Вопросы по PCA

1. PCA используется для:

Снижения размерности данных при сохранении максимально возможной вариации.

2. PCA находит:

Ортогональные направления (главные компоненты), вдоль которых данные имеют наибольшую дисперсию.

3. Результатом работы PCA являются:

Главные компоненты (направления векторного пространства) и преобразованные данные в новом базисе.

4. Главные компоненты являются:

Линейными комбинациями исходных признаков.

5. Параметр n\_components в PCA определяет:

Количество главных компонент, которые нужно оставить.

6. Параметр svd\_solver в PCA определяет:

Алгоритм для выполнения сингулярного разложения (например, «auto», «full», «arpack», «randomized»).

7. explained\_variance\_ratio\_ в PCA показывает:

Долю дисперсии, объясненную каждой главной компонентой.

8. Перед применением PCA рекомендуется:

Стандартизировать данные, чтобы все признаки имели одинаковый масштаб.

Дополнительные вопросы по методу главных компонент (PCA)

1. Основная цель PCA:

Упростить данные путем снижения размерности с минимальной потерей информации.

2. PCA ищет:

Направления наибольшей вариации данных в многомерном пространстве.

3. Результатом PCA являются:

Матрица преобразованных данных, собственные значения и собственные векторы.

4. Главные компоненты:

Ортогональны друг другу и упорядочены по убыванию дисперсии.

5. n\_components в PCA определяет:

Число главных компонент или долю сохраненной дисперсии (если задано как число от 0 до 1).

6. explained\_variance\_ratio\_ в PCA показывает:

Какую часть общей дисперсии данных объясняет каждая главная компонента.

7. Перед применением PCA рекомендуется:

Привести данные к нулевому среднему и единичной дисперсии.

8. PCA чувствителен к масштабу признаков?

Да, поэтому стандартизация данных обязательна.

9. Какой метод чаще всего используется для вычисления главных компонент в PCA?

Сингулярное разложение (SVD).

10. Что такое собственные значения в контексте PCA?

Меры дисперсии, объясняемой каждой главной компонентой.

11. components\_ в PCA содержит:

Коэффициенты линейных комбинаций признаков для главных компонент.

12. mean\_ в PCA содержит:

Средние значения для каждого признака в данных.

13. PCA может использоваться для:

Визуализации данных, уменьшения размерности, устранения многоколлинеарности.

14. Если n\_components задан как число от 0 до 1, то это:

Доля дисперсии, которую необходимо сохранить.

15. whiten=True в PCA:

Преобразует данные так, чтобы каждая главная компонента имела единичную дисперсию.

16. PCA является:

Линейным методом снижения размерности.

17. inverse\_transform в PCA:

Восстанавливает данные из пространства главных компонент в исходное пространство.

18. Какой параметр отвечает за случайность в PCA?

Параметр random\_state при использовании randomized SVD.

19. svd\_solver=«arpack» рекомендуется для:

Малых объемов данных и низкого числа компонент.

20. PCA гарантирует:

Максимальную сохранность дисперсии данных в первых компонентах.

Вопросы по Kernel PCA

1. Какая из следующих функций ядра НЕ является допустимой для Kernel PCA?

Любая, которая не является положительно определенной (например, несимметричные функции).

2. Что из перечисленного является целью Kernel PCA?

Нелинейное снижение размерности путем преобразования данных в более высокомерное пространство.

3. Какое из следующих утверждений о Kernel PCA является верным?

Kernel PCA использует ядра для учета нелинейных зависимостей в данных.

4. Какая из следующих матриц используется для вычисления собственных значений и собственных векторов в Kernel PCA?

Матрица ядра (kernel matrix).

5. Что из перечисленного является преимуществом Kernel PCA?

Способность выявлять сложные, нелинейные структуры данных.