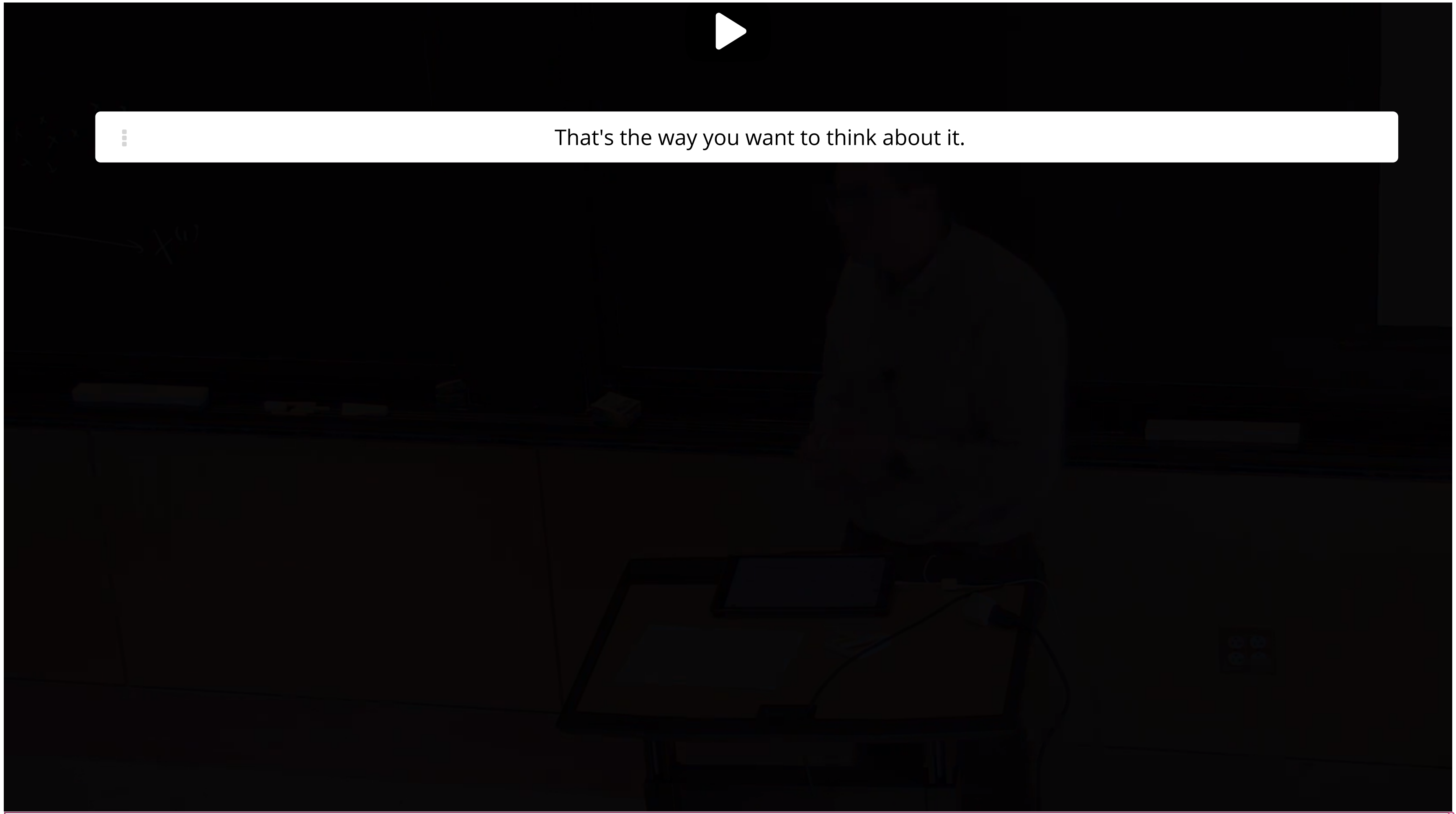
2. Introduction

Introduction to Principal Component Analysis



4:49 / 4:49 1.50x

Video

[Download video file]

Transcripts

[Download SubRip (.srt) file]

[Download Text (.txt) file]

Concept Check: Purpose of Principal Component Analysis

1/1 point (ungraded)

What is the goal of Principal Component Analysis (PCA)?

- [ ] To predict the outcome of a dependent variable based on covariates.
- [x] To visualize a high-dimensional point cloud (for example, 100-dimensional) in a lower dimension (for example, dimension 2 or 3).
- [ ] To find a non-informative prior that can be updated as we receive more information about our point cloud.

✔

Solution:

We examine the choices in order.

- The first choice **"To predict the outcome of a dependent variable based on covariates."** is incorrect. This was the goal of linear regression from a previous chapter. In PCA, we are not interested in making predictions but rather want to better understand a complicated data-set.

