

1. Select the option that best completes the following sentence:

1 / 1 point

For data with many features, principal components analysis

- ☐ identifies which features can be safely discarded
- ☐ reduces the number of features without losing any information.
- ☐ establishes a minimum number of viable features for use in the analysis.
- ☒ generates new features that are linear combinations of the original features.

☒ Correct

Correct! You can find more information in the lesson on Dimensionality Reduction.

2. Which option correctly lists the steps for implementing PCA in Python?

1 / 1 point

1. Fit PCA to data
2. Scale the data
3. Determine the desired number of components based on total explained variance
4. Define a PCA object

- ☐ 4, 1, 3, 2
- ☐ 2, 1, 3, 4
- ☒ 2, 4, 1, 3
- ☐ 4, 1, 2, 3

☒ Correct

Correct! Note that we need to scale the data prior to fitting a PCA object and obtain the total explained variance afterwards based on the principal components built.

3. Given the following matrix for lengths of singular vectors, how do we rank the vectors in terms of importance?

1 / 1 point

$$\begin{bmatrix} 11 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$v_1, v_2, v_3, v_4$

- ☒  $v_1, v_2, v_3, v_4$
- ☐  $v_4, v_3, v_2, v_1$
- ☐  $v_1, v_4, v_3, v_2$
- ☐  $v_2, v_3, v_4, v_1$

☒ Correct

Correct! The bigger the eigenvalue (value on the diagonal), the more important it is.

4. Given two principal components  $v_1, v_2$ , let's say that feature  $f_1$  contributed 0.15 to  $v_1$  and 0.25 to  $v_2$ . Feature  $f_2$  contributed -0.11 to  $v_1$  and 0.4 to  $v_2$ .

1 / 1 point

Which feature is more important according to their total contribution to the components?

- ☐  $v_1$  because  $0.15 + 0.25 > -0.11 + 0.4$
- ☐ Neither
- ☐  $v_2$  because  $-0.11 + 0.4 < 0.15 + 0.25$
- ☒  $v_2$  because  $|-0.11| + |0.4| > |0.15| + |0.25|$

✓ Correct  
Correct!

5. (True/False) In PCA, the first principal component represents the most important feature in the dataset.

1 / 1 point

- ☐ True
- ☒ False

✓ Correct  
Correct! Each principal component in PCA is a linear combination of features in the dataset, so the first one doesn't necessarily correspond to the single most important original feature.