% (Y-Y2hat)

+               [,1]
+ [1,] 2.220446e-16

t(Y1hat) %*% (Y-Y1hat)
```

```
+                [,1]
+ [1,] -6.661338e-16
```

```
t(Y2hat) %*% (Y-Y2hat)
```

```
+                [,1]
+ [1,] 2.220446e-16
```