Module 6.6 : PCA : Interpretation 3

- We started off with the following wishlist

- We started off with the following wishlist
- We are interested in representing the data using fewer dimensions such that

- We started off with the following wishlist
- We are interested in representing the data using fewer dimensions such that
  - the dimensions have low covariance

- We started off with the following wishlist
- We are interested in representing the data using fewer dimensions such that
  - the dimensions have low covariance
  - the dimensions have high variance

- We started off with the following wishlist
- We are interested in representing the data using fewer dimensions such that
  - the dimensions have low covariance
  - the dimensions have high variance
- So far we have paid a lot of attention to the covariance

- We started off with the following wishlist
- We are interested in representing the data using fewer dimensions such that
  - the dimensions have low covariance
  - the dimensions have high variance
- So far we have paid a lot of attention to the covariance
- It has indeed played a central role in all our analysis

- We started off with the following wishlist
- We are interested in representing the data using fewer dimensions such that
  - the dimensions have low covariance
  - the dimensions have high variance
- So far we have paid a lot of attention to the covariance
- It has indeed played a central role in all our analysis
- But what about variance? Have we achieved our stated goal of high variance along dimensions?

- We started off with the following wishlist
- We are interested in representing the data using fewer dimensions such that
    - the dimensions have low covariance
    - the dimensions have high variance
- So far we have paid a lot of attention to the covariance
- It has indeed played a central role in all our analysis
- But what about variance? Have we achieved our stated goal of high variance along dimensions?
- To answer this question we will see yet another interpretation of PCA

The i^{th} dimension of the transformed data $\hat{X}$ is given by

$$\hat{X}_i = Xp_i$$

The i^{th} dimension of the transformed data $\hat{X}$ is given by

$$\hat{X}_i = X p_i$$

The variance along this dimension is given by

The i^{th} dimension of the transformed data $\hat{X}$ is given by

$$\hat{X}_i = Xp_i$$

The variance along this dimension is given by

$$\frac{\hat{X}_i^T \hat{X}_i}{m} = \frac{1}{m} p_i^T \underbrace{X^T X p_i}$$

The i^{th} dimension of the transformed data $\hat{X}$ is given by

$$\hat{X}_i = X p_i$$

The variance along this dimension is given by

$$\begin{aligned} \frac{\hat{X}_i^T \hat{X}_i}{m} &= \frac{1}{m} p_i^T \underbrace{X^T X p_i} \\ &= \frac{1}{m} p_i^T \lambda_i p_i \qquad [\because p_i \text{ is the eigen vector of } X^T X] \end{aligned}$$

The i^{th} dimension of the transformed data $\hat{X}$ is given by

$$\hat{X}_i = X p_i$$

The variance along this dimension is given by

$$\begin{aligned}
\frac{\hat{X}_i^T \hat{X}_i}{m} &= \frac{1}{m} p_i^T \underbrace{X^T X p_i} \\
&= \frac{1}{m} p_i^T \lambda_i p_i \qquad [\because p_i \text{ is the eigen vector of } X^T X] \\
&= \frac{1}{m} \lambda_i \underbrace{p_i^T p_i}_{=1}
\end{aligned}$$

The i^{th} dimension of the transformed data $\hat{X}$ is given by

$$\hat{X}_i = X p_i$$

The variance along this dimension is given by

$$\begin{aligned}
\frac{\hat{X}_i^T \hat{X}_i}{m} &= \frac{1}{m} p_i^T \underbrace{X^T X p_i} \\
&= \frac{1}{m} p_i^T \lambda_i p_i \qquad [\because p_i \text{ is the eigen vector of } X^T X] \\
&= \frac{1}{m} \lambda_i \underbrace{p_i^T p_i}_{=1} \\
&= \frac{\lambda_i}{m}
\end{aligned}$$

The i^{th} dimension of the transformed data $\hat{X}$ is given by

$$\hat{X}_i = X p_i$$

The variance along this dimension is given by

$$\begin{aligned}
\frac{\hat{X}_i^T \hat{X}_i}{m} &= \frac{1}{m} p_i^T \underbrace{X^T X p_i} \\
&= \frac{1}{m} p_i^T \lambda_i p_i \qquad\qquad [\because p_i \text{ is the eigen vector of } X^T X] \\
&= \frac{1}{m} \lambda_i \underbrace{p_i^T p_i}_{=1} \\
&= \frac{\lambda_i}{m}
\end{aligned}$$

- Thus the variance along the i^{th} dimension (i^{th} eigen vector of $X^T X$) is given by the corresponding (scaled) eigen value.

The i^{th} dimension of the transformed data $\hat{X}$ is given by

$$\hat{X}_i = X p_i$$

The variance along this dimension is given by

$$\begin{aligned}
\frac{\hat{X}_i^T \hat{X}_i}{m} &= \frac{1}{m} p_i^T \underbrace{X^T X p_i} \\
&= \frac{1}{m} p_i^T \lambda_i p_i && [\because p_i \text{ is the eigen vector of } X^T X] \\
&= \frac{1}{m} \lambda_i \underbrace{p_i^T p_i}_{=1} \\
&= \frac{\lambda_i}{m}
\end{aligned}$$

- Thus the variance along the i^{th} dimension (i^{th} eigen vector of $X^T X$) is given by the corresponding (scaled) eigen value.
- Hence, we did the right thing by discarding the dimensions (eigenvectors) corresponding to lower eigen values!

A Quick Summary

We have seen 3 different interpretations of PCA

A Quick Summary

We have seen 3 different interpretations of PCA

- It ensures that the covariance between the new dimensions is minimized

A Quick Summary

We have seen 3 different interpretations of PCA

- It ensures that the covariance between the new dimensions is minimized
- It picks up dimensions such that the data exhibits a high variance across these dimensions

A Quick Summary

We have seen 3 different interpretations of PCA

- It ensures that the covariance between the new dimensions is minimized
- It picks up dimensions such that the data exhibits a high variance across these dimensions
- It ensures that the data can be represented using less number of dimensions