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

**Теоретические вопросы**

1. Перечислите основные задачи машинного обучения.
2. Перечислите примеры прикладных задач, которые решаются с помощью алгоритмов машинного обучения.
3. Метрические методы классификации и регрессии. Метод ближайших соседей.
4. Линейные модели классификации (бинарная и многоклассовая). Логистическая регрессия и метод опорных векторов.
5. Линейная регрессия. Метод опорных векторов в задачах регрессии. Метод наименьших квадратов.
6. Решающие деревья. Критерии ветвления. Борьба с переобучением. Регуляризация и подрезка деревьев.
7. Ансамбли. Бэггинг. Голосование/усреднение. Разложение ошибки смещение (bias) и разброс (variance). Случайный лес.
8. Градиентный бустинг.
9. Оценивание качества алгоритмов. Матрица ошибок (confusion matrix). Кросс-валидация.
10. Матричные разложения, понижение размерности. Метод главных компонент. t-SNE.
11. Кластеризация. Алгоритмы кластеризации. Иерархическая кластеризация. Метод K-средних.
12. Рекомендательные системы. Коллаборативная фильтрация, модели со скрытыми переменными. Методы на основе матрицы оценок: item-based, user-based. Матричные разложения.
13. Временные ряды. Компоненты временного ряда (тренд, сезонность, цикличность). Декомпозиция временного ряда. Плавающее среднее, экспоненциальное сглаживание. ARIMA. Линейные модели.

**Примеры тестовых заданий**

1. Решающие деревья обладают следующими свойствами (выберете верные утверждения):
   1. Строят разделяющую гиперплоскость, не чувствительны к шуму, не переобучаются
   2. Плохо интерпретируются, чувствительны к шуму, иногда отказывают в классификации данных
   3. Легко обрабатывают пропуски данных, хорошо интерпретируются, исключают отказы классификации, обрабатывают разнотипные и неотмасштабированные данные
2. Выберете верные утверждения:
   1. Логистическая регрессия решает задачу регрессии, а линейная регрессия решает задачу классификации
   2. Линейная регрессия строит гиперплоскость, пользуясь методом наименьших квадратов, а логистическая регрессия решает задачу классификации
   3. И линейная, и логистическая регрессия решают задачу регрессии
3. Метод опорных векторов (Support Vectors Machine, SVM) (выберете верные утверждения):
   1. Строит оптимальную разделяющую гиперплоскость, максимизируя ширину полосы между двумя классами
   2. Аппроксимирует искомую функцию, обучаясь при помощи градиентных методов, решает задачу регрессии
   3. Проводит гиперповерхность, минимизируя сумму квадратов расстояний элементов обучающей выборки до этой гиперповерхности.
4. Градиентный бустинг - это (выберете верные утверждения):
   1. Усиление обобщающей способности модели посредством использования нескольких градиентных методов оптимизации
   2. Композиционный алгоритм (обычно над решающими деревьями), основанный на идее построения нового элемента композиции на каждой итерации таким образом, что следующее приближение вектора ответов алгоритма отличается от текущего на градиент ошибки
   3. Композиционный алгоритм над решающими деревьями, основанный на идее голосования классификаторов, проводимого особым способом, с приписыванием константных весов каждому классификатору
5. Метод K-Means - это (выберете верные утверждения) :
   1. Метрический метод классификации, основанный на усреднении расстояний до k ближайших соседей
   2. Метод кластеризации, основанный на применении EM-алгоритма
   3. Метод валидации модели, основанный на усреднении ошибки по к прогонам модели на тестовых данных
6. Перечислите основные недостатки стандартного градиентного спуска.