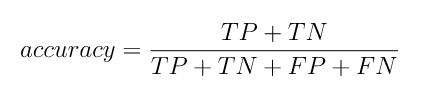
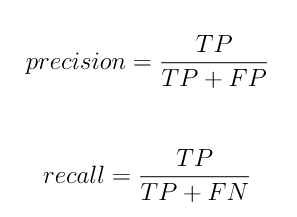
**1. Задачи классификации и регрессии в машинном обучении. Отличия. Метрики (accuracy, recall, precision, f1, roc-auc).**

* Регрессии – построение модели, способной предсказывать численную величину на основе набора признаков объекта.
* Классификация – определение категории объекта на основе его признаков



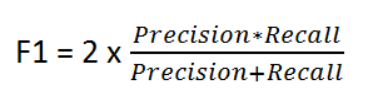


accuracy — доля правильных ответов алгоритма:

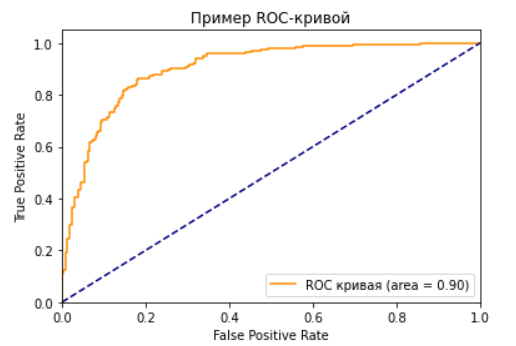


Precision можно интерпретировать как долю объектов, названных классификатором положительными и при этом действительно являющимися положительными,

recall показывает, какую долю объектов положительного класса из всех объектов положительного класса нашел алгоритм.

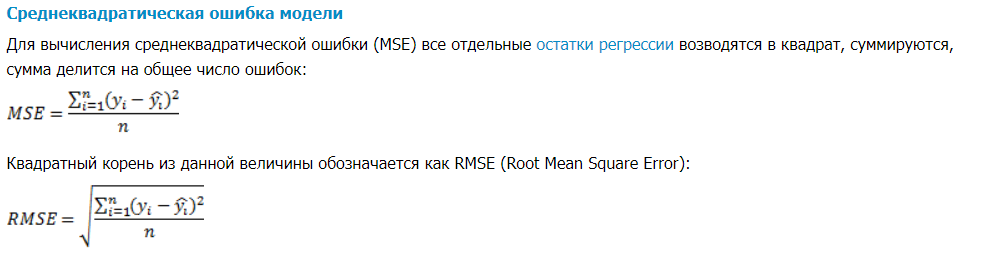


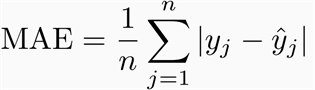
F1 Score поддерживает **баланс** между точностью и отзывчивостью. Мы используем его при неравномерном распределении классов, поскольку точность и отзыв могут привести к ошибочным результатам!



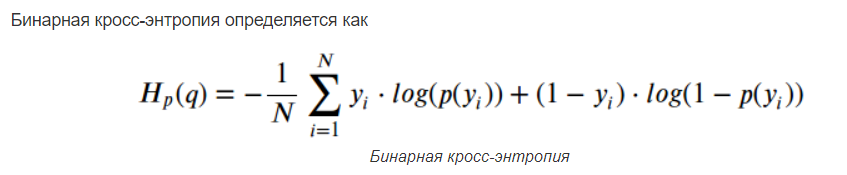
**2. Что такое функция потерь(ошибок)? MSE, MAE, cross-entropy, gini.**

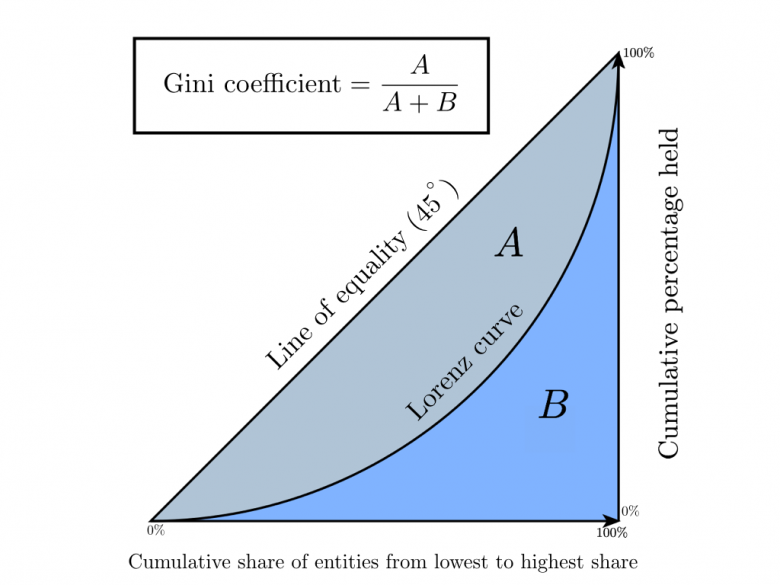
**Функция потерь** — [функция](https://ru.wikipedia.org/wiki/%D0%A4%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D1%8F_(%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0)), которая в [теории статистических решений](https://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D0%BE%D1%80%D0%B8%D1%8F_%D0%BF%D1%80%D0%B8%D0%BD%D1%8F%D1%82%D0%B8%D1%8F_%D1%80%D0%B5%D1%88%D0%B5%D0%BD%D0%B8%D0%B9) характеризует потери при неправильном принятии решений на основе наблюдаемых данных.





