

Variances of higher-dimensional data sets

Given a data set $\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, $\mathbf{x}_n \in \mathbb{R}^D$, we compute the variance of the data set as

$$\mathbb{V}[\mathcal{D}] = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^\top \in \mathbb{R}^{D \times D}$$

where $\boldsymbol{\mu} \in \mathbb{R}^D$ is the mean value of the data set.