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
- ○ $\frac{14}{3}$

Solution: $z = 5$.

The projection onto the direction of maximum variance is given by the dot product with $\vec{u}$:

$$z = \vec{x} \cdot \vec{u} = (4, 2, 3) \cdot \left(\frac{2}{3}, \frac{2}{3}, \frac{1}{3}\right) = \frac{8 + 4 + 3}{3} = \frac{15}{3} = 5$$

This was similar to Practice Problem 59.

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

Solution: 8.

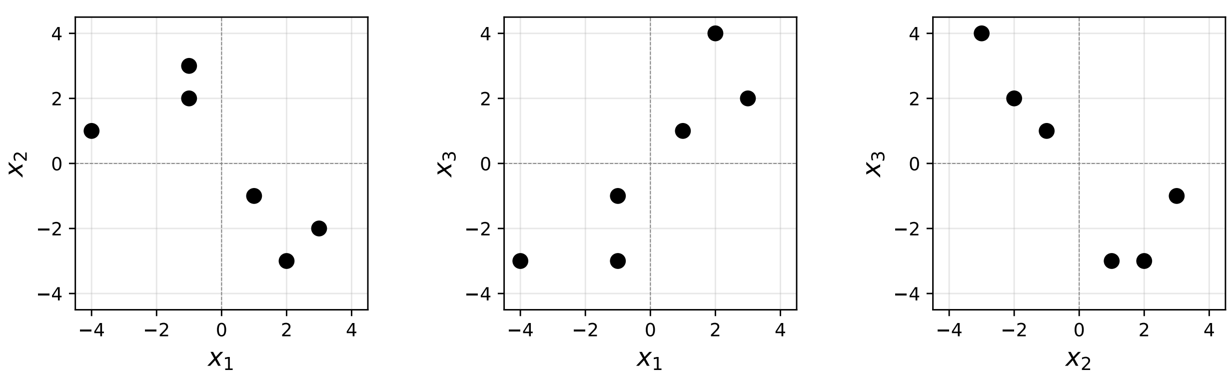
The variance in the direction of a unit vector $\vec{u}$ is given by $\vec{u}^T C \vec{u}$.

$$\begin{aligned}
\vec{u}^T C \vec{u} &= \frac{1}{\sqrt{5}}(1,2)^T \begin{pmatrix} 4 & 2 \\ 2 & 7 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\
&= \frac{1}{5}(1,2)^T \begin{pmatrix} 4+4 \\ 2+14 \end{pmatrix} \\
&= \frac{1}{5}(1,2)^T \begin{pmatrix} 8 \\ 16 \end{pmatrix} \\
&= \frac{1}{5}(8+32) \\
&= \frac{40}{5} = 8
\end{aligned}$$

This was similar to Practice Problem 67.

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?

- ○ $\begin{pmatrix} 7 & 3 & -2 \\ 3 & 5 & -4 \\ -2 & -4 & 8 \end{pmatrix}$

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

Solution: $\begin{pmatrix} 6 & -4 & 3 \\ -4 & 5 & -5 \\ 3 & -5 & 9 \end{pmatrix}$

From the scatter plots, we can determine the signs of the covariances. The (x_1, x_2) plot shows a negative correlation, so $C_{12} < 0$. The (x_1, x_3) plot shows a positive correlation, so $C_{13} > 0$. The (x_2, x_3) plot shows a negative correlation, so $C_{23} < 0$. The only matrix with $C_{12} < 0$, $C_{13} > 0$, and $C_{23} < 0$ is the second choice.

This was similar to Practice Problem 65.

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:

$$\begin{aligned} z^{(1)} &= 2 \\ z^{(2)} &= -4 \\ z^{(3)} &= 6 \\ z^{(4)} &= -4 \end{aligned}$$

What is the largest eigenvalue of C?

- [ ] 9
- [ ] 14
- [x] 18
- [ ] 72

Solution: 18.

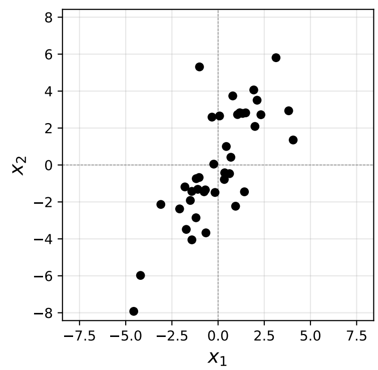
The largest eigenvalue of C equals the variance of the projected data along the first principal component. This data has mean zero (we can verify: $(2 - 4 + 6 - 4)/4 = 0$). So the variance is:

$$\begin{aligned} \lambda_1 &= \frac{1}{4} \sum_{i=1}^{4} (z^{(i)})^2 \\ &= \frac{1}{4}(4 + 16 + 36 + 16) \\ &= \frac{72}{4} = 18 \end{aligned}$$

This was similar to Practice Problem 69.

