

Основные алгоритмы задание 1

Бунин Кирилл

March 27

Task 1

Какие слабые модели рекомендуется включать в бэггинг-ансамбль

Ответ: Модели, имеющие высокий bias и малый variance

Task 2

Как изменяется bias откликов дерева решений при увеличении числа решающих вершин?

Ответ: Уменьшается

Task 3

Значение логит-функции для некоторого примера оказалось равным нулю. Что можно сказать о вероятности принадлежности этого примера к позитивному классу?

Ответ: 0.5

Task 4

Какой вид имеет функция регрессии, формируемая классическим регрессионным деревом?

Ответ: Кусочно-постоянная функция

Task 5

Выберите верные утверждения о числе решающих и терминальных вершин дерева решений.

Ответ: Число решающих вершин всегда меньше числа терминальных вершин

Task 6

Возможно ли, чтобы граница между классами, формируемая классификационным деревом, была линейной?

Ответ: Нет

Task 7

Что происходит с variance KNN-регрессии при увеличении параметра K ?

Ответ: Уменьшается

Task 8

Чему равно отношение шансов для бинарного классификатора?

Ответ: Отношению априорных вероятностей

Task 9

Пусть x - точка, лежащая на границе между классами, формируемыми логистической регрессией. Чему равно отношение шансов для этой точки?

Ответ: 1

Task 10

Зачем нужно проводить обрезку дерева решений?

Ответ: Для повышения обобщающих способностей дерева

Task 11

Если для некоторого обучающего примера отношение вероятностей близко к 0, о чём это говорит?

Ответ: Вероятность близка к 0

Task 12

Если мы хотим отобрать признаки, нам стоит использовать L_1 или L_2 регуляризацию?

Ответ: L_1

Task 13

Предположим, что наша модель $a(x)$ всегда выдаёт константное значение. Тогда у этой модели будет...

Ответ: Низкий bias, высокий variance

Task 14

Мы обучаем дерево решений. При увеличении глубины дерева у модели...

Ответ: bias снижается, variance растёт

Task 15

Предположим, что мы обучили алгоритм Random Forest и Gradient Decision Trees. В каком ансамбле базовые алгоритмы, как правило, имеют большую глубину?

Ответ: Gradient Boosted Decision Trees

Task 16

Предположим, что мы обучили модель бинарной классификации. Какие ошибки хуже?

Ответ: нельзя дать точный ответ, зависит от задачи

Task 17

Изменится ли ROC AUC, если ко всем предсказаниям применить логарифм? // Ответ: Нет

Task 18

Что будет со значением ROC AUC, если все предсказанные значения разделить на 2?

Ответ: Не изменится

Task 19

Рассмотрим нормализацию признака x с помощью стандартизации. Признак x - разреженный.

Ответ: Нет

Task 20

Вам необходимо разбить всех существующих клиентов по этим категориям. Какая это задача?

Ответ: Частичное обучение (semi-supervised learning)

Task 21

Что быстрее отработает на большом датасете (100 Гб) K-Fold или Leave-One-Out (LOO) CV?

Ответ: K-Fold CV

Task 22

Какой алгоритм использует ленивое обучение?

Ответ: kNN - k ближайших соседей

Task 23

Ответ: 2/3

Task 24

Какой наиболее подходящий способ работы с категориальными признаками при построении модели линейной регрессии?

Ответ: Закодировать их целыми числами от 0 до $n - 1$, где n - количество категорий

Task 25

Какую задачу решает логистическая регрессия?

Ответ: Классификация