Gram-Schmidt Orthonormalization

Gram-Schmidt Orthonormalization:

Suppose $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ are LI vectors in $\mathbb{R}^n$.

We seek mutually orthogonal vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$ in $\mathbb{R}^n \ni$

$$span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\} = span\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\} \quad \forall\, k = 1, \ldots, p.$$

Define

$$\boldsymbol{U}_0 = \underset{n\times 1}{\boldsymbol{0}} \quad \text{and} \quad \boldsymbol{U}_k = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k] \quad k = 1, \ldots, p,$$

where

$$\boldsymbol{u}_k = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{U}_{k-1}})\boldsymbol{x}_k \quad \forall\, k = 1, \ldots, p.$$

We will show $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$ have the desired properties.

First note that $\boldsymbol{P}_{\boldsymbol{U}_0}$ is the orthogonal projection matrix onto $\mathcal{C}(\boldsymbol{U}_0) = \mathcal{C}(\mathbf{0})$, i.e.,

$$\boldsymbol{P}_{\boldsymbol{U}_0} = \mathbf{0}(\mathbf{0}'\mathbf{0})^{-}\mathbf{0}' = \underset{n \times n}{\mathbf{0}}.$$

$$\therefore \boldsymbol{u}_1 = (\boldsymbol{I} - \mathbf{0})\boldsymbol{x}_1 = \boldsymbol{x}_1.$$

$$\boldsymbol{u}_2 = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{U}_1})\boldsymbol{x}_2 = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{x}_1})\boldsymbol{x}_2$$

$= \text{residual vector from the regression of } \boldsymbol{x}_2 \text{ on } \boldsymbol{x}_1.$

Likewise, $\boldsymbol{u}_k$ is the residual vector from the regression of $\boldsymbol{x}_k$ on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{k-1} \; \forall \; k = 3, \ldots, p$.

(This will follow if we can show $\mathcal{C}(\boldsymbol{U}_k) = span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\} \quad \forall \; k = 1, \ldots, p.$)

Now can you show

$$span\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\} = span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\} \quad \forall\, k = 1, \ldots, p?$$

Because $\boldsymbol{x}_1 = \boldsymbol{u}_1$, the result holds for $k = 1$.

Now suppose

$$span\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_l\} = span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l\}$$

for some $l \in \{1, \ldots, p-1\}$.

If we can show

$$span\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{l+1}\} = span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{l+1}\},$$

the result will follow by induction.

Recall that

$$\boldsymbol{u}_{l+1} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{U}_l})\boldsymbol{x}_{l+1} = \boldsymbol{x}_{l+1} - \boldsymbol{P}_{\boldsymbol{U}_l}\boldsymbol{x}_{l+1} \tag{1}$$

which is equivalent to

$$\boldsymbol{x}_{l+1} = \boldsymbol{u}_{l+1} + \boldsymbol{P}_{\boldsymbol{U}_l}\boldsymbol{x}_{l+1}. \tag{2}$$

We know

$$\begin{aligned} \boldsymbol{P}_{\boldsymbol{U}_l}\boldsymbol{x}_{l+1} \in \mathcal{C}(\boldsymbol{U}_l) &= span\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_l\} \\ &= span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l\}. \end{aligned}$$

Therefore,

$$(1) \Rightarrow \boldsymbol{u}_{l+1} \in span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{l+1}\} \quad \text{and}$$

$$(2) \Rightarrow \boldsymbol{x}_{l+1} \in span\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{l+1}\}.$$

We have

$$\boldsymbol{u}_{l+1} \in span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{l+1}\}$$
$$\Rightarrow span\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{l+1}\} \subseteq span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{l+1}\}.$$

Likewise,

$$\boldsymbol{x}_{l+1} \in span\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{l+1}\}$$
$$\Rightarrow span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{l+1}\} \subseteq span\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{l+1}\},$$

and the result follows by induction.

Now can you prove mutually orthogonality of $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$?

