PCA in higher dimensions

Dataset $\mathcal{D} = \{x_1, \dots, x_n\}$ $x_i \in \mathbb{R}^d$, $i = 1 \dots n$

feature dimension d $\gg$ number of datapoints n

much larger

(or, it is easier to handle $n \times n$ matrices than $d \times d$ matrices)

PCA requires finding the eigenvectors of $C = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$, $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

a $d \times d$ matrix

Goal: Re-formulate the problem as finding the eigenvectors of a $n \times n$ matrix.

Notice that $\text{rank}(C) \leq n$ (Why?)

Think about rank of $x_i x_i^T$ & then rank of $\sum_{i=1}^{n} x_i x_i^T$

$\Rightarrow$ $(d-n)$ eigenvalues of C are zero.

Thus, It is not necessary to find $(d-n)$ eigenvectors.

Let $A = \begin{bmatrix} (x_1 - \bar{x})^T \\ \vdots \\ (x_n - \bar{x})^T \end{bmatrix}$ Then, $C = \frac{1}{n} A^T A$

Let u_i be an eigenvector of C corresponding to eigenvalue $\lambda_i > 0$

Claim: λ_i is an eigenvalue of $\frac{1}{n} A A^T$ $\longrightarrow$ Then is a $n \times n$ matrix

Proof:

$$\lambda_i (A u_i) = A (\lambda_i u_i)$$

$$= A \left(\frac{1}{n} A^T A u_i \right) \rightarrow \text{since } \lambda_i \text{ is a eigenvalue of } C = \frac{1}{n} A^T A$$

$$\lambda_i (A u_i) = \frac{1}{n} A A^T (A u_i)$$

i.e, $\left(\frac{1}{n} A A^T \right) (A u_i) = \lambda_i (A u_i) \Rightarrow \lambda_i$ is an eigenvalue of $\frac{1}{n} A A^T$.

" It is enough to find eigenvectors of $\frac{1}{n} A A^T$ " because

Suppose v_i is an eigenvector of $\frac{1}{n} A A^T$, i.e.,

$$\frac{1}{n} A A^T v_i = \lambda_i v_i$$

$$\frac{1}{n} A^T A A^T v_i = \lambda_i (A^T v_i)$$

(or) $$\left(\frac{1}{n} A^T A\right)\left(A^T v_i\right) = \lambda_i (A^T v_i)$$

$\Rightarrow$ $A^T v_i$ is an eigenvector of $C = \frac{1}{n} A^T A$.

So, instead of working with a $d \times d$-matrix $C = \frac{1}{n} A^T A$ to find its eigenvalues/eigenvectors, it is enough to find eigenvalues/eigenvectors of $\frac{1}{n} A A^T$, which is a $n \times n$ matrix.

(2) PCA can be implemented efficiently in higher dimensions