Гиперпараметры в контексте машинного обучения - это

a) параметры модели, которые оптимизируются в процессе обучения

b) параметры модели, которые задаются до начала процесса обучения и не изменяются в процессе обучения +

c) параметры модели, которые определяют количество эпох обучения

d) параметры модели, которые определяют структуру модели

Кросс-валидация в контексте качества моделей - это метод, при котором

a) обучающие данные разделяются на несколько частей, и каждая часть поочередно используется как тестовый набор данных +

b) обучающие данные дублируются для улучшения качества модели

c) тестовые данные используются для обучения модели

d) обучающие данные используются для тестирования модели

Переобучение - это ситуация, когда

a) модель слишком хорошо подстроилась под обучающие данные и плохо обобщает на новые данные +

b) модель недостаточно подстроилась под обучающие данные и плохо обобщает на новые данные

c) модель не способна обучаться на данных

d) модель не может адаптироваться к изменяющимся данным

MAE, MSE и R^2 используются для оценки

a) точности модели

b) полноты модели

c) качества предсказаний модели +

d) скорости обучения модели

Для оценки классификации используются

a) точность и полнота +

b) средняя абсолютная ошибка и средняя квадратичная ошибка

c) коэффициент детерминации R^2

d) средняя абсолютная ошибка и коэффициент детерминации R^2

Функция потерь - это функция, которая измеряет

a) точность предсказаний модели

b) ошибку между предсказанными значениями и реальными значениями +

c) скорость обучения модели

d) сложность модели

Методы обучения с учителем - это такие методы,

a) которые не требуют наличия меток в данных

b) которые требуют наличия меток в данных +

c) которые не требуют наличия обучающих данных

d) которые не требуют наличия тестовых данных

Линейная регрессия … использоваться для задач классификации

а) может +

b) не может

Недообучение возникает из-за

a) слишком хорошей подгонки модели под обучающие данные

b) недостаточной подгонки модели под обучающие данные +

c) отсутствия тестовых данных

d) недостаточного количества эпох обучения

Для минимизации функции потерь можно использовать

a) градиентный спуск +

b) метод опорных векторов (SVM)

c) случайный лес

d) метод k-ближайших соседей

Метрика полнота (recall) используется для оценивания возможности модели...

a) правильно классифицировать положительные примеры +

b) правильно классифицировать отрицательные примеры

c) правильно классифицировать все примеры

d) правильно классифицировать случайные примеры

Логистическая регрессия - это

a) метод для группировки данных

b) метод для предсказания вероятности принадлежности к классу +

c) метод для обучения без учителя

d) метод для измерения сходства данных

Полиномиальные модели используются для

a) группировки данных

b) предсказания вероятности принадлежности к классу

c) моделирования нелинейных зависимостей +

d) измерения сходства данных

Зазор (margin) в терминах SVM - это

a) расстояние между классами в методе k-ближайших соседей

b) расстояние между центроидами в кластеризации k-средних

c) расстояние от разделяющей гиперплоскости до ближайшего объекта обучающей выборки +

d) расстояние от центра класса до его ближайшего соседа

Линейно неразделимые данные - это такие данные, которые

a) можно разделить гиперплоскостью на два класса

b) нельзя разделить прямой линией на два класса +

c) имеют большой зазор между классами

d) имеют одинаковые признаки у всех объектов