- The second choice **"To visualize a high-dimensional point cloud in a lower dimension (for example, dimension 2 or 3)."** is correct. In practice, it would be impossible to visualize a data set where each data point is a 100-dimensional vector. However, if we can somehow embed our data set in a much lower dimension (e.g. $d = 2$ or 3) while preserving the data set's salient statistical features, then it would be possible to visualize a very complicated data set. **Remark**: PCA attempts to find a low-dimensional representation of the data while preserving the *covariance structure* of our point cloud as accurately as possible.

- The third choice **"To find a non-informative prior that can be updated as we receive more information about our point cloud."** is incorrect. This more so describes a Bayesian statistical set-up, which is not our focus here.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Number of Dimensions Required: Intuition I

0/1 point (ungraded)

Note: This problem builds intuition for what we tackle in principal component analysis later in the lecture. Choose an answer based on intuition and the given numbers, and not based on rigorous mathematical equations or proofs.

Let $\mathbf{X}_i, i = 1, \ldots, n$ be iid random vectors in $\mathbb{R}^d$. Say $X_i^{(1)}, \ldots, X_i^{(d-k)}$, where $0 < k < d$, take values in the real interval $[10, 25]$ and that the variances of $X_i^{(1)}, \ldots, X_i^{(d-k)}$ are of the order 10^{-5}. How many dimensions would we need to visualize the data $\mathbf{X}_i, i = 1, \ldots, n$ if we wish to reduce the dimensionality of the data?

1

✖ **Answer:** k+0*d

1

STANDARD NOTATION

Solution:

It is given that the first $d - k$ random variables of the random vector take values in the interval $[10, 25]$ and that their variances are very small (of the order 10^{-5}) compared to the values that they take (in $[10, 25]$). We would therefore expect that we should be able to visualize the data points $\mathbf{X}_i, i = 1, \ldots, n$ in k dimensions.

Submit

You have used 2 of 2 attempts

Answers are displayed within the problem

Number of Dimensions Required: Intuition II

1/1 point (ungraded)

Let $\mathbf{X}_i, i = 1, \ldots, n$ be iid random vectors in $\mathbb{R}^3$. Let $X_i^{(1)}$ and $X_i^{(2)}$ be such that $X_i^{(1)} + X_i^{(2)} = c$, for some constant c. How many dimensions, **at most**, would we need to visualize the data $\mathbf{X}_i, i = 1, \ldots, n$ if we wish to reduce the dimensionality of the data?

2

✔ **Answer:** 2

Solution:

Since $X_i^{(1)} + X_i^{(2)} = c$, i.e. one of the random variables is a deterministic function of another, we expect that we should be able to visualize the data in at most 2 dimensions.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Review of Covariance of a Random Vector

5:36 / 5:36 1.50x

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Covariance Matrix: Review

1/1 point (ungraded)
Let $\mathbf{X}$ be a random vector in $\mathbb{R}^d$. The elements of this random vector are obtained as follows.

Let us toss a coin with p as its probability of heads d times independently. Let $X^{(i)}, i = 1, \ldots, d$ be the number of heads obtained in the first i tosses. Let Σ be the covariance matrix of $\mathbf{X}$.

Let σ_i^2 denote the variance of $X^{(i)}$.

Compute the covariance of $X^{(i)}$ and $X^{(j)}$, $\textsf{Cov}\left(X^{(i)}, X^{(j)}\right)$, when $j > i$.

sigma_i^2

Answer: sigma_i^2

σ_i^2

STANDARD NOTATION

Solution:

The covariance of $X^{(i)}$ and $X^{(j)}$ when $i \neq j$, denoted $\textsf{Cov}\left(X^{(i)}, X^{(j)}\right)$, can be computed as follows. Let $j > i$, without loss of generality. This means, $X^{(j)} = X^{(i)} + \sum_{k=i+1}^{j} Y_k$, where Y_k are iid with distribution $\textsf{Ber}(p)$ and $\left\{X^{(i)}, Y_{i+1}, \ldots, Y_j\right\}$ are independent (by definition of the experiment).

Using the bilinearity of covariance,

$$\textsf{Cov}\left(X^{(i)}, X^{(j)}\right) = \textsf{Cov}\left(X^{(i)}, \left(X^{(i)} + \sum_{k=i+1}^{j} Y_k\right)\right)$$

$$
\begin{aligned}
&= \mathsf{Cov}\left(X^{(i)}, X^{(i)}\right) + \sum_{k=i+1}^{j} \mathsf{Cov}\left(X^{(i)}, Y_k\right) \\
&= \mathsf{Var}\left(X^{(i)}\right) = ip\,(1-p)\,,
\end{aligned}
$$

since $X^{(i)} \sim \mathsf{Binom}\,(i, p)$.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: (Optional) Unit 8 Principal component analysis:(Optional) Lecture 23: Principal Component Analysis / 2. Introduction

Add a Post

Show all posts ▼ by recent activity ▼

There are no posts in this topic yet.

✖