**1. Опишите задачу машинного обучения. Дайте определение объекту, целевой переменной, признакам, модели, функционалу ошибки.**

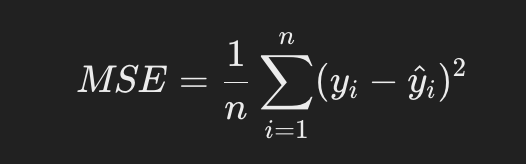
* **Задача машинного обучения**: обучение модели на основе данных для выполнения предсказаний или принятия решений без явного программирования.
* **Объект**: отдельная запись или наблюдение в наборе данных (например, человек, товар, изображение).
* **Целевая переменная**: переменная, которую модель пытается предсказать или классифицировать (например, цена дома, тип товара).
* **Признаки**: характеристики объекта, используемые для описания (например, возраст, вес, цвет).
* **Модель**: математическая функция, которая на основе признаков делает предсказания целевой переменной.
* **Функционал ошибки**: мера, показывающая, насколько хорошо модель предсказывает целевую переменную. Он показывает суммарную ошибку модели на наборе данных.

**2. Чем отличается функция потерь от функционала ошибки?**

* **Функция потерь**: измеряет ошибку для одного объекта.
* **Функционал ошибки**: агрегирует ошибки по всем объектам в наборе данных (суммирует или усредняет значения функции потерь

**3. Какие функции потерь используются при решении задачи регрессии (с формулами)?**

1. **MSE (Mean Squared Error, Среднеквадратическая ошибка)**:
   * MSE измеряет средний квадрат разницы между предсказанными значениями и истинными значениями .
   * **Формула**:



1. **MAE (Mean Absolute Error, Средняя абсолютная ошибка)**:
   * MAE измеряет среднюю абсолютную разницу между предсказанными и истинными значениями.
   * **Формула**:

​Изображение выглядит как текст, часы, Шрифт, Цифровые часы

Автоматически созданное описание

1. **Huber Loss (Губерова функция потерь)**:
   * Huber Loss сочетает в себе свойства MSE для малых ошибок и MAE для больших ошибок, что делает её устойчивой к выбросам.
   * **Формула**:

Изображение выглядит как текст, Шрифт, рукописный текст, черный

Автоматически созданное описание

1. **Log-cosh**:
   * Log-cosh — это гладкая функция потерь, которая ведёт себя как MSE для малых ошибок и как MAE для больших ошибок, но без резких переключений.
   * **Формула**:

Изображение выглядит как текст, рукописный текст, Шрифт, черный

Автоматически созданное описание

1. **Quantile Loss (Функция потерь квантилей)**:
   * Используется для предсказания квантилей распределения. Определяет, насколько предсказанное значение отклоняется от фактического значения, с учетом квантиля α\alphaα.
   * **Формула**:

Изображение выглядит как текст, Шрифт, рукописный текст, черный

Автоматически созданное описание

**4. Запишите формулу для линейной модели регрессии.**

Линейная модель регрессии:

Изображение выглядит как текст, снимок экрана, Шрифт, алгебра

Автоматически созданное описание

**5. Чем отличаются функционалы MSE и MAE? В каких случаях лучше использовать MSE, а в каких MAE?**

* **MSE**: измеряет среднее квадрата отклонений, чувствительна к выбросам, используется, когда важны крупные ошибки.
* **MAE**: измеряет среднее абсолютных отклонений, менее чувствительна к выбросам, используется, когда важна устойчивость к выбросам.

**6. Чем отличается MAE от MAPE? Что более понятно заказчику продукта?**

* **MAE**: измеряет абсолютную ошибку в тех же единицах, что и данные.
* **MAPE**: измеряет ошибку в процентах, что делает её понятной для заказчика.
* **Понятность заказчику**: MAPE более понятна, так как представляет ошибку в относительных величинах (процентах).
* **Коэффициент детерминации**, также известный как R2R^2R2 (R-квадрат), — это показатель качества модели регрессии, который измеряет, какую долю вариации зависимой переменной (например, yyy) можно объяснить независимыми переменными (предикторами) модели.
* **Формула:**

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

**R2=1: модель идеально предсказывает.**

**R2=0: модель не объясняет вариацию.**

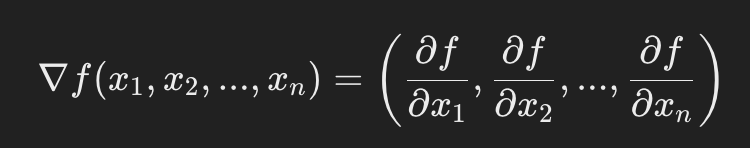
**R2<0: модель хуже среднего предсказания**.

