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. 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.

3. Какие методы были реализованы в созданной модели логистической регрессии?

1. Определение функции для стохастического градиентного спуска.
2. Реализация подсчета градиента функции ошибки для логистической регрессии. +
3. Разработка методов для обработки изображений.
4. Программирование для вычисления среднего арифметического значений признаков.
5. Ответы 1 и 2

4. 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.

5. 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.

6. Как регуляризация решает проблему переобучения в моделях?

1. Путем увеличения параметров модели, чтобы сделать ее более сложной и способной уловить больше шума в данных.
2. Путем увеличения количества признаков и объектов обучающей выборки.
3. Путем уменьшения параметров модели, сохраняя все признаки и снижая высокую дисперсию или переобучение. +
4. Исключением полезных признаков из модели для снижения ее сложности.
5. Применением кросс-валидации для оценки точности модели.

7. Для чего может потребоваться использование метода Stratified k-Fold при кросс-валидации?

1. Чтобы уменьшить количество фолдов и ускорить процесс кросс-валидации.
2. Для обучения модели только на определенных признаках, исключая остальные из фолдов.
3. В случае, когда требуется сохранить пропорции классов в обучающем и тестовом наборах данных. +
4. Для увеличения количества фолдов и улучшения точности модели при кросс-валидации.
5. Для выбора оптимального количества признаков, которые вносят наибольший вклад в обучение модели.

8. Что из перечисленного характеризует "чистые данные"?

1. Использование датасета с данными, которые описывают различные типы объектов, но без численного представления.
2. Наличие отсутствующих (пропущенных) значений, которые могут быть проигнорированы при анализе данных.
3. Отсутствие внутренней согласованности и структурированности в датасете.
4. Использование данных, которые представлены только текстовой информацией, без числовых значений.
5. Отсутствие отсутствующих (пропущенных) значений и наличие внутренней согласованности в датасете. +

9. 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.

10. 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.

11. 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.

12. 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()

13. . 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.

14. 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.