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?

- ○ $(1, -2)^T$
- ○ $(2, 1)^T$
- ● $(1, 2)^T$
- ○ $(0, 1)^T$

Solution: $(1, 2)^T$

The top eigenvector of the covariance matrix points in the direction of greatest variance. Looking at the scatter plot, the data is elongated along a direction with positive slope, rising roughly twice as fast in the x_2 direction as in the x_1 direction. The vector $(1, 2)^T$ points in this direction.

$(2, 1)^T$ has too shallow a slope. $(1, -2)^T$ points in the wrong direction (negative slope). $(0, 1)^T$ is vertical, which doesn't match the data's elongation.

This was similar to Practice Problem 70.

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 $\vec{x}$?

- ○ $\begin{pmatrix} \sqrt{2} \\ 2\sqrt{3} \end{pmatrix}$

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

Solution: $\begin{pmatrix} 2\sqrt{2} \\ 2\sqrt{3} \end{pmatrix}$

In PCA, to reduce from d dimensions to k dimensions, we project each data point onto the top k eigenvectors (those with the largest eigenvalues).

Here, the top 2 eigenvectors are $\vec{u}^{(1)}$ (with $\lambda_1 = 7$) and $\vec{u}^{(2)}$ (with $\lambda_2 = 4$).

The new representation is obtained by computing the dot product of $\vec{x}$ with each of the top k eigenvectors:

$$\vec{x} \cdot \vec{u}^{(1)} = \frac{3}{\sqrt{2}} + \frac{1}{\sqrt{2}} + 0 = \frac{4}{\sqrt{2}} = 2\sqrt{2}$$
$$\vec{x} \cdot \vec{u}^{(2)} = \frac{3}{\sqrt{3}} - \frac{1}{\sqrt{3}} + \frac{4}{\sqrt{3}} = \frac{6}{\sqrt{3}} = 2\sqrt{3}$$

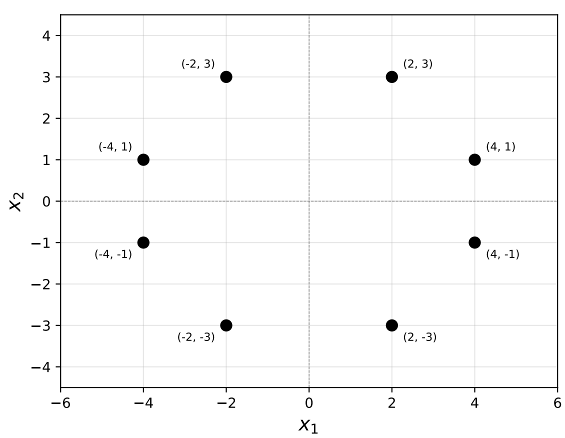
Therefore, the new representation is:

$$\vec{z} = \begin{pmatrix} \vec{x} \cdot \vec{u}^{(1)} \\ \vec{x} \cdot \vec{u}^{(2)} \end{pmatrix} = \begin{pmatrix} 2\sqrt{2} \\ 2\sqrt{3} \end{pmatrix}$$

This was similar to Practice Problem 79.

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

Solution: 40.

The first thing to figure out is the first principal component (first eigenvector of the covariance matrix), since this is the direction onto which the data will be projected. The first eigenvector points in the direction of maximum variance, and in this plot the data is wider horizontally than vertically. Thus, the first principal component is $(1, 0)^T$ (or $(-1, 0)^T$).

When we project all of the data onto the x_1-axis, each point's x_2-coordinate is lost and becomes zero, while the x_1-coordinate remains unchanged.

The reconstruction error is the sum of squared distances between each point's original position and its projected position. Since only the x_2-coordinate changes, each point contributes $(x_2)^2$ to the error.

