Тестовые вопросы по Sparse PCA

1. Какой параметр в SparsePCA используется для управления степенью разреженности? - **"alpha"** или **"regularization parameter"**. Он определяет уровень регуляризации, который применяется для достижения разреженных компонент.

2. Какой метод в SparsePCA используется для вычисления компонент? - метод **"Principal Component Analysis" (PCA)** с добавлением регуляризации, чтобы обеспечить разреженность компонент. Это может быть реализовано через методы, такие как **Lasso** или другие подходы к регуляризации.

3. Какой атрибут в SparsePCA содержит коэффициенты разреженных компонент? - **"components\_"**. Этот атрибут хранит значения весов для каждой из компонент.

4. Какой метод в SparsePCA используется для восстановления исходных данных из спроецированных данных? - метод **"inverse transform"**, который позволяет преобразовать спроецированные данные обратно в исходное пространство.

5. Что означает "разреженность" в контексте Sparse PCA? - "Разреженность" в контексте Sparse PCA означает наличие большого количества нулевых или близких к нулю значений в коэффициентах компонент. Это позволяет выделить только наиболее значимые признаки и игнорировать менее важные, что делает модель более интерпретируемой и устойчивой к шуму.

6. Какой из следующих методов регуляризации чаще всего используется в Sparse PCA? - метод регуляризации **L1**, который способствует созданию разреженных решений за счет добавления штрафа за величину коэффициентов.

7. Как влияет увеличение параметра регуляризации (например, alpha в Scikit-Learn) на разреженность компонент в Sparse PCA? - Увеличение параметра регуляризации (alpha) приводит к увеличению разреженности компонент, так как более высокий уровень регуляризации заставляет модель устанавливать больше коэффициентов равными нулю, что уменьшает количество активных признаков.

8. В каких случаях Sparse PCA может быть предпочтительнее стандартного PCA? - в случаях, когда необходимо выделить лишь несколько значимых признаков из большого набора данных, особенно если данные содержат много шумов или нерелевантных признаков.

9. Какой из следующих недостатков характерен для Sparse PCA? - Один из недостатков Sparse PCA заключается в том, что он может быть более чувствителен к выбору параметров регуляризации и может требовать больше вычислительных ресурсов по сравнению со стандартным PCA, особенно при работе с большими наборами данных.

Тестовые вопросы по "Факторный анализ"

1. Какова основная цель факторного анализа? - Основная цель факторного анализа заключается в уменьшении размерности данных и выявлении скрытых факторов, которые объясняют взаимосвязи между наблюдаемыми переменными. Это позволяет упростить данные и выявить структуры, которые могут быть неочевидны при прямом анализе.

2. Что такое "фактор" в факторном анализе? - "Фактор" в факторном анализе представляет собой скрытую переменную, которая объясняет корреляции между наблюдаемыми переменными. Факторы помогают понять, какие группы переменных имеют схожие характеристики и как они взаимодействуют друг с другом.

3. Какой тип данных подходит для факторного анализа? - Для факторного анализа подходят данные, которые являются количественными (например, интервальные или отношения). Однако также возможно применение методов для категориальных данных после их соответствующей кодировки.

4. Что такое "факторная нагрузка"? - Факторная нагрузка — это коэффициент, который показывает степень влияния конкретной переменной на определенный фактор. Она указывает на то, насколько сильно переменная связана с фактором, и помогает интерпретировать значение фактора.

5. Какой метод вращения факторов используется для упрощения интерпретации факторов? - Для упрощения интерпретации факторов часто используется метод **оборачивания (rotation)**, такой как **Varimax**, который максимизирует количество высоких нагрузок и минимизирует низкие нагрузки на факторы, делая их более интерпретируемыми.

6. Что такое "собственное значение" в факторном анализе? - "Собственное значение" (eigenvalue) в факторном анализе отражает количество вариации в данных, объясняемой конкретным фактором. Чем выше собственное значение, тем больше информации этот фактор содержит.

7. Какое правило используется для определения количества факторов, которые следует оставить? - Одно из распространенных правил — оставить те факторы, у которых собственное значение больше 1 (правило Кайзера). Также можно использовать график "локтя" для визуального определения точки, где добавление новых факторов не дает значительного увеличения объясненной вариации.

8. Что такое "общность" (communality) в факторном анализе? - Общность — это доля вариации каждой переменной, которая объясняется факторами в модели. Она показывает, насколько хорошо факторы объясняют каждую переменную; высокие значения общности указывают на то, что переменная хорошо представлена факторами.

9. Какой из следующих методов НЕ является методом извлечения факторов? - Методы извлечения факторов включают методы главных компонент (PCA), максимального правдоподобия и других подходов. Метод **регрессии** не является методом извлечения факторов.

10. Факторный анализ является \_\_\_ методом. - Факторный анализ является **методом уменьшения размерности** и **методом многомерной статистики**, который используется для анализа взаимосвязей между переменными и выявления скрытых структур в данных.

Тестовые вопросы по sklearn.decomposition.FactorAnalysis

1. Какой метод используется в ***sklearn.decomposition.FactorAnalysis*** для оценки параметров модели? - В sklearn.decomposition.FactorAnalysis используется метод **максимального правдоподобия** для оценки параметров модели.

2. Какой параметр в ***FactorAnalysis*** отвечает за количество факторов, которые необходимо извлечь? - Параметр, отвечающий за количество факторов, которые необходимо извлечь, называется **n\_components**.

3. Какой параметр в ***FactorAnalysis*** управляет типом вращения факторов? - Параметр, который управляет типом вращения факторов, называется **rotation**. Он может принимать значения, такие как 'varimax', 'promax' и другие.

4. Какой атрибут обученного объекта ***FactorAnalysis*** содержит матрицу факторных нагрузок? - Атрибут, который содержит матрицу факторных нагрузок в обученном объекте FactorAnalysis, называется **components\_**.

5. Какой метод в ***FactorAnalysis*** используется для трансформации новых данных с использованием обученной модели? - Метод, используемый для трансформации новых данных с использованием обученной модели, называется **transform()**.

6. Какой параметр в ***FactorAnalysis*** отвечает за инициализацию факторных нагрузок? - Параметр, отвечающий за инициализацию факторных нагрузок, называется **init**. Он определяет метод инициализации (например, 'random' или 'pca').

7. Какой параметр в ***FactorAnalysis*** позволяет контролировать максимальное количество итераций при оценке параметров? - Параметр, который позволяет контролировать максимальное количество итераций при оценке параметров, называется **max\_iter**.

8. Какой атрибут обученного объекта ***FactorAnalysis*** содержит ковариационную матрицу шума? - Атрибут, который содержит ковариационную матрицу шума в обученном объекте FactorAnalysis, называется **noise\_variance\_**.

9. Какой параметр в ***FactorAnalysis*** отвечает за допуск сходимости алгоритма? - Параметр, который отвечает за допуск сходимости алгоритма, называется **tol** (tolerance).

10. Какой метод в ***FactorAnalysis*** используется для вычисления логарифмической

вероятности данных? - Метод, используемый для вычисления логарифмической вероятности данных в FactorAnalysis, называется **score()**.