Suppose $i, j \in \{1, \ldots, p\}$ with $i < j$.

Then,

$$\begin{aligned}
\boldsymbol{u}_j'\boldsymbol{u}_i &= [(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{U}_{j-1}})\boldsymbol{x}_j]'\boldsymbol{u}_i \\
&= \boldsymbol{x}_j'(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{U}_{j-1}})'\boldsymbol{u}_i \\
&= \boldsymbol{x}_j'(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{U}_{j-1}})\boldsymbol{u}_i \\
&= \boldsymbol{x}_j'(\boldsymbol{u}_i - \boldsymbol{P}_{\boldsymbol{U}_{j-1}}\boldsymbol{u}_i) \\
&= \boldsymbol{x}_j'(\boldsymbol{u}_i - \boldsymbol{u}_i) \\
&= 0.
\end{aligned}$$

Now let $d_k = \|\boldsymbol{u}_k\| \quad \forall\, k = 1, \ldots, p.$

Define $\boldsymbol{q}_k = \frac{1}{d_k}\boldsymbol{u}_k \quad \forall\, k = 1, \ldots, p.$

Note that

$$\begin{aligned} \boldsymbol{q}_k'\boldsymbol{q}_k &= \frac{1}{d_k^2}\boldsymbol{u}_k'\boldsymbol{u}_k \\ &= \frac{1}{d_k^2}\|\boldsymbol{u}_k\|^2 \\ &= 1 \quad \forall\, k = 1, \ldots, p. \end{aligned}$$

Also $\boldsymbol{q}_i'\boldsymbol{q}_j = \frac{1}{d_i d_j}\boldsymbol{u}_i'\boldsymbol{u}_j = 0 \quad \forall\, i \neq j$.

Thus, $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_p$ are mutually orthonormal.

Furthermore,

$$\begin{aligned} span\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k\} &= span\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\} \\ &= span\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\} \quad \forall\, k = 1, \ldots, p. \end{aligned}$$

Show that $\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{R}$, where

$$\boldsymbol{X} \equiv [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p]$$

$$\boldsymbol{Q} \equiv [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_p] \quad \text{and}$$

$\boldsymbol{R}$ is an upper triangular matrix.

To see the intuition behind this result, let's look at a special case:

$$[\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3] = [\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3] \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix}.$$

A general proof is as follows.

$$\begin{aligned} X = P_X X = P_Q X &= Q(Q'Q)^- Q'X \\ &= QQ'X = QR, \quad \text{where} \end{aligned}$$

$$R = Q'X = [q_i' x_j].$$

Now, $\forall\, j = 1, \ldots, p; \quad \exists\, c_1, \ldots, c_j \in \mathbb{R} \ni \boldsymbol{x}_j = c_1\boldsymbol{q}_1 + \cdots + \boldsymbol{c}_j\boldsymbol{q}_j$.

Thus,

$$\begin{aligned} \boldsymbol{q}_i'\boldsymbol{x}_j &= c_1\boldsymbol{q}_i'\boldsymbol{q}_1 + \cdots + c_j\boldsymbol{q}_i'\boldsymbol{q}_j \\ &= 0 + \cdots + 0 \\ &= 0 \quad \forall\, i = j+1, \ldots, p. \end{aligned}$$

$\therefore r_{ij} = \boldsymbol{q}_i'\boldsymbol{x}_j = 0$ whenever $i > j$. Thus, $\boldsymbol{R} = [\boldsymbol{q}_i'\boldsymbol{x}_j]$ is upper triangular.

Note that $\boldsymbol{R}$ is unique:

Suppose $\boldsymbol{QR}_1 = \boldsymbol{QR}_2 = \boldsymbol{X}$.

Then $\boldsymbol{Q}'\boldsymbol{QR}_1 = \boldsymbol{Q}'\boldsymbol{QR}_2 \Rightarrow \boldsymbol{R}_1 = \boldsymbol{R}_2$.