**8. Чем log-cosh лучше функции потерь Хубера? Опишите обе функции потерь.**

* **Huber Loss**: сочетает квадратичное и линейное поведение; переключается на MAE для больших ошибок. Подходит для данных с выбросами.
* **Log-cosh**: гладкая функция, которая ведет себя как MSE для малых ошибок и как MAE для больших, но без резких переключений. **Лучше**, так как она обеспечивает более плавное поведение градиента.

**9. Что такое градиент? Какое его свойство используется при минимизации функций?**

* **Градиент**: вектор частных производных функции, указывающий направление наибольшего увеличения.
* **Свойство**: направление, противоположное градиенту, указывает путь к уменьшению функции (используется в градиентном спуске для минимизации).



**Метод градиентного спуска:**

Метод градиентного спуска заключается в том, что для минимизации функции (например, функции потерь в машинном обучении) мы движемся в направлении, противоположном градиенту, чтобы достичь точки минимума. Обновление параметров выполняется следующим образом:

Изображение выглядит как Шрифт, Графика, логотип, черный

Автоматически созданное описание

**10. Что такое градиентный спуск? Опишите процесс алгоритма.**

* **Градиентный спуск** — это итеративный метод оптимизации, который находит минимум функции путём движения в направлении, противоположном градиенту.
* **Процесс**:
  1. Инициализация весов модели случайными значениями.
  2. Вычисление градиента функции потерь.
  3. Обновление весов с шагом, пропорциональным градиенту и learning rate.
  4. Повторение шагов до достижения сходимости.

**11. Почему не всегда можно использовать полный градиентный спуск? Какие способы оценивания градиента вы знаете? Почему в стохастическом градиентном спуске важно менять длину шага по мере итераций? Какие стратегии изменения шага вы знаете?**

* **Полный градиентный спуск** не всегда используется из-за высоких вычислительных затрат для больших наборов данных.
* **Оценки градиента**:
  + Стохастический градиентный спуск (SGD),
  + Мини-батч градиентный спуск.
* **Изменение шага** важно для улучшения сходимости: на начальных этапах — большие шаги, ближе к минимуму — маленькие.
* **Стратегии изменения шага**:
  + Линейное уменьшение,
  + Экспоненциальное уменьшение,
  + Адаптивные методы (Adam, RMSprop).

**12. Что такое переобучение? Как можно отследить переобучение модели?**

* **Переобучение**: модель хорошо работает на тренировочных данных, но плохо обобщается на новые данные.
* **Способы отслеживания**:
  + Сравнение ошибок на тренировочном и тестовом наборе,
  + Использование кросс-валидации.

**13. Что такое кросс-валидация? На что влияет количество блоков в кросс-валидации?**

* **Кросс-валидация**: метод оценки модели путём разбиения данных на несколько блоков, где каждый блок поочередно используется для тестирования, а остальные — для обучения.
* **Количество блоков** влияет на стабильность оценки: чем больше блоков, тем меньше разброс результатов, но больше вычислительных затрат.

**14. Как построить итоговую модель после того, как по кросс-валидации подобраны оптимальные гиперпараметры?**

После подбора гиперпараметров на кросс-валидации:

1. Обучите модель с оптимальными гиперпараметрами на всём наборе данных.
2. Оцените модель на тестовом наборе для окончательной проверки.

**15. Что такое регуляризация? Для чего используется?**

* **Регуляризация** — метод добавления штрафа за сложность модели, чтобы предотвратить переобучение.
* **Использование**: снижает веса моделей, уменьшая влияние сложных зависимостей и шумов в данных.

