Контрольные вопросы

Введение

1. Дать определение машинному обучению.
2. Дать определение Data Mining.
3. Какие три компоненты включает в себя машинное обучение?
4. Какая цель машинного обучения?
5. Что такое пространство признаков?
6. Перечислить этапы машинного обучения.
7. Перечислить задачи машинного обучения.
8. Что включает в себя процесс решения задачи машинного обучения?
9. Какие категории включает в себя классическое обучение?
10. Перечислить популярные алгоритмы классификации.
11. Для чего используется регрессия?
12. При решении каких задач используется обучение без учителя?
13. Что означает задача кластеризации?
14. Какие цели преследует кластеризация?
15. В чем состоит задача сокращения размерности данных?
16. Перечислить популярные алгоритмы сокращения размерности данных.
17. В чем заключается метод ассоциаций?
18. Дать определение метода обучения с подкреплением.
19. Что такое ансамбль?
20. В чем состоит процедура стекинга?
21. В чем состоит процедура беггинга??
22. В чем состоит процедура бустинга?
23. Дать определение нейронной сети.

Глава 1. Числовые атрибуты

1. В чем смысл однофакторного анализа?
2. Описать структуру эмпирическое кумулятивной функции распределения.
3. Как определяется функция обратного кумулятивного распределения?
4. Что описываю меры центральной тенденции?
5. Что означает устойчивость данных?
6. Дать определение квартилю.
7. В чем смысл двухфакторного анализа?
8. Что означает ковариантность и корреляция между атрибутами?
9. Как задается общая и обобщенная дисперсия?
10. Что означает нормализация диапазона?
11. Какие параметры полностью описывают нормальное распределение?
12. Как определяется вероятностная масса нормального распределения?
13. Как определяется расстояние, называемое расстоянием Махаланобиса?

Глава 2. Нечеткая алгебра

1. Каким аксиомам удовлетворяет булева алгебра?
2. Дать определение инвертора.
3. Дать определение *t-*нормы.
4. Привести пример *t-*нормы.
5. Дать определение *s*-нормы.
6. Дать определение импликатора.
7. Привести пример имплькатора.
8. Дать определение нечеткому высказыванию.
9. Дать определение четкому множеству.
10. От чего зависит выбор функции принадлежности?
11. Привести примеры функции принадлежности.
12. С помощью какого подхода определяются операции над нечеткими отношениями?
13. Дать определение нечеткому числу.
14. Дать определение носителю нечеткого множества.
15. В чем заключается принцип расширения?

Глава 2. Ядерные методы

1. В чем состоит основная идея использования ядерных методов?
2. Когда функция скалярного произведения является положительно полуопределённым ядром?
3. Какое отображение осуществляет отображение воспроизводящего ядра?
4. Какие ядра называются векторными?
5. Какие полиномиальные ядра называются однородными и неоднородными?
6. Какая функция является ядром нормального распределения?
7. Перечислить основные операции ядра в пространстве признаков.
8. Дать определение нормы точки.
9. Как определяется среднее в пространстве признаков?
10. Как осуществляется центрирование в пространстве признаков?
11. Как осуществляется нормализация в пространстве признаков?
12. Пути какой длины образуют положительно определенные ядра?

Глава 3. Снижение размерности

1. Перечислить основные методы снижения размерности.
2. Дать определение методу главных компонент.
3. На какие матрицы разлагается ковариационная матрица с помощью метода главных компонент?
4. В чем смысл критерия среднеквадратичной ошибки?
5. В чем заключается ядерный анализ главных компонент?
6. Как соотносятся метод главных компонент и сингулярного разложения?
7. На какие матрицы разлагается ковариационная матрица с помощью метода сингулярного разложения?
8. Как определяется матрица Фробениуса?

Глава 4. Поиск наборов объектов

1. Что включает в себя кортеж транзакции?
2. Дать определение двоичной базы данных.
3. Дать определение поддержки набора в базе данных.
4. Дать определение относительной поддержки правила.
5. Перечислить наиболее распространенные алгоритмы майнинга набора данных?
6. Что понимается пол вычислительной сложностью алгоритма?
7. Перечислить основные способы улучшения brute-force алгоритма.
8. В чем состоит основная идея поуровневого подхода?

Глава 5. Суммирование наборов

1. Какое множество называется максимальным?
2. Каким свойствам удовлетворяет оператор замыкания?
3. Какие проверки надо осуществить при поиске максимальных наборов?
4. Какой набор данных называется невыводимым?
5. В чем заключается принцип включения-исключения?
6. Какой набор данных называется выводимым?
7. Дать определение обобщенному набору объектов.

