Orthogonal Projection

Jeremy Teitelbaum

Orthogonal Decomposition

Let W be a subspace of $\mathbf{R}^n$. Then every vector $y \in \mathbf{R}^n$ can be written

$$y = \hat{y} + z$$

where $\hat{y} \in W$ and $z \in W^{\perp}$.

Orthogonal decomposition

To compute the decomposition, let $\{u_1, u_2, \dots, u_k\}$ be an orthogonal basis of W. Let

$$\hat{y} = \sum_{i=1}^{k} \frac{y \cdot u_i}{u_i \cdot u_i} u_i.$$

Let $z = y - \hat{y}$.

Notice that, for any $i = 1, \dots, n$,

$$z \cdot u_i = (y - \hat{y}) \cdot u_i = y \cdot u_i - \hat{y} \cdot u_i = 0$$

so $z \in W^{\perp}$.

The vector $\hat{y}$ is called the orthogonal projection of y onto W.

Example

Let

$$y = \begin{bmatrix} -1 \\ 4 \\ 3 \end{bmatrix}$$

and

$$u_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, u_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

Check that $u_1 \cdot u_2 = 0$ and then find the orthogonal projection of y into the span of $\{u_1, u_2\}$.

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2 = \frac{3}{2} u_1 + \frac{5}{2} u_2$$

so

$$\hat{y} = \begin{bmatrix} -1 \\ 4 \\ 0 \end{bmatrix}$$

Best approximation

Let W be a subspace of $\mathbf{R}^n$. Then $\hat{y} = \mathrm{proj}_W(y)$ is the point in W that is closest to y among all points in W.

If $v \in W$, then

$$\|v - y\| \geq \|\hat{y} - y\|$$

for all $v \in W$.

$$\|y - v\|^2 = \|(y - \hat{y}) + (\hat{y} - v)\|^2 = \|(y - \hat{y})\|^2 + \|(\hat{(y)} - v)\|^2 + 2(y - \hat{y}) \cdot (\hat{y} - v)$$

The dot product is zero since $y - \hat{y}$ is in $W^\perp$ and $\hat{y} - v$ is in W.

Therefore the minimum value occurs when $\hat{y} - v = 0$.

Distance to a subspace

The distance from a point y to a subspace W is by definition the length of $y - \hat{y}$ where $\hat{y}$ is the orthogonal projection onto W.

Orthonomal bases

If $\{u_1, \dots, u_p\}$ is an orthonormal basis (meaning orthogonal, but all vectors have length one), then we can let

$$U = \begin{bmatrix} u_1 & \cdots & u_p \end{bmatrix}$$

be the matrix whose columns are the u_i.

So U has n rows and p columns.

Then

$$\hat{y} = UU^T y$$

for any $y \in \mathbf{R}^n$. This is because $U^T y$ is the vector whose entries are the $u_i \cdot y$.

$U^T y$ is a vector with p entries, and $UU^T y$ is the sum of the columns of U – the u_i weighted by the elements of $U^T y$.