# Projection

For any vector $\mathbf{y} \in \mathbb{R}^n$ and a subspace $\mathcal{M} \subseteq \mathbb{R}^n$, there exists a unique vector $\hat{\mathbf{y}}$ such that

1. $\hat{\mathbf{y}} \in \mathcal{M}$, and
2. $(\mathbf{y} - \hat{\mathbf{y}}) \perp \mathcal{M}$.

We call $\hat{\mathbf{y}}$ the **projection** of $\mathbf{y}$ onto $\mathcal{M}$.

$$\mathbf{y} = \underbrace{\hat{\mathbf{y}}}_{\in \mathcal{M}} + \big(\underbrace{\mathbf{y} - \hat{\mathbf{y}}}_{\in \mathcal{M}^\perp}\big)$$

# How to Compute $\hat{\mathbf{y}}$?

Approach I: Suppose $\mathcal{M}$ has an ONB $\{\mathbf{e}_1, \ldots, \mathbf{e}_p\}$. Then any vector $\mathbf{v} \in \mathcal{M}$ can be uniquely expressed as

$$\mathbf{v} = a_1\mathbf{e}_1 + \cdots a_p\mathbf{e}_p, \quad \text{with } a_j = \mathbf{v}^t\mathbf{e}_j.$$

Since $\hat{\mathbf{y}} \in \mathcal{M}$, we have

$$\hat{\mathbf{y}} = a_1\mathbf{e}_1 + \cdots a_p\mathbf{e}_p, \quad \text{with } a_j = \hat{\mathbf{y}}^t\mathbf{e}_j.$$

Note that we can replace $\hat{\mathbf{y}}$ by $\mathbf{y}$ when computing $a_j$:

$$\mathbf{y}^t\mathbf{e}_j = (\hat{\mathbf{y}} + \mathbf{y} - \hat{\mathbf{y}})^t\mathbf{e}_j = \hat{\mathbf{y}}^t\mathbf{e}_j + \underbrace{(\mathbf{y} - \hat{\mathbf{y}})^t\mathbf{e}_j}_{=0} = \hat{\mathbf{y}}^t\mathbf{e}_j.$$

So we have

$$
\begin{aligned}
\hat{\mathbf{y}} &= \big(\mathbf{e}_1^t\mathbf{y}\big)\mathbf{e}_1 + \cdots + \big(\mathbf{e}_p^t\mathbf{y}\big)\mathbf{e}_p \\
&= \mathbf{e}_1\big(\mathbf{e}_1^t\mathbf{y}\big) + \cdots + \mathbf{e}_p\big(\mathbf{e}_p^t\mathbf{y}\big) \\
&= \mathbf{e}_1\mathbf{e}_1^t\mathbf{y} + \cdots + \mathbf{e}_p\mathbf{e}_p^t\mathbf{y} \\
&= \big(\mathbf{e}_1\mathbf{e}_1^t + \cdots + \mathbf{e}_p\mathbf{e}_p^t\big)\mathbf{y} \\
&= \mathbf{M}_{n\times 1}\mathbf{y}.
\end{aligned}
$$

Approach II: Suppose $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ is a basis for $\mathcal{M}$. Then any vector $\mathbf{v} \in \mathcal{M}$ can be expressed as a unique linear combination of $\mathbf{x}_j$'s. In particular, assume

$$\hat{\mathbf{y}} = b_1\mathbf{x}_1 + \cdots + b_p\mathbf{x}_p = \mathbf{X}_{n\times p}\mathbf{b}_{p\times 1}.$$

By the definition of projection, we have

$$\begin{aligned}
& \mathbf{y} - \hat{\mathbf{y}} \perp \mathcal{M} \\
\Longleftrightarrow \quad & \mathbf{y} - \hat{\mathbf{y}} \perp \mathbf{x}_j, \qquad \text{for } j = 1, \ldots, p \\
\Longleftrightarrow \quad & \mathbf{x}_j^t(\mathbf{y} - \hat{\mathbf{y}}) = 0, \qquad \text{for } j = 1, \ldots, p \\
\Longleftrightarrow \quad & \mathbf{X}^t(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{0}_{p\times 1} \Longleftrightarrow \mathbf{X}^t(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{0}
\end{aligned}$$

Solving the equation array above, we have

$$\mathbf{X}^t\mathbf{y} = \big(\mathbf{X}^t\mathbf{X}\big)\mathbf{b} \Longrightarrow \mathbf{b} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y},$$

therefore

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} = \mathbf{X}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y} = \mathbf{M}_{n\times n}\mathbf{y}.$$

# Projection Matrix

For a vector space $\mathcal{M}$, the projection of any vector $\mathbf{y}$ can be computed as $\hat{\mathbf{y}} = \mathbf{M}\mathbf{y}$, and $\mathbf{M}$ is called the **projection matrix** for $\mathcal{M}$.