**16. Опишите, как работают L1- и L2-регуляризаторы.**

**L1- и L2-регуляризаторы — это методы регуляризации, которые используются для предотвращения переобучения моделей машинного обучения путем добавления штрафа к функции потерь. Оба метода стремятся уменьшить сложность модели, но делают это разными способами.**

**1. L1-регуляризация (Lasso-регуляризация)**

**L1-регуляризация добавляет сумму абсолютных значений весов модели в функцию потерь. Она заставляет некоторые веса становиться равными нулю, тем самым приводя к разреженным моделям (то есть с меньшим количеством активных признаков).**

**Формула функции потерь с L1-регуляризацией:**

**Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание**

**Основные свойства L1-регуляризации:**

* **Приводит к разреженным решениям, так как штраф за абсолютное значение весов вынуждает многие веса становиться нулевыми. Это помогает автоматическому отбору признаков.**
* **Полезна в задачах, где есть много признаков, но не все из них значимы, поскольку она может "удалить" ненужные признаки, установив их веса в ноль.**

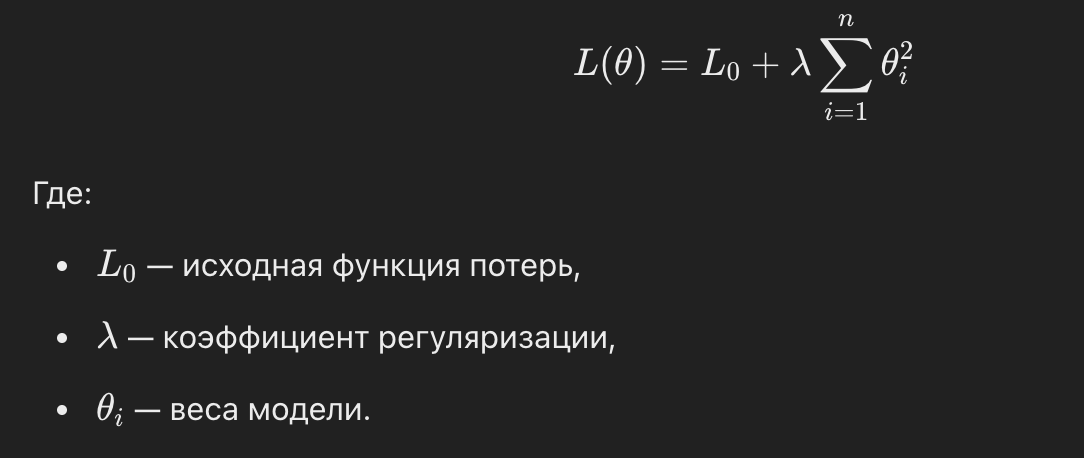
**Применение:**

* **Используется, когда важно получить интерпретируемую модель с небольшим количеством активных признаков.**
* **Популярна в линейных моделях, таких как Lasso-регрессия.**

**2. L2-регуляризация (Ridge-регуляризация)**

L2-регуляризация добавляет сумму квадратов весов модели в функцию потерь. Это штрафует большие веса, заставляя модель иметь более гладкие, меньшие значения параметров, но, в отличие от L1, она не обнуляет веса.

Формула функции потерь с L2-регуляризацией:

****

**Основные свойства L2-регуляризации:**

* **Штрафует большие веса: L2-регуляризация стремится сделать все веса маленькими, но не равными нулю.**
* **Гладкость решений: Модель с L2-регуляризацией обычно менее подвержена сильным колебаниям параметров и переобучению.**
* **Не приводит к разреженности модели, так как веса уменьшаются, но не становятся равными нулю.**

**Применение:**

* **L2-регуляризация полезна, когда все признаки могут быть потенциально важны, и важно контролировать масштаб весов.**
* **Широко используется в таких моделях, как Ridge-регрессия, и в нейронных сетях через метод веса L2 (weight decay).**

**Комбинированный метод: Elastic Net**

**Это комбинация L1- и L2-регуляризации, которая использует преимущества обоих методов. Elastic Net накладывает как штраф на абсолютные значения весов (L1), так и на их квадраты (L2).**

**Формула для Elastic Net:**

**Изображение выглядит как текст, Шрифт, снимок экрана, черный

Автоматически созданное описание**

**17. Почему L1-регуляризация отбирает признаки?**

L1-регуляризация создает линейный штраф на веса, что приводит к обнулению некоторых весов. Это помогает исключать несущественные признаки, оставляя только важные для предсказаний.

**18. Почему плохо накладывать регуляризацию на свободный коэффициент?**

Свободный коэффициент (интерсепт) отвечает за сдвиг модели и не связан с зависимостью между признаками и целевой переменной. Регуляризация на интерсепт может исказить общее предсказание модели, не улучшая её обобщающую способность.