

Машинное обучение.

Домашнее задание №2

Задача 1. Ответьте на вопросы:

1. Что такое решающее дерево? Как по построенному дереву найти прогноз для объекта?
2. Зачем в вершинах нужны предикаты? Какие типы предикатов вы знаете? Приведите примеры.
3. Почему для любой выборки можно построить решающее дерево, имеющее нулевую ошибку на ней?
4. Почему не рекомендуется строить небинарные деревья (т.е. имеющие больше двух потомков у каждой вершины)?
5. Как устроен жадный алгоритм построения дерева? Какие у него параметры?
6. Зачем нужны критерии информативности? Чем объясняется их большое разнообразие для задач классификации?
7. Как задается критерий ошибки классификации? Индекс Джини? Энтропийный критерий? Какой у них смысл?
8. Как задается критерий информативности в задачах регрессии?
9. Какие критерии останова вы знаете?
10. Что такое стрижка дерева? В чем идея cost-complexity pruning?
11. Какие методы обработки пропущенных значений вы знаете?
12. Как можно учитывать категориальные признаки в решающем дереве?
13. Как можно свести задачу перебора всех разбиений категориального признака к задаче поиска оптимального разбиения для вещественного признака?