Properties of $\mathbf{M}$:

- Unique; Exist
- For any vector $\mathbf{v} \in \mathcal{M}$, $\mathbf{M}\mathbf{v} = \mathbf{v}$ and for any vector $\mathbf{v} \perp \mathcal{M}$, $\mathbf{M}\mathbf{v} = \mathbf{0}$.
- Symmetric and Idempotent: $\mathbf{M}^t = \mathbf{M}, \mathbf{M}^2 = \mathbf{M}$. Further we have $\mathbf{M}^k = \mathbf{M}$ and $\mathbf{M}\big(\mathbf{I} - \mathbf{M}\big)$.
- $\mathbf{I} - \mathbf{M}$ is the projection matrix for vector space $\mathcal{M}^\perp$.
- Trace of a projection matrix tells us the dim of the subspace:

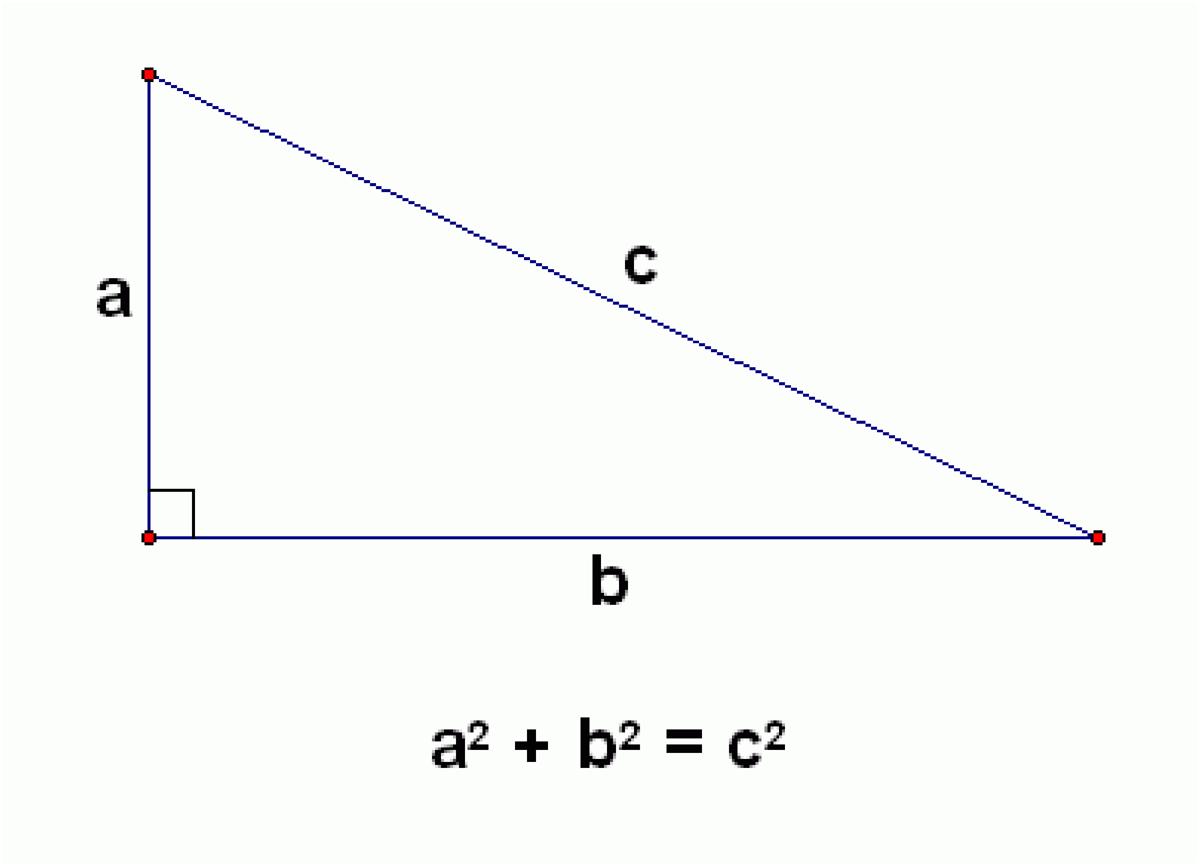
$$\text{tr}(\mathbf{M}) = p.$$

# Pythagorean Theorem

If $\mathbf{v}_1 \perp \mathbf{v}_2$ then $\|\mathbf{v}_1 + \mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2$.

In particular

$$\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}} + \mathbf{y} - \hat{\mathbf{y}}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\mathbf{y} - \hat{\mathbf{y}}\|^2$$



# LS and Projection

Recall the LS problem: find a vector $\mathbf{v}$ in $C(\mathbf{X})$, which minimizes $\|\mathbf{y}-\mathbf{v}\|^2$, i.e.,

$$\min_{\mathbf{v}\in C(\mathbf{X})} \|\mathbf{y}-\mathbf{v}\|^2.$$

Let $\mathbf{y}$ denote the projection of $\mathbf{y}$ onto $C(\mathbf{X})$. We have

$$\|\mathbf{y}-\mathbf{v}\|^2 = \| \underbrace{\mathbf{y}-\hat{\mathbf{y}}}_{\in C(\mathbf{X})^\perp} + \underbrace{\hat{\mathbf{y}}-\mathbf{v}}_{\in C(\mathbf{X})} \|^2 = \|\mathbf{y}-\hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}}-\mathbf{v}\|^2 \geq \|\mathbf{y}-\hat{\mathbf{y}}\|^2.$$

So the LS solution is the projection of $\mathbf{y}$ onto the space $C(\mathbf{X})$:

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y} = \mathbf{H}\mathbf{y}.$$

The projection matrix $\mathbf{H}$ is also called the hat matrix in many textbooks.

- $C(\mathbf{X})$ is often called the estimation space, and $C(\mathbf{X})^{\perp}$ the error space.

- $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{M})\mathbf{y}$ is the projection of $\mathbf{y}$ onto $C(\mathbf{X})^{\perp}$, the space orthogonal to $C(\mathbf{X})$. The corresponding projection matrix is $(\mathbf{I}_n - \mathbf{M})$ and dim is $n - p$.

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

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{r},$$

where $\hat{\mathbf{y}}$ in the estimation space and $\mathbf{r}$ in the error space.