

Ungraded: Dimensionality Reduction

Ungraded: Dimensionality Reduction
Practice Assignment • 10 min

Exit

Your grade: 100%
Your latest: 100% • Your highest: 100% • To pass you need at least 66%. We keep your highest score.

Next item →

1. What is the purpose of dimensionality reduction in enterprise datasets? 1 / 1 point

☐ To predict the target with the best accuracy.

☒ To improve model performance by reducing the number of features used.

Correct! This is accomplished by either selecting a subset of the original features or by creating new features from them.

☐ To create clusters for grouping data points.

☐ To improve model performance by providing a ranking of the features and maximizing the features used.

2. (True/False) Principal Component Analysis reduces dimensions by identifying features that can be excluded. 1 / 1 point

☐ True

☒ False

Correct! Instead, PCA creates new features that are linear combinations of the original ones.

3. Let's say that PCA found two principal components v_1 and v_2 . v_1 accounts for 0.5 of the total amount of variance in our dataset and v_2 accounts for 0.24. Which one is more important and why? 1 / 1 point

☒ v_1 because we will be able to maintain more of the original variance in the dataset.

☐ v_2 because it reduces 50% of the total variance in the dataset.

Correct! We are able to retain more information about the original dataset by projecting

Graded: Module 4 Quiz

Your grade: 100%

Your latest: 100% • Your highest: 100% • To pass you need at least 80%. We keep your highest score.

Next item →

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

☒ 2, 4, 1, 3

☐ 2, 1, 3, 4

☐ 4, 1, 2, 3

☐ 4, 1, 3, 2

✔ 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_3 contributed -0.11 to v_1 and 0.4 to v_2 .

1 / 1 point

Which feature is more important according to their total contr

- ☐ Neither
- ☐ v_1 because $0.15 + 0.25 > -0.11 + 0.4$
- ☐ 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.

Lưu ý rằng chúng ta cần chia tỷ lệ dữ liệu trước khi đưa chính để tương PCA và thu được tổng phương sai được giải thích sau đó dựa trên các thành phần chính được xây dựng.