Кросс-энтропия (Перекрестная энтропия) – это Функция потерь, которую можно использовать для количественной оценки разницы между двумя Распределениями вероятностей.





Индекс Джини — это статистический показатель, с помощью которого можно описывать характер изменения одной величины относительно изменения другой. Кривая Лоренса показывает отклонения от настоящих предсказаний.

**3. Общий вид любой задачи машинного обучения.**



**4. Как работает и обучается решающее дерево для классификации и регрессии.**

****

Процесс построения деревьев решений заключается в последовательном, рекурсивном разбиении обучающего множества на подмножества с применением решающих правил в узлах. Процесс разбиения продолжается до тех пор, пока все узлы в конце всех ветвей не будут объявлены листьями.

**5. Как работает и обучается kNN**

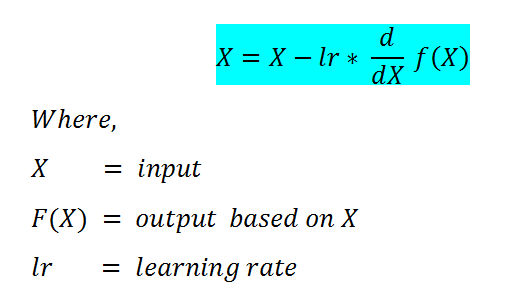
Для каждой записи в ДатаФрейме определяем ее класс по след алгоритму.

Поиск k ближайших соседей. Класс определяется большинством определенного класса соседей.

**6. Градиентный спуск. Для чего конкретно нужен в ML?**

Для минимизации функции ошибок.

Принцип: у нас есть функция ошибок мы пытаемся найти минимум этой функции. Берем начаальное приближение х0 и на каждой итерации уточняем его по формуле.



**(КРИО = «как работает и обучается»)**

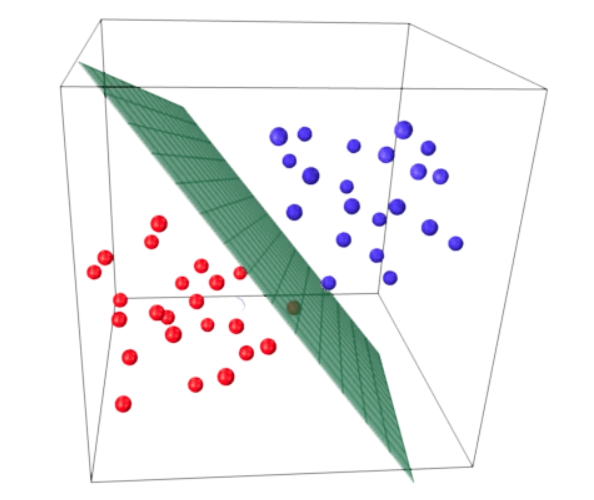
**7. КРИО Линейная регрессия.**

Задача регрессии в [машинном обучении](https://skillfactory.ru/machine-learning) — это предсказание одного параметра (Y) по известному параметру X, где X — набор параметров, характеризующий наблюдение.

Цель линейной регрессии — поиск линии, которая наилучшим образом соответствует этим точкам. Напомним, что общее уравнение для прямой есть f (x) = m⋅x + b, где m — наклон линии, а b — его y-сдвиг. Таким образом, решение линейной регрессии определяет значения для m и b, так что f (x) приближается как можно ближе к y.

