

ПИШ 2023. Блок 2 «Машинное обучение». Вопросы аттестации

1. Объясните, как вычисляется качество модели с задачей классификации: назовите основные и вторичные метрики, приведите пример.
2. Что такое тестовый набор данных (test set) и для чего он нужен?
3. Каково назначение валидационного множества данных (validation set)?
4. В чем разница между параметром модели и гиперпараметром алгоритма обучения? Приведите примеры.
5. Что может пойти не так, если вы подгоните гиперпараметры модели, на тестовом наборе?
6. Что такое перекрестная проверка (cross-validation) и чем она лучше использования одного тестового множества?
7. Что такое онлайн-режим работы системы машинного обучения?
8. Опишите классификатор по методу ближайших соседей (KNN): в чем заключаются режимы его обучения и предсказания (вывода)?
9. Опишите модель линейной регрессии: в чем заключаются режимы ее обучения и предсказания (вывода)?
10. В чем заключается разница между обучением с учителем и обучением без привлечения учителя? Какие ещё есть «промежуточные» типы обучения.
11. Что такое переобучение?
12. Опишите кратко метод работы дерева решений и ансамбля таких деревьев в режиме взаимного усиления (мы знаем его как «бустинг»).