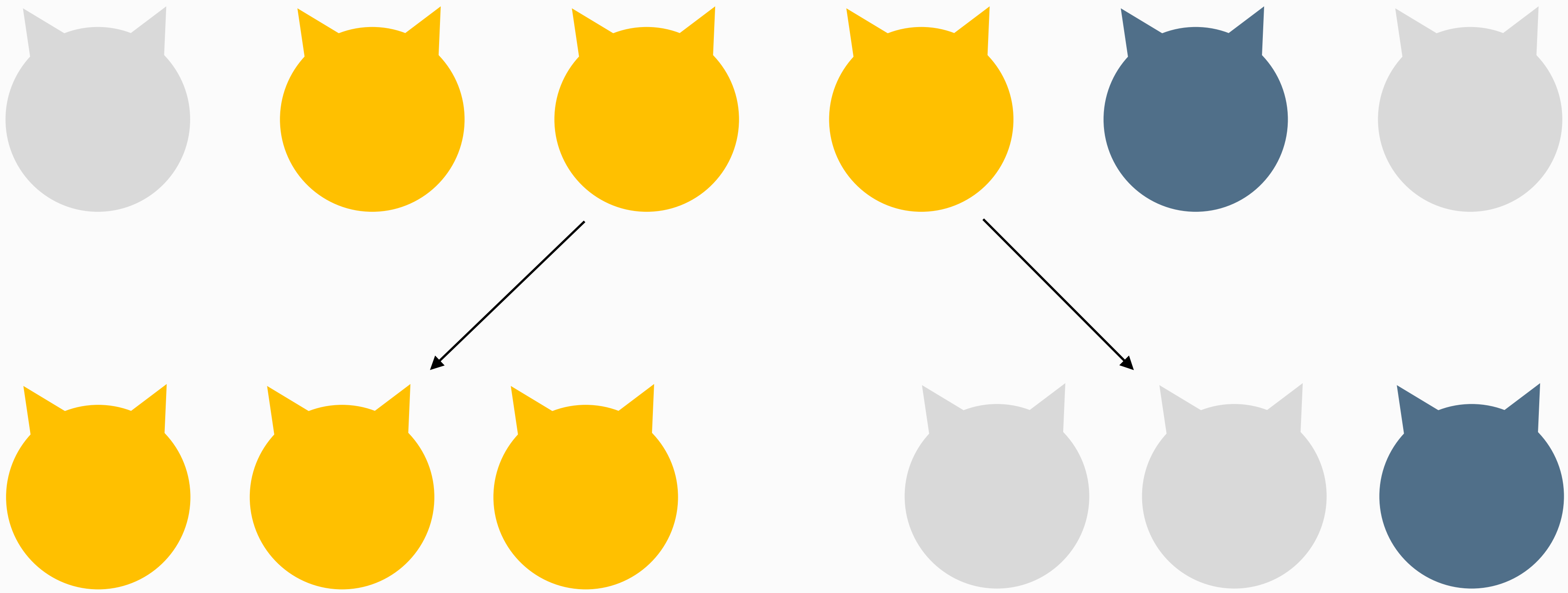
H(R) в задаче мягкой классификации

Skillbox
Елена Кантонистова

Пример