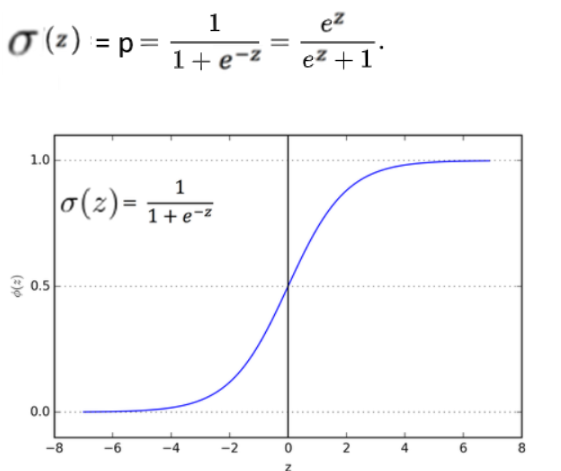
**8. Основная идея линейных классификаторов. Разделяющая гиперплоскость. Отступ объекта.**

Основная идея линейного классификатора заключается в том, что признаковое пространство может быть разделено гиперплоскостью на два полупространства, в каждом из которых прогнозируется одно из двух значений целевого класса.

****

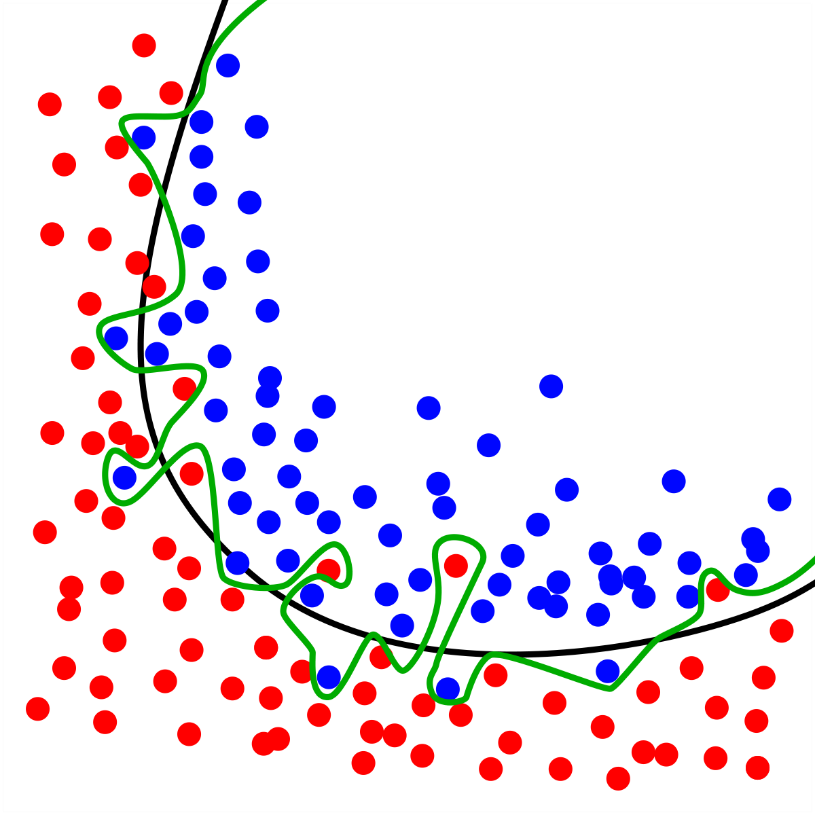
**Отступ (для классификатора)** — эвристика, оценивающая то, насколько объект "погружён" в свой класс, насколько эталонным представителем он является. Чем меньше значение отступа, тем ближе объект находится к границе класса, соответственно тем выше вероятность ошибочного прогноза.

**9. КРИО Логистическая регрессия.**

****

Мы можем видеть, что значение сигмоидной функции всегда лежит между 0 и 1. Значение точно равно 0,5 при X = 0. Мы можем использовать 0,5 в качестве порога вероятности для определения классов. Если вероятность больше 0,5, мы классифицируем ее как**Класс-1 (Y = 1)**или как**Класс-0 (Y = 0)**,

**10. Что такое переобучение? Как распознать? Способы борьбы.**



**Переобучение** — негативное явление, возникающее, когда алгоритм обучения вырабатывает предсказания, которые слишком близко или точно соответствуют конкретному набору данных и поэтому не подходят для применения алгоритма к дополнительным данным или будущим наблюдениям.

**11. Ансамблирование - бэггинг, бустинг, стэкинг. В чём суть каждого подхода.**

**12. КРИО случайный лес.**

**13. Персептрон(полносвязная нейронная сеть). Модель искусственного нейрона.**

**14. Зачем нужны функции активации в нейронах? Каким свойством они обладают**

**15. Метод обратного распространения ошибки (концептуально, зазубривать формулы не надо).**

**16. Как работает свёртка, свёрточный нейрон, свёрточный слой. Пример архитектуры свёрточной нейросети для классификации.**

**17. Что делать, если мало данных для обучения сети? Transfer learning, аугментации, dropout как ещё один метод борьбы с переобучением.**