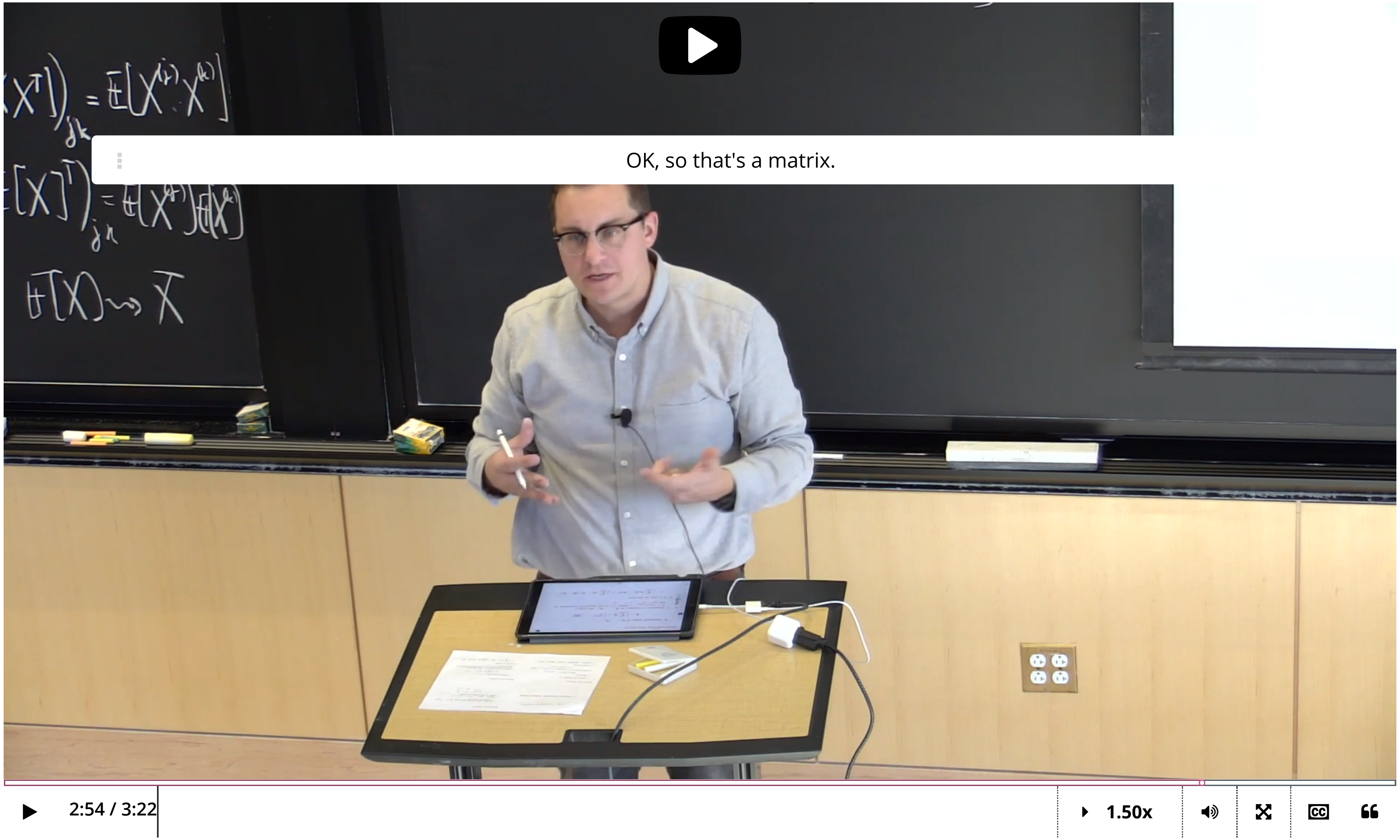
3. Multivariate Statistics and Geometry Behind the Empirical Covariance

Empirical Covariance Matrix



OK, so that's a matrix.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Bias

1/1 point (ungraded)

Let $\mathbf{X}_i, i = 1, \ldots, n$ be iid data points in $\mathbb{R}^d$. As presented in the lecture and given in the slides, let S be the empirical covariance matrix

$$S \triangleq \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i\mathbf{X}_i^T\right) - \overline{\mathbf{X}}\,\overline{\mathbf{X}}^T,$$

where $\overline{\mathbf{X}}$ is the empirical or sample mean $\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i$.

Is the following **true or false**. "S is an unbiased estimator of the covariance matrix Σ of $\mathbf{X}_i$'s."

