DSC 140B - Quiz 04

January 29, 2026

Name: ____________________ PID: ____________

About the quizzes:

- Quizzes in DSC 140B are *optional* and graded pass/fail.
- A score of 70% or higher earns a "pass" and 1.5 credits toward your final grade.
- If you don't pass, no credits are earned, but it doesn't hurt your grade.
- You have 30 minutes to complete the quiz.
- At least one of the questions below will be on an exam (probably with slight changes, such as different numbers).
- Unfortunately, we can't answer clarifying questions during the quiz. If you think a question has a bug or is unclear, please let us known in a private post on Campuswire after the quiz, and we'll take it into account when grading.

Problem 1.

Suppose the direction of maximum variance in a centered data set is

$$\vec{u} = \left(\frac{2}{3}, \frac{2}{3}, \frac{1}{3}\right)^T$$

Let $\vec{x} = (4, 2, 3)^T$ be a centered data point.

Reduce $\vec{x}$ to one dimension by projecting onto the direction of maximum variance. What is the new feature z obtained from this projection?

- ○ 3
- ○ 5
- ○ 7
- ○ $\frac{14}{3}$

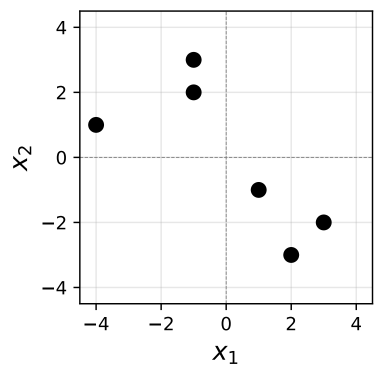
Problem 2.

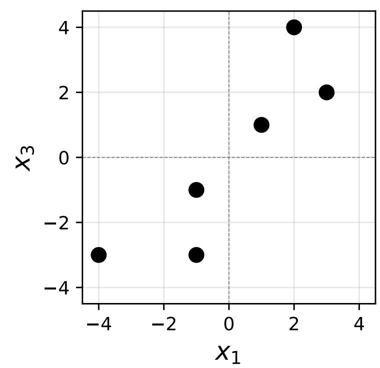
Suppose $C = \begin{pmatrix} 4 & 2 \\ 2 & 7 \end{pmatrix}$ is the empirical covariance matrix for a centered data set. What is the variance in the direction given by the unit vector $\vec{u} = \frac{1}{\sqrt{5}}(1, 2)^T$?

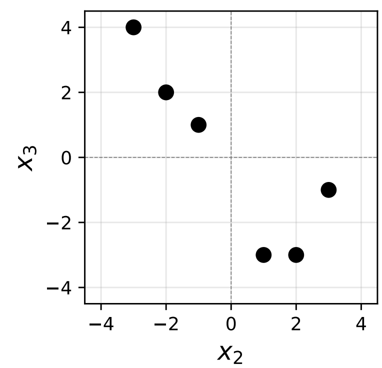
- ○ 4
- ○ 6
- ○ 8
- ○ 11

Problem 3.

Let $\vec{x}^{(1)}, \ldots, \vec{x}^{(6)}$ be a data set of 6 points in $\mathbb{R}^3$. Shown below are scatter plots of each pair of coordinates (pay close attention to the axis labels):







Which one of the following could possibly be the data's sample covariance matrix?

- ◯ $\begin{pmatrix} 7 & 3 & -2 \\ 3 & 5 & -4 \\ -2 & -4 & 8 \end{pmatrix}$
- ◯ $\begin{pmatrix} 6 & -4 & 3 \\ -4 & 5 & -5 \\ 3 & -5 & 9 \end{pmatrix}$
- ◯ $\begin{pmatrix} 8 & -2 & -3 \\ -2 & 6 & 4 \\ -3 & 4 & 7 \end{pmatrix}$
- ◯ $\begin{pmatrix} 5 & 4 & 5 \\ 4 & 7 & -3 \\ 5 & -3 & 6 \end{pmatrix}$

Problem 4.

Let C be the sample covariance matrix of a centered data set $\mathcal{X}$ consisting of four points. Suppose that PCA is performed to reduce the dimensionality of $\mathcal{X}$ to one dimension. The results are:

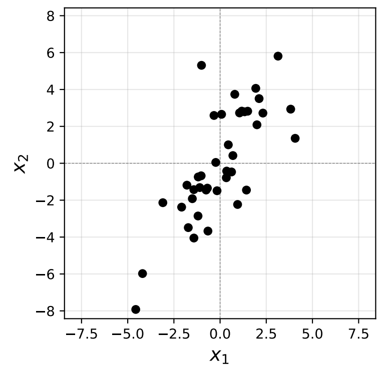
$$z^{(1)} = 2$$
$$z^{(2)} = -4$$
$$z^{(3)} = 6$$
$$z^{(4)} = -4$$

What is the largest eigenvalue of C?

- ◯ 9
- ◯ 14
- ◯ 18
- ◯ 72

Problem 5.

Consider the data set shown below:



Which of the following could possibly be the top eigenvector of the data's sample covariance matrix?

- ◯ $(1, -2)^T$
- ◯ $(2, 1)^T$
- ◯ $(1, 2)^T$
- ◯ $(0, 1)^T$

Problem 6.

