**1.Какая функция обычно используется в методе градиентного спуска для минимизации ошибки при реализации парной линейной регрессии?**

**1.Экспоненциальную функцию ошибки (exponential error).**

**2.Математическую функцию ошибки первого порядка (mean absolute error, MAE).**

**3.Квадратичную функцию ошибки (mean squared error, MSE). +**

**4.Функцию ошибки Хьюбера (Huber loss).**

**5.Показательную функцию ошибки (exponential loss).**

**2. Какую функцию используют в логистической регрессии для преобразования выхода в интервал (0, 1)?**

**1.Квадратичную функцию ошибки (mean squared error, MSE).**

**2.Сигмоидальную функцию (логистическую функцию). +**

**3.Экспоненциальную функцию ошибки (exponential error).**

**4.Линейную функцию активации.**

**5.Гиперболический тангенс.**

**3.** **Какие действия предпринимаются для визуализации данных в задаче классификации после генерации признаков и целевой переменной?**

1. **Применяется метод plot() для каждого признака.**
2. **Используется scatterplot() для отображения двух классов в зависимости от их признаков. +**
3. **Применяется метод histogram() для анализа распределения данных.**
4. **Используется lineplot() для визуализации линий регрессии.**
5. **Метод circule()**

**4. Какие признаки вводятся в модель при использовании полиномиальной регрессии для аппроксимации нелинейных зависимостей?**

1. **Новые признаки, вычисленные на основе разности существующих атрибутов.**
2. **Логарифмы существующих атрибутов.**
3. **Полиномы всех признаков в датасете.**
4. **Степени существующих атрибутов. +**
5. **Случайные значения, добавляемые к существующим атрибутам.**

**5.** **Что является целью модели SVM при работе с линейно разделимыми классами?**

1. **Минимизация ширины разделяющей полосы.**
2. **Поиск гиперплоскости с максимальной шириной разделяющей полосы (margin). +**
3. **Построение гиперплоскости, ближайшей ко всем объектам обучающего набора.**
4. **Создание гиперплоскости, минимально приближенной к обучающим данным.**

**6.** **Что такое недообучение в контексте проблемы переобучения?**

1. **Проблема выбора гипотезы, когда функция h недостаточно точно отражает тренд данных из-за слишком простой или слишком малой функции, неспособной хорошо обобщить предсказания на новые данные. +**
2. **Проблема выбора гипотезы, когда функция h слишком сложная и хорошо подгоняет обучающие данные, но плохо обобщает предсказания на новые данные.**
3. **Ситуация, когда количество признаков превышает количество обучающих данных, что делает модель менее точной.**
4. **Проблема отсутствия признаков в обучающих данных, в результате чего модель не может сделать точные предсказания.**
5. **Обучение модели только на части данных, игнорируя другую часть, что ведет к несбалансированным результатам.**

**7****. Что представляет собой метод Stratified k-Fold в контексте кросс-валидации?**

1. **Метод разделения выборки на несколько фолдов, учитывающий баланс классов в каждом фолде для сохранения соотношения классов, как и в исходном наборе данных. +**
2. **Метод, который рандомно перемешивает данные и разделяет их на k фолдов для обучения и тестирования.**
3. **Метод, который использует только один фолд для обучения модели, игнорируя остальные фолды.**
4. **Метод, который увеличивает количество фолдов до k+1 для более точной оценки модели.**
5. **Метод, использующий тестовый набор данных для обучения модели с целью более высокой точности предсказаний.**

**8. Что означает понятие "чистых данных" в контексте датасета?**

1. **Это данные, которые содержат информацию только о численных объектах.**
2. **Это данные, где каждый объект описан числовым значением, без каких-либо атрибутов.**
3. **Это данные, которые не содержат отсутствующих (пропущенных) значений и имеют внутреннюю согласованность. +**
4. **Это данные, где каждый объект характеризуется множеством атрибутов, включающих текстовые и числовые значения.**
5. **Это данные, которые содержат только информацию об объектах, не связанную с экспериментами или наблюдениями.**

**9. Что включает в себя оценка источников и объемов данных для задач машинного обучения?**

1. **Определение только открытых источников данных.**
2. **Определение объема данных, исключительно основанного на сложности задачи.**
3. **Определение репрезентативности выборки источников данных относительно генеральной совокупности. +**
4. **Установление исключительно пакетных источников данных для обучения моделей.**
5. **Процесс генерации данных без учета существующих источников.**

**10. Что включает в себя оценка объемов данных для эффективного обучения моделей машинного обучения?**

1. **Определение объема данных, основанного только на сложности задачи.**
2. **Установление репрезентативности выборки источников данных относительно генеральной совокупности. +**
3. **Определение только открытых источников данных для обучения.**
4. **Определение и использование только закрытых источников данных для обучения.**
5. **Процесс генерации данных без учета существующих источников.**

**11. Какие характеристики являются важными при определении источников и объемов данных для моделирования?**

1. **Использование только открытых источников данных для обучения.**
2. **Использование только закрытых источников данных для обучения.**
3. **Выбор источников данных, представляющих репрезентативную выборку генеральной совокупности. +**
4. **Пренебрежение объемом данных и сосредоточение только на сложности задачи.**
5. **Использование исключительно пакетных данных для обучения моделей.**

**12. Какая характеристика степени полиномиальной регрессии является источником увеличения её сложности и вычислительной затратности?**

1. **Уменьшение количества признаков в модели.**
2. **Увеличение числа наблюдений в датасете.**
3. **Уменьшение степени полинома, используемого в модели.**
4. **Увеличение степени полинома. +**
5. **Усредненные значения всех атрибутов датасета.**

**13.Что обозначает граница принятия решения в контексте логистической регрессии?**

1. **Линия, которая разделяет область, где y = 0 и где y = 2.**
2. **График функции ошибки.**
3. **Граница, где x = 0 и y = 1.**
4. **Линия, разделяющая области прогнозирования классов 0 и 1. +**
5. **Отрезок прямой, где производная функции равна нулю.**

**14. В каких случаях применение модели SVM с линейной разделяющей гиперплоскостью является предпочтительным?**

1. **Когда классы объектов перекрываются в признаковом пространстве.**
2. **Когда объекты обучающей выборки имеют сложную структуру.**
3. **Когда существует возможность провести гиперплоскость таким образом, чтобы она максимально разделяла классы. +**
4. **Только в случае, когда данные имеют нелинейную зависимость.**