○ True

● False

✓

Solution:

The answer is **False**. We have seen before in Lecture 10 that the correct fraction for obtaining an unbiased estimator for the sample variance or the sample covariance is $\frac{1}{n-1}$ and not $\frac{1}{n}$. The same carries over to covariance matrices, as well (Self-exercise: Verify this statement).

Submit You have used 1 of 1 attempt

ℹ Answers are displayed within the problem

Projection Onto a Subspace: Example

0/1 point (ungraded)
Note: This problem is discussed briefly at the beginning of the following video.

Recall the matrix solution to the linear regression problem $\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}\|^2$:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbb{X}^T\mathbb{X}\right)^{-1}\mathbb{X}^T\mathbf{Y},$$

where $\mathbb{X} \in \mathbb{R}^{n\times d}$, $\mathbf{Y} \in \mathbb{R}^n$, and $\boldsymbol{\beta} \in \mathbb{R}^d$ and where we assume that $\mathbb{X}$ is full column rank, i.e. $\operatorname{rank}(\mathbb{X}) = d$.

Choose from the following the correct geometric interpretation of $\mathbb{X}\widehat{\boldsymbol{\beta}}$. That is,

○ $\mathbb{X}\widehat{\boldsymbol{\beta}}$ is a vector in the column space of $\mathbb{X}$ and is the orthogonal projection of $\mathbf{Y}$ onto the column space of $\mathbb{X}$. ✓

● $\mathbb{X}\widehat{\boldsymbol{\beta}}$ is a vector in the row space of $\mathbb{X}$ and is the orthogonal projection of $\mathbf{Y}$ onto the row space of $\mathbb{X}$.

✗

Solution:

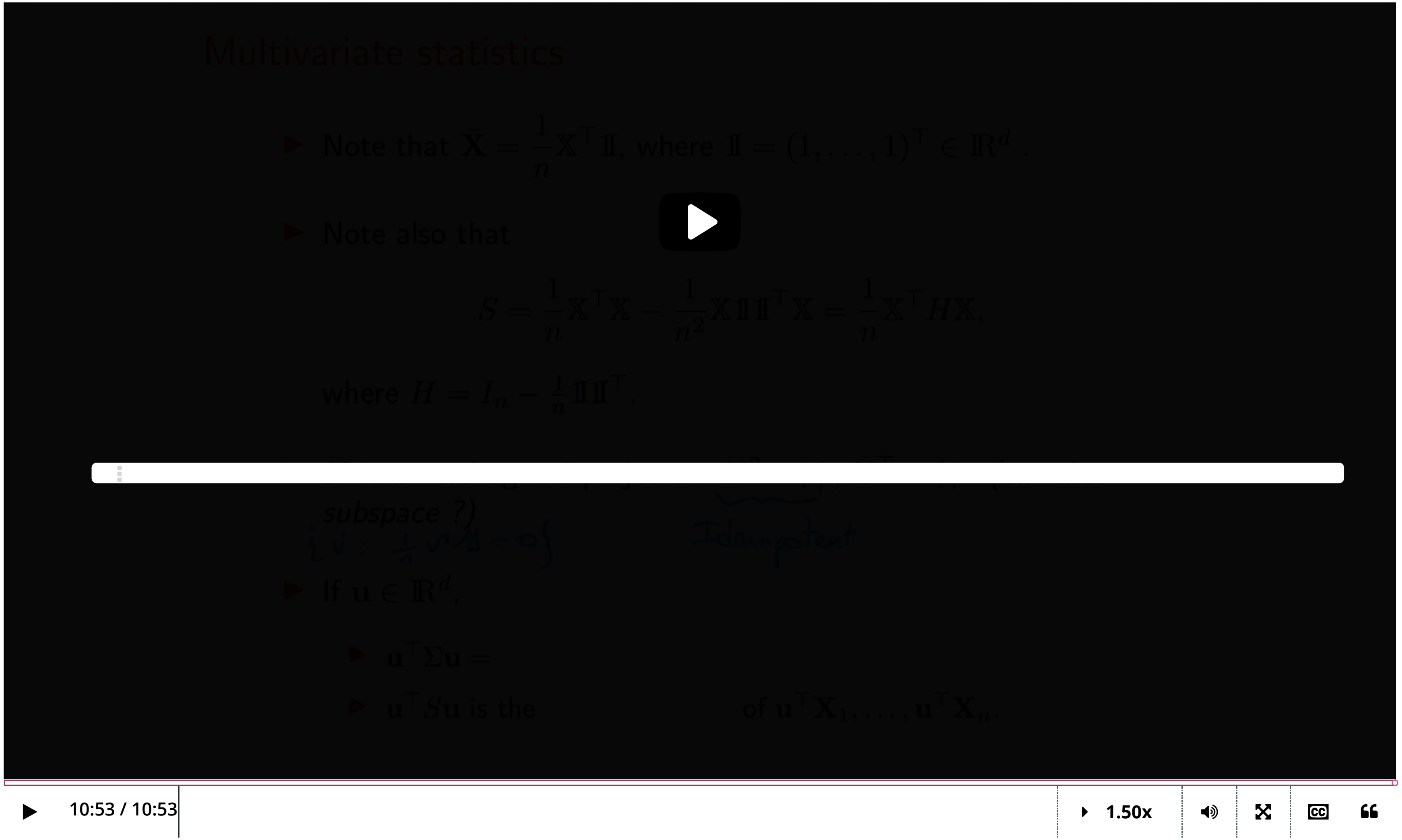
By definition of column space, $\mathbb{X}\widehat{\boldsymbol{\beta}}$ is a vector in the column space of $\mathbb{X}$.

