

Теоретические вопросы зачёта по машинному обучению

1. Базовые общие понятия машинного обучения и искусственного интеллекта: в чём их разница? Назовите основные виды машинного обучения.
2. Какова общая последовательность действий при создании системы машинного обучения (решателя)? В чём назначение каждой из этих операций?
3. Машинное обучение с учителем (supervised learning): чем оно отличается от других видов машинного обучения и что специфическое для него необходимо? Какими средствами это достигается?
4. Методы решения задач обучения с подкреплением (reinforcement learning). Опишите задачи, решаемые методами обучения с подкреплением.
5. Как готовят числовые и категориальные данные для использования в системах машинного обучения? Перечислите операции и их назначение. В чём принципиальная разница предобработки естественным образом числовых и категориальных данных?
6. К какому типу машинного обучения относится регрессия? Что такое регрессор? Какими свойствами хороший регрессор должен быть наделён? По каким параметрам судят о качестве регрессионной модели?
7. Опишите метод наименьших квадратов для линейной регрессии. Опишите и сравните методы ridge regression, LASSO и эластичной сети (elastic net). Для чего в них служит регуляризация?
8. Что такое переобучение при регрессии? В чём оно проявляется? Как с ним бороться?
9. Перечислите метрики качества моделей бинарной классификации.
10. Задача снижения размерности. Метод главных компонент: идея и основные свойства.
11. Кластеризация. (Опишите методы агломеративной кластеризации, К-средних, DBSCAN). Как происходит оценка количества кластеров?

12. Принцип работы ансамблей решателей. Какие предъявляются требования к слабым решателям для возможности построения сильного решателя? Опишите принципы организации бэггинга (bootstrap aggregation).
13. Объясните, что такое простое голосование, взвешенное голосование, комитет экспертов (mixture of experts) при создании ансамбля. Стекинг (stacking, stacked generalization), его основная идея и назначение.
14. Бустинг (boosting). Основные идеи и принципы работы AdaBoost (adaptive boosting, адаптивный бустинг).
15. Бустинг (boosting). Основные идеи и принципы работы градиентного бустинга (gradient boost).
16. Балансирование наборов данных: проблема несбалансированных данных и методы её преодоления.
17. Деревья решений. Процесс построения. Регуляризация деревьев решений.
18. Что такое случайный лес (random forest)? Для чего в него введена случайность? В чём она состоит и как влияет на методы построения случайного леса?