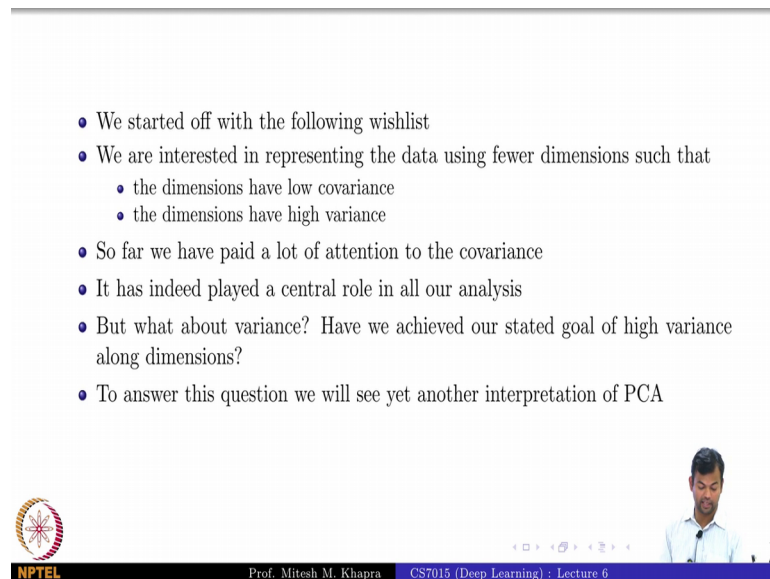


Deep Learning
Prof. Mithesh M. Khapra
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Module – 6.6
Lecture - 06
PCA: Interpretation 3

And now you go to the third interpretation, where we will try to say something about the variance ok.

(Refer Slide Time: 00:18)

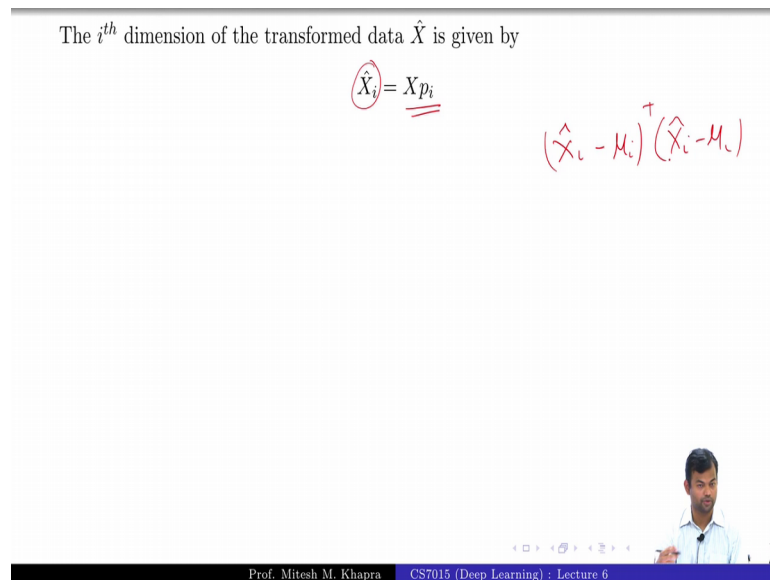


- We started off with the following wishlist
- We are interested in representing the data using fewer dimensions such that
 - the dimensions have low covariance
 - the dimensions have high variance
- So far we have paid a lot of attention to the covariance
- It has indeed played a central role in all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

So, we started off with the following the wish list, that we wanted low covariance and we wanted high variance. So, far we have paid attention to the covariance because everything was revolving around this covariance matrix in both the solutions ah, but what about variance have we achieved the goal with respect to high variance.

(Refer Slide Time: 00:40)

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

$$\hat{X}_i = X p_i$$
$$(\hat{X}_i - \mu_i) (\hat{X}_i - \mu_i)^{\top}$$


Prof. Mitesh M. Khapra CS7015 (Deep Learning) : Lecture 6

So, let us see. So, what is the i^{th} dimension of the transformed data it is this, you take your data and project it onto the i^{th} dimension right. So, \hat{X} is equal to X into p_i now what is the variance along this dimension, how do you compute the variance? So, this is my projected data and let me just call it \hat{X}_i .

So, this is the i^{th} column after projection is that fine everyone is with this? Now, for this i^{th} column I want to compute the variance; how will I do that? Remember that the data is zero mean what is the formula actually? It is going to be \hat{X}_i minus μ_i into \hat{X}_i minus μ_i right, but μ_i is 0. So, it just turns out to be the dot product dot product of \hat{X}_i with itself and of course, divided by m is this fine?

(Refer Slide Time: 01:34)



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

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

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



NPTEL Prof. Mitesh M. Khapra CS7015 (Deep Learning) : Lecture 6

I can write this as $X p_i$ and then when I take the transpose I will get this ok. Now what is this quantity? This is exactly the moment where I feel like saying [FL] what is this quantity?

Student: (Refer Time: 02:00).

No look at the circle, what is $X^T X p_i$?

Student: (Refer Time: 02:08).

What is p_i with respect to $X^T X$?

Student: Eigenvector eigenvector

Eigenvector so what is this product going to be?

Student: λ_i (Refer Time: 02:14).

$\lambda_i p_i$ is that fine, what is $p_i^T p_i$?

Student: 1.

Ok so what is actually the variance along the i^{th} dimension?

Student: λ_i (Refer Time: 02:26).

What is λ_i ?

Student: Eigenvalue.

So, what will happen if I retain the highest eigenvalues?

Student: (Refer Time: 02:33).

I will get the highest variance dimensions right fine. So, all roads lead to.

Student: (Refer Time: 02:39).

Eigenvectors eigenvalues right. So, Andrew Ng in one of his lecture says that there are 10 different interpretations of PCA, I only know 3 of these I do not know the remaining 7. Maybe he was bluffing so that people like ask him keep busy how this is getting recorded fine.



So ya, so, you get this. So, we have satisfied everything in our wish list variance, covariance and also did this detour where we saw that it actually amounts to minimizing the error in reconstruction where we are throwing away the dimensions, along which reconstruction did not add much to our knowledge about the data ok. So, these are the three different interpretations that I have right. So, hence we did the right thing by throwing away those dimensions, which correspond to the lowest eigenvalues because lowest eigenvalues is nothing the lowest variance also ok.

(Refer Slide Time: 03:24)

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



NPTEL Prof. Mitesh M. Khapra CS7015 (Deep Learning) : Lecture 6

So, this is the quick summary, the covariance between the new dimensions, you can leave actually those you can just read it later on.