Глава 6. Анализ последовательностей

1. Дать определение последовательности.
2. Как определяется длина последовательности?
3. Какая частая последовательность называется максимальной?
4. Какая частая последовательность называется максимальной?
5. Описать сущность метода уровневого поиска GSP?
6. Описать сущность метода вертикального поиска последовательности Spade?
7. Описать сущность метода поиска последовательностей на основе проекций: PrefixSpan:?
8. Что представляет из себя суффиксное дерево?
9. В чем заключается метод Укконена?

Глава 7. Оценка паттернов и правил

1. Для чего проводится оценка паттернов и правил?
2. Дать определение лифта.
3. Что измеряет усиление?
4. Что измеряет коэффициент Жаккара?
5. Что измеряет коэффициент убежденности?
6. Перечислить показатели оценки паттернов.
7. Какой набор элементов является продуктивным?
8. Дать определение неизбыточного правила.
9. В чем заключается сущность точного теста Фишера.
10. Что позволяет определить рандомизация?

Глава 8. Репрезентативная кластеризация

1. В чем сущность алгоритма К-средних?
2. В чем сущность алгоритма К-средних++?
3. Дать точное описание алгоритма К-средних++?
4. Какие недостатки имеет метод К-средних?
5. В чем сущность алгоритма Х-средних?
6. Дать определение нечеткой кластеризации С-средних?
7. Описать алгоритм С-средних?
8. Дать описание ядерного алгоритма К-средних.
9. Дать описание кластеризации с максимизацией ожиданий.
10. Дать описание гауссовской смеси?

Глава 9. Иерархическая кластеризация

1. В чем состоит цель иерархической кластеризации?
2. Чему соответствует количество иерархических кластеризаций?
3. Пояснить суть агломеративной иерархической кластеризации?
4. Какая мера используется для определения расстояния между двумя кластерами?
5. Пояснить в чем состоит суть метода полной связи.
6. Пояснить в чем состоит суть метода средней связи.
7. Пояснить в чем состоит суть цендроидного метода.
8. Дать определение алгоритма BIRCH.

Глава 10. Кластеризация, основанная на плотности

1. Пояснить суть кластеризации, основанной на плотности.
2. Что является целью оценки плотности?
3. Пояснить суть алгоритма DBSCAN.
4. Пояснить суть алгоритма OPTICS.
5. Какие ограничения имеет алгоритм DBSCAN?
6. Дать определение аттрактора плотности.
7. Дать определение центрального аттрактора.
8. Пояснить сущность алгоритма распространения сходства.
9. Пояснить сущность алгоритма кластеризации на основе сдвига среднего значения.
10. Пояснить сущность спектрального метода кластеризации.

Глава 11. Валидация кластеризации

1. Какие задачи включает в себя проверка и оценка кластера?
2. На какие типы делятся критерии оценки достоверности и статистики?
3. Описать сущность внешней меры достоверной кластеризации.
4. Дать определение частоты кластеризации.
5. Дать определение меры максимального соответствия.
6. Как определяется энтропия кластеризации?
7. Что определяет взаимная информация?
8. На чем основан критерий вариации информации?
9. Какую информацию используют парные меры?
10. Что измеряет статистика Рэнда?
11. Что определяет мера Фаулкса-Мальлоу?
12. Как определяется статистика Гильберта?
13. Пояснить сущность внутренних мер?
14. Как определяется индекса Дэнна?
15. В чем состоит сущность метода статистики разрыва?
16. В чем заключается основная идея стабильности кластера?
17. Дать определение статистики Хопкинса?

Глава 12. Бикластеризация

1. Пояснить, что понимается под бикластером.
2. Что используется для визуализации бикластеризации?
3. Перечислить алгоритмы бикластеризации.
4. Какой бикластер называется идеальным?
5. На чем основан алгоритм удаления узлов?
6. На чем основан алгоритм FLOC?
7. Пояснить сущность бикластеризации на основе спектрального разбиения.
8. Пояснить смысл последовательной бикластеризация с помощью 0-1 программирования.
9. Пояснить сущность бикластеризация посредством гиббсовского сэмплирования.
10. Что такое гиббсовское сэмплирование?
11. Пояснить сущность статистико-алгоритмического метода бикластерного анализа?
12. В чем сущность классификации с учителем посредством максимальных δ-валидных паттернов?
13. Где применяется бикластеризация?

Глава 13. Трикластеризация

1. Когда, кем и для чего был метод трикластеризации?
2. Из чего состоит триадический контекст?
3. Пояснить сущность метода TRIAS.
4. Перечислить основные достоинства алгоритма TRIAS.
5. Какой трикластер называется полным?
6. Каким полезным свойством обладает определение OAC-трикластера?
7. На чем основан метод спектральной трикластеризации?
8. Перечислить критерии качества для оценки трикластеров.
9. Как определяется мощность трикластера?
10. Как определяется покрытие трикластера?
11. Дать определение устойчивости алгоритма к шуму.