The orthogonal projection of a vector $\mathbf{Y}$ onto the column space of a matrix such as $\mathbb{X}$, which is full column rank, is (in linear algebra theory) equal to $\mathbb{X}\left(\mathbb{X}^T\mathbb{X}\right)^{-1}\mathbb{X}^T\mathbf{Y}$. The important observation in the linear least squares problem is that the solution $\widehat{\boldsymbol{\beta}}$ to the linear least squares problem turns out to have the property that $\mathbb{X}\widehat{\boldsymbol{\beta}}$ is the orthogonal projection of $\mathbf{Y}$ onto the column space of $\mathbb{X}$.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Geometric View of Empirical Covariance



Multivariate statistics

- Note that $\bar{\mathbf{X}} = \frac{1}{n}\mathbb{X}^\top \mathbb{1}$, where $\mathbb{1} = (1,\dots,1)^\top \in \mathbb{R}^d$.
- Note also that

$$S = \frac{1}{n}\mathbb{X}^\top\mathbb{X} - \frac{1}{n^2}\mathbb{X}\mathbb{1}\mathbb{1}^\top\mathbb{X} = \frac{1}{n}\mathbb{X}^\top H\mathbb{X},$$

where $H = I_n - \frac{1}{n}\mathbb{1}\mathbb{1}^\top$.

subspace ?)

$\{v : \frac{1}{n} v^\top \mathbb{1} = 0\}$

Idempotent

- If $\mathbf{u} \in \mathbb{R}^d$,
  - $\mathbf{u}^\top \Sigma \mathbf{u} =$
  - $\mathbf{u}^\top S \mathbf{u}$ is the of $\mathbf{u}^\top\mathbf{X}_1, \dots, \mathbf{u}^\top\mathbf{X}_n$.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

A matrix $P \in \mathbb{R}^{n\times n}$ is called an **orthogonal projection matrix** if and only if $P^2 = P = P^T$. A matrix with the property that $P^2 = P$ is also called **idempotent** .

Projection: Zero Mean Vectors in 2 Dimensions

1/1 point (ungraded)

Consider the projection matrix introduced in the above video:

$$H = \mathbf{I}_n - \frac{1}{n}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}_{n\times 1}\begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}_{1\times n}$$

Let $n = 2$. What is the subspace that H, with $n = 2$, projects any vector $\mathbf{x} \in \mathbb{R}^2$ onto?

○ $\{\mathbf{y} : y^{(1)} - y^{(2)} = 0\}$

● $\{\mathbf{y} : \frac{y^{(1)}+y^{(2)}}{2} = 0\}$

○ $\{\mathbf{y} : y^{(1)} = \frac{1}{2}y^{(2)}\}$

✓

Solution:

As shown in the video, the projection matrix H projects any given vector $\mathbf{x}$ onto the space of mean-removed vectors, which is choice 2.

The same solution can also be arrived at by looking at the columns of H:

$$H = \begin{bmatrix}\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

Notice that the two columns are linearly dependent (through a sign change). The column space of H is all vectors that are scalar multiples of

$$\begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix},$$

which is the line $y^{(1)} = -y^{(2)}$.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: (Optional) Unit 8 Principal component analysis:(Optional) Lecture 23: Principal Component Analysis / 3. Multivariate Statistics and Geometry Behind the Empirical Covariance

Add a Post

Show all posts ▼ by recent activity ▼

There are no posts in this topic yet.

✖