PCA (Principal Component Analysis)

Geometry Processing (GPR)

1 Procedure

So let us enumerate the steps of the PCA method. We want to find the principal components (vectors that best represent our data) of a $d-$dimensional point set, $p_1, \cdots, p_n \in \mathbb{R}^d$. The steps are:

1. Compute the centroid (center of mass) of the data points:

$$\tilde{\mathbf{p}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{p_i}$$

2. Translate all the points so that their origin is at $\tilde{\mathbf{p}}$:

$$\tilde{\mathbf{p}}_\mathbf{i} = \mathbf{p_i} - \tilde{\mathbf{p}}$$

3. Build the $d \times d$ covariance matrix (also known as scatter matrix):

$$\mathbf{C} = \sum_{i=1}^{n} \tilde{\mathbf{p}}_\mathbf{i} \tilde{\mathbf{p}}_\mathbf{i}^T$$

4. Compute the spectral decomposition (eigenvalues and eigenvectors) of $\mathbf{C}$:

$$\mathbf{C} = \mathbf{V} \cdot \mathbf{\Lambda} \cdot \mathbf{V}^T$$

5. Sort the eigenvalues in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$$

6. The corresponding orthonormal eigenvectors, $\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_d}$, are the principal components, sorted in order of significance (alignment to the input point set).