



Question 1

Imagine that we apply `X = PCA(n_components=5).fit_transform(data)` and data has shape (5000, 53). What is the shape of X?

Correct answers:

- `(5000, 5)`. Yes, it should be (n_samples, n_components).

Incorrect answers:

- `(5, 53)`. No, we cannot transform 5000 samples into 5.
- `(5, 5000)`. No, we cannot transform 5000 samples into 5.
- `(53, 5)`. No, we cannot transform 5000 samples into 53.

Question 2

To which data NMF is NOT applicable?

Correct answers:

- Standardized matrix. "Standardized" means that every feature column has zero mean and unit variance. This implies that we have negative values and cannot apply NMF.

Incorrect answers:

- Bag-of-words matrix. Since BoW matrix is non-negative matrix, we can apply NMF to it.
- One-Hot encoded feature. Since this matrix contains only 0's and 1's -- it is non-negative and we can apply NMF.

Question 3

Suppose we have 2 categorical features: **f1** with A possible values and **f2** with B possible values. How many values will their interaction have?

Correct answers:

- Less or equal to A * B. True. Sometimes some value (e.g. *a*) from A cannot be used with some value (eg *b*) from B. In this case, we have no change to see *ab* combination. If all value from A can be used with all values from B -- we will get A*B new possible values.

Incorrect answers:

- Exactly A + B. No, it's too small.
- Exactly A * B. Not exactly, sometimes some value (e.g. *a*) from A cannot be used with some value (eg *b*) from B. In this case, we have no change to see *ab* combination.
- max(A, B). No, it's way too small.

Question 4

Imagine we have 2 categorical features represented as integers: **f1** with all values in range [0, 1000] and **f2** with values in range [0, 100]. What is the correct way to build their interaction?

Correct answers:

- `f1.astype(str) + " " + f2.astype(str)`. Yes, this is the right answer

Incorrect answers:

- `f1 + f2`. There are some problems. For example, if $f1+f2=100$: it is an interaction of 0 and 100, 100 and 0, or 90 and 10?
- `f1.astype(str) + f2.astype(str)`. There is some problems. For example, "123": it is an d interaction of "1" and "23" or "12" and "3"?
- `(f1 + f2).astype(str)`. It is essentially the same way as just $f1+f2$.

Question 5

What is a correct way to get t-SNE projection of train and test data?

Correct answers:

- Apply t-SNE to concatenation of train and test and split projection back. This is the right way since train and test will be projected in the same way.

Incorrect answers:

- Apply t-SNE to the train and after that to the test. In this case we will have 2 different projections.

- Apply t-SNE to the test first and after to train. This this case we will have 2 different projections.
- Doesn't matter, all variants will produce the same result. No, since tSNE results in randomized projection

Question 6

Is it possible to do t-SNE projection into 20-dimensional space?

Correct answers:

- Yes, why not. You can do tSNE projection into arbitrary space.

Incorrect answers:

- No, only 2-dim or 3-dim projections are possible. Wrong! Despite tSNE is quite often used for visualization purposes, you don't limit to use only 2- or 3-dimensional projections.

Mark as completed

