

The pca algorithm

Algorithm 1 principal component analysis

- 0 input:** m properties each measured n times
represented as m total $n \times 1$ vectors $\vec{x}^{(1)}, \dots, \vec{x}^{(m)}$
-
- 1** Populate the $m \times n$ data matrix X to have column vectors $\vec{x}^{(1)}, \dots, \vec{x}^{(m)}$, i.e. $X = [\vec{x}^{(1)} \mid \dots \mid \vec{x}^{(m)}]$
- 2** Center each column $\vec{x}^{(j)}$ of X by subtracting off its mean, i.e.

$$\vec{x}^{(j)} \leftarrow \vec{x}^{(j)} - \overline{\vec{x}^{(j)}} \quad \text{where} \quad \overline{\vec{x}^{(j)}} = \frac{1}{n} \sum_{i=1}^n X_{ij}$$

- 3** Either
- (3.1) Compute the **eigendecomposition** of the **covariance matrix**
 $C \propto X^T X$
 - (3.2) Decompose X via the **SVD** directly as $X = U \Sigma V^T$
-