**Вопросы к зачету по** **курсу “Машинное обучение для решения прикладных задач с помощью Python”**

1. Метрические методы

- Что такое косинусная схожесть?

- Верно ли, что функция расстояния в метрических алгоритмах классификации и регрессии всегда удовлетворяет неравенству треугольника?

(доп. вопросы: что такое неравенство треугольника? Можете ли привести пример такой функции расстояния?)

- Манхэттенское расстояние между двумя точками.

- Идея алгоритма k-NN. В каких задачах применяется. Эффективность. Влияние значения k на время работы и и качество получаемого решения. Недостатки k-NN.

2. Линейные классификаторы

- Почему классификаторы называются линейными? Какое предположение о зависимости в данных делается в этом случае?

- Support Vector Machine, или метод опорных векторов.

- Какую задачу решает логистическая регрессия? Что возвращается в результате применения логистической регрессии?

- Расшифруйте TF-IDF (TF(t, d) - term frequency, IDF(t, D) - inverse document frequency, tf-idf(t, d, D) = tf(t, d) \* idf(t, D))

- Градиентный спуск. Градиент.

- Отличия SVM от Логистической регрессии

- Недостатки SVM

3. Линейная регрессия

- Какие метрики можно применять для оценки качества построенной линейной модели?

- Для чего еще, кроме решения самой задачи регрессии, довольно часто используют алгоритм построения линейной регрессии?

- Одномерный случай лин. регрессии. Сколько параметров у модели?

4. Решающие деревья

- Это алгоритм без учителя, с учителем?

- Недостатки решающих деревьев.

- Достоинства решающих деревьев.

5. Ансамбли

- Что такое ассамблирование нескольких моделей?

- Мотивация для использования ансамблей.

- Какими должны быть базовые алгоритмы в ансамбле?

- В чем отличие градиентного бустинга от случайного леса?

6. Метрики качества классификации

- Когда метрика accuracy может сбить с толку?

- Что такое F1 мера?

- PR-кривая.

- ROC-кривая.

- Многоклассовая классификация. One vs All, All vs All. Качество многоклассовой классификации.

7. Понижение размерности

- Проклятие размерности. Проблемы.

- SVD разложение. Применение.

- Что такое эмбеддинг?

- PCA, или метод главных компонент.

- t-SNE. Что это такое?

8. Кластеризация

- К какому типу задач относится кластеризация?

- Проблемы постановки задачи кластеризации?

- K-Means. Сходимость. Недостатки.

- Иерархическая кластеризация.

9. Рекомендательные системы

- Коллаборативная фильтрация.

- Knowledge-based и User-based подходы.

- Матрица User-Item, разреженность матрицы.

- ALS разложение

- SVD vs ALS разложение.

10. Временные ряды

- Что такое временной ряд?

- Компоненты временного ряда?

- Стационарный временной ряд?

- Автокорреляция.

- Приведите примеры алгоритмов прогнозирования временных рядов.

11. Нейронные сети

- Перцептрон Розенблатта

- Приведите пример функций активации. Почему они имеют такую форму?

- Многослойный перцептрон.

- XOR проблема и ее решение.

- Применение нейронных сетей.