Let C be the sample covariance matrix of a data set in $\mathbb{R}^3$, and suppose $\vec{u}^{(1)}, \vec{u}^{(2)}, \vec{u}^{(3)}$ are orthonormal eigenvectors of C with eigenvalues $\lambda_1 = 7, \lambda_2 = 4, \lambda_3 = 1$, respectively, where:

$$\vec{u}^{(1)} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}, \quad \vec{u}^{(2)} = \begin{pmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}, \quad \vec{u}^{(3)} = \begin{pmatrix} 1/\sqrt{6} \\ -1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}$$

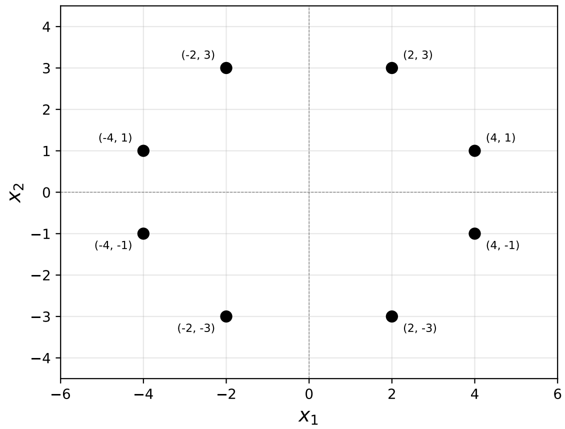
Suppose a data point is $\vec{x} = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix}$.

If PCA is performed to reduce the dimensionality from 3 to 2, what is the new representation of $\vec{x}$?

- ◯ $\begin{pmatrix} \sqrt{2} \\ 2\sqrt{3} \end{pmatrix}$
- ◯ $\begin{pmatrix} 4\sqrt{2} \\ \sqrt{3} \end{pmatrix}$
- ◯ $\begin{pmatrix} 2\sqrt{2} \\ 2\sqrt{3} \end{pmatrix}$
- ◯ $\begin{pmatrix} 2\sqrt{3} \\ 2\sqrt{2} \end{pmatrix}$

Problem 7.

Suppose PCA is used to reduce the dimensionality of the centered data shown below from 2 dimensions to 1.

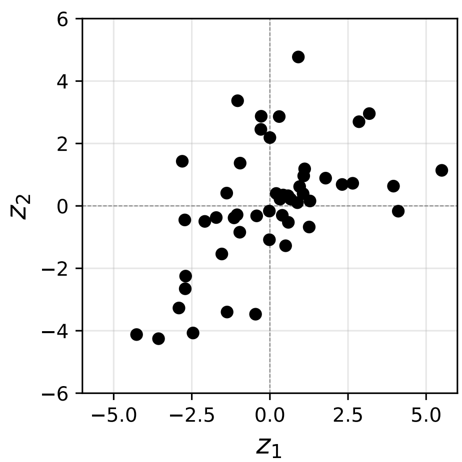


What will be the reconstruction error?

- ◯ 20
- ◯ 32
- ◯ 40
- ◯ 80

Problem 8.

Your friend claims that they have performed PCA on a 100-dimensional data set and reduced its dimensionality to two dimensions. They show you the following scatter plot of the their result:



True or false: this plot could show the data after PCA has been performed. That is, z_1 and z_2 could be the first two PCA features of the data.

- ◯ True
- ◯ False

Problem 9.

You and your friend are both given the same centered data set in $\mathbb{R}^{100}$. You perform PCA to reduce the dimensionality directly from 100 to 50. Your friend does this the hard way, in two steps: first, they reduce from 100 to 75 dimensions with PCA, then they run PCA *again* on this 75-dimensional data set to reduce it to 50 dimensions.

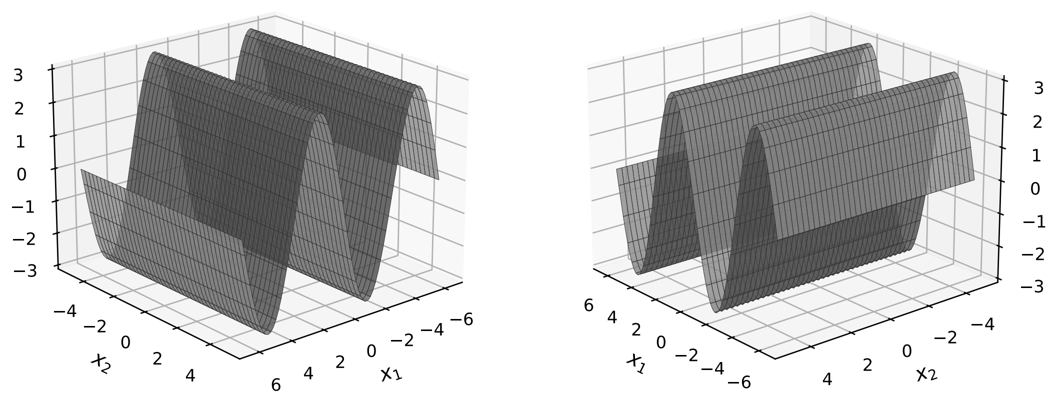
True or false: you and your friend must arrive at the exact same 50-dimensional data set.

You may assume that whenever eigenvectors are computed, there is no sign ambiguity. That is, you and your friend always obtain the exact same eigenvectors.

- ○ True
- ○ False

Problem 10.

Suppose data lies exactly on the surface shown below. The surface is shown from two angles for clarity.



a) What is the ambient dimension?

b) What is the intrinsic dimension?