

Вопросы ко второй части курса по «Многомерному анализу данных» в сокращенном виде.

1. Факторный анализ. Модель. Параметры в факторном анализе. Общность и уникальность.
2. Что минимизируется в факторном анализе (отличие от анализа главных компонент).
3. Вращение в факторном анализе. Для чего нужны, пример. Метод varimax.
4. Линейный дискриминантный анализ. Модель. Классифицирующие функции. Значимость (какая гипотеза проверяется).
5. Почему линейный дискриминантный анализ называется линейным, а квадратичный – квадратичным?
6. Общий подход к классификации через апостериорные вероятности.
7. Какая ошибка минимизируется в подходе через максимизацию апостериорных вероятностей? Чему соответствует доля неправильных классификаций в матрице классификации?
8. Две группы, граница между двумя классами. Что происходит с границей при изменении априорных вероятностей?
9. Как проверяют качество построенной классифицирующей процедуры (cross-validation)?
10. Что такое ROC-кривая и AUC, для чего используются?
11. Разные модели для классов, число параметров в моделях и возможный overfitting.
12. Кластерный анализ, пример model-based подхода.
13. Кластерный анализ (partitioning): k-means (целевая функция, алгоритм, свойства, какие предположения о кластерах).
14. Кластерный анализ иерархический. Расстояния между точками и между кластерами. Разница между complete и single linkage.