**Вопросы к экзамену / зачету по машинному обучению**

**Билет № 1**

1. Линейная регрессия

Область применения

Описание метода

Гиперпараметры

Сравните линейную и логистическую регрессию

1. Конструирование признаков (Feature Engineering) и подготовка данных. Кросс-валидация, регуляризация, поиск по сетке

**Билет № 2**

1. Логистическая регрессия

Область применения

Описание метода. Уравнение лог функции. Шансы. Log odds

Гиперпараметры

Сравните линейную и логистическую регрессию

1. Метрики оценки модели классификации

Дайте определение машинному обучению и объясните его важность в современных технологиях.

**Билет № 3**

1. Метод опорных векторов

Область применения

Описание метода: определение гиперплоскости и формула, классификатор максимального зазора, классификатор опорных векторов, метод опорных векторов, kernel trick,

Гиперпараметры

Сравните и сопоставьте методы k-ближайших соседей (KNN) и опорных векторов (SVM).

1. Каковы основные различия между обучением с учителем и обучением без учителя?

Опишите этапы процесса выбора признаков и объясните их значение.

**Билет № 4**

1. Метод k-ближайших соседей

Область применения

Описание метода: суть метода, 4 метрики расстояния, выбор числа – k,

Гиперпараметры

Сравните и сопоставьте методы k-ближайших соседей (KNN) и опорных векторов (SVM).

1. Что такое переобучение в машинном обучении? Как его недопустить?

Какова цель регуляризации в машинном обучении? Приведите примеры часто используемых методов регуляризации.

**Билет № 5**

1. Деревья решений

Область применения

Описание метода: терминология, операции с деревьями, gini impurity.

Гиперпараметры

1. Сравните и сопоставьте методы деревья решений

и случайных лесов.

Объясните разницу между задачами классификации и регрессии в машинном обучении.

**Билет № 6**

1. Случайные леса

Область применения

Описание метода: решение задачи классификации, решение задачи регрессии, bootstrapping, bagging,

Гиперпараметры

1. Сравните и сопоставьте методы деревья решений

и случайных лесов.

Объясните разницу между задачами классификации и регрессии в машинном обучении.

**Билет № 7**

1. Бустинг - AdaBoost

Область применения

Описание метода: Формула бустинга, AdaBoost, Процесс AdaBoost,

Гиперпараметры

Сравните и сопоставьте методы градиентный бустинг

и AdaBoost.

1. Объясните концепцию обнаружения аномалий в данных и работу с ними.

Как бы вы поступили с отсутствующими данными в наборе данных?

**Билет № 8**

1. Бустинг - градиентный бустинг

Область применения

Описание метода: процесс градиентного бустинга.

Гиперпараметры

Сравните и сопоставьте методы градиентный бустинг

и AdaBoost.

1. Как бы вы поступили с отсутствующими данными в наборе данных?

Какими способами можно уменьшить размерность набора данных?

**Билет № 9**

1. Кластеризацию К-средних (K-Means Clustering)

Область применения

Описание метода

Сравнение K-Means и DBSCAN

Метод локтя

Метод силуэтов

Опишите процесс настройки гиперпараметров и его важность для оптимизации производительности для выбранной вами модели.

1. Конструирование признаков (Feature Engineering) и подготовка данных

**Билет № 10**

1. DBSCAN

Область применения

Описание метода

Метод локтя

Сравнение K-Means и DBSCAN

Опишите процесс настройки гиперпараметров и его важность для оптимизации производительности для выбранной вами модели.

1. Кросс-валидация, регуляризация, поиск по сетке

**Билет № 11**

1. Иерархическая кластеризация

Область применения

Описание метода

Сравнение Иерархической кластеризации и DBSCAN

Опишите процесс настройки гиперпараметров и его важность для оптимизации производительности для выбранной вами модели.

1. Метрики оценки модели классификации

**Билет № 12**

1. Метод главных компонент

Область применения

Описание метода

Сравнение Метода главных компонент и Иерархической кластеризации

Опишите процесс настройки гиперпараметров и его важность для оптимизации производительности для выбранной вами модели.

1. Дайте определение машинному обучению и объясните его важность в современных технологиях.

**2 вопрос:**

1. Конструирование признаков (Feature Engineering) и подготовка данных
2. Кросс-валидация, регуляризация, поиск по сетке
3. Метрики оценки модели классификации
4. Дайте определение машинному обучению и объясните его важность в современных технологиях.
5. Каковы основные различия между обучением с учителем и обучением без учителя?
6. Опишите этапы процесса выбора признаков и объясните их значение.
7. Сравните и сопоставьте методы k-ближайших соседей (KNN) и опорных векторов (SVM).
8. Что такое переобучение в машинном обучении? Как его недопустить?
9. Какова цель регуляризации в машинном обучении? Приведите примеры часто используемых методов регуляризации.
10. Объясните принципы работы деревьев решений и случайных лесов.
11. Объясните разницу между задачами классификации и регрессии в машинном обучении.
12. Опишите процесс настройки гиперпараметров и его важность для оптимизации производительности для выбранной вами модели.
13. Объясните концепцию обнаружения аномалий в данных и работу с ними.
14. Что такое случайный лес?
15. Объясните недостатки линейной модели?
16. Несколько маленьких деревьев решений лучше, чем одно большое? Обоснуйте.
17. На каких предположениях основана линейная регрессия?
18. Какие библиотеки Python обычно используются в машинном обучении?
19. Как бы вы поступили с отсутствующими данными в наборе данных?
20. Какими способами можно уменьшить размерность набора данных?