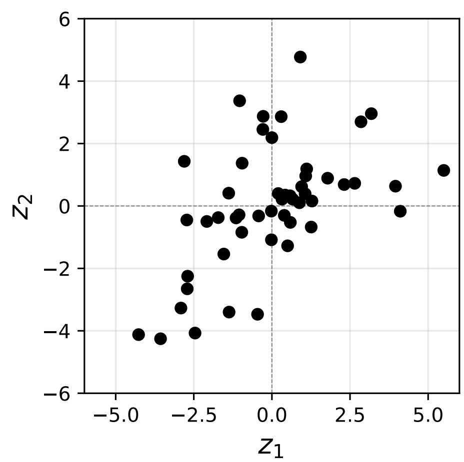
There are four points with $|x_2| = 1$ and four points with $|x_2| = 3$:

$$4 \cdot 1^2 + 4 \cdot 3^2 = 4 + 36 = 40$$

This was similar to Practice Problem 86.

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

Solution: False.

After PCA, the data is represented in the eigenbasis of the covariance matrix. In this basis, the covariance matrix is diagonal, which means the coordinates are uncorrelated. But the scatter plot clearly shows a positive correlation between x_1 and x_2 (the data trends upward from left to right), so the covariance matrix is not diagonal. Therefore, this cannot be data in its PCA coordinates.

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

Solution: True.

When you perform PCA directly from 100 to 50 dimensions, you project the data onto the top 50 eigenvectors of the original 100×100 covariance matrix.

Your friend's first step projects onto the top 75 eigenvectors of the same covariance matrix, producing 75-dimensional data. The covariance matrix of this projected data is diagonal, with the top 75 eigenvalues $\lambda_1 \geq \cdots \geq \lambda_{75}$ on the diagonal. But note that the first 50 coordinates they get after this first projection are exactly the same as the coordinate you get after your projection, since both are projections onto the same top 50 eigenvectors.

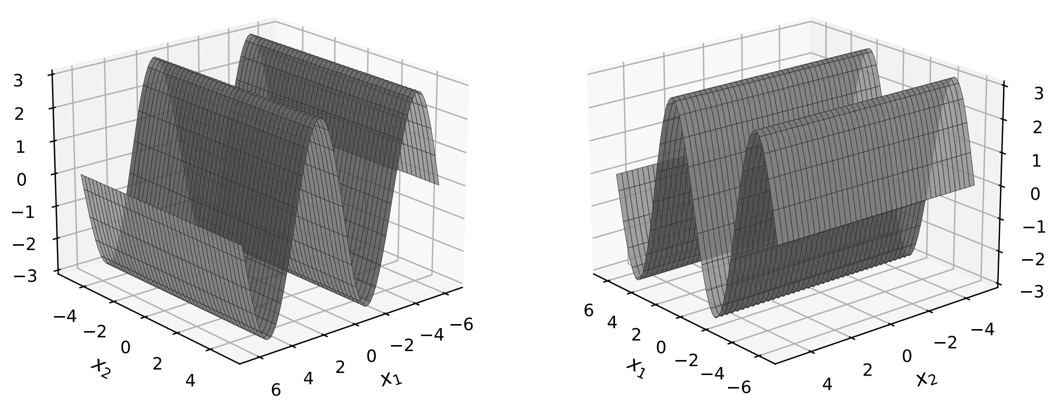
When your friend performs PCA a second time on this 75-dimensional data, the eigenvectors of a diagonal matrix are the standard basis vectors, and the top 50 correspond to the 50 largest eigenvalues $\lambda_1, \ldots, \lambda_{50}$. Selecting these is equivalent to keeping the first 50 coordinates of the 75-dimensional data, which are exactly the projections onto the original top 50 eigenvectors.

Therefore, both procedures yield the same 50-dimensional data set.

For this problem, it was important that we could assume that you and your friend always obtained the same eigenvectors. If there were sign ambiguities (or if there were eigenvalues with multiplicity greater than 1), then your friend could have obtained a different set of eigenvectors in the second PCA step, leading to a different final result.

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?

3

Solution: 3.

The ambient dimension is the dimension of the space in which the surface lives. Here, the surface has three coordinates (x_1, x_2, and x_3), so the ambient dimension is 3.

b) What is the intrinsic dimension?

2

Solution: 2.

The intrinsic dimension is the number of parameters needed to describe a point's location on the manifold. This shape is a 2D sine wave: a flat 2D sheet that has been bent into a wave in 3D space. If you were to "flatten" it, it would become a flat rectangle, which is 2-dimensional. Therefore, the intrinsic dimension is 2.

This was similar to Practice Problem 92.