C. SVD and PCA

With similar computations it is evident that the two methods are intimately related. Let us return to the original $m \times n$ data matrix $\mathbf{X}$. We can define a new matrix $\mathbf{Y}$ as an $n \times m$ matrix[9].

$$\mathbf{Y} \equiv \frac{1}{\sqrt{n-1}} \mathbf{X}^T$$

where each *column* of $\mathbf{Y}$ has zero mean. The definition of $\mathbf{Y}$ becomes clear by analyzing $\mathbf{Y}^T\mathbf{Y}$.

$$\begin{aligned}
\mathbf{Y}^T\mathbf{Y} &= \left(\frac{1}{\sqrt{n-1}}\mathbf{X}^T\right)^T \left(\frac{1}{\sqrt{n-1}}\mathbf{X}^T\right) \\
&= \frac{1}{n-1}\mathbf{X}^{TT}\mathbf{X}^T \\
&= \frac{1}{n-1}\mathbf{X}\mathbf{X}^T \\
\mathbf{Y}^T\mathbf{Y} &= \mathbf{C}_{\mathbf{X}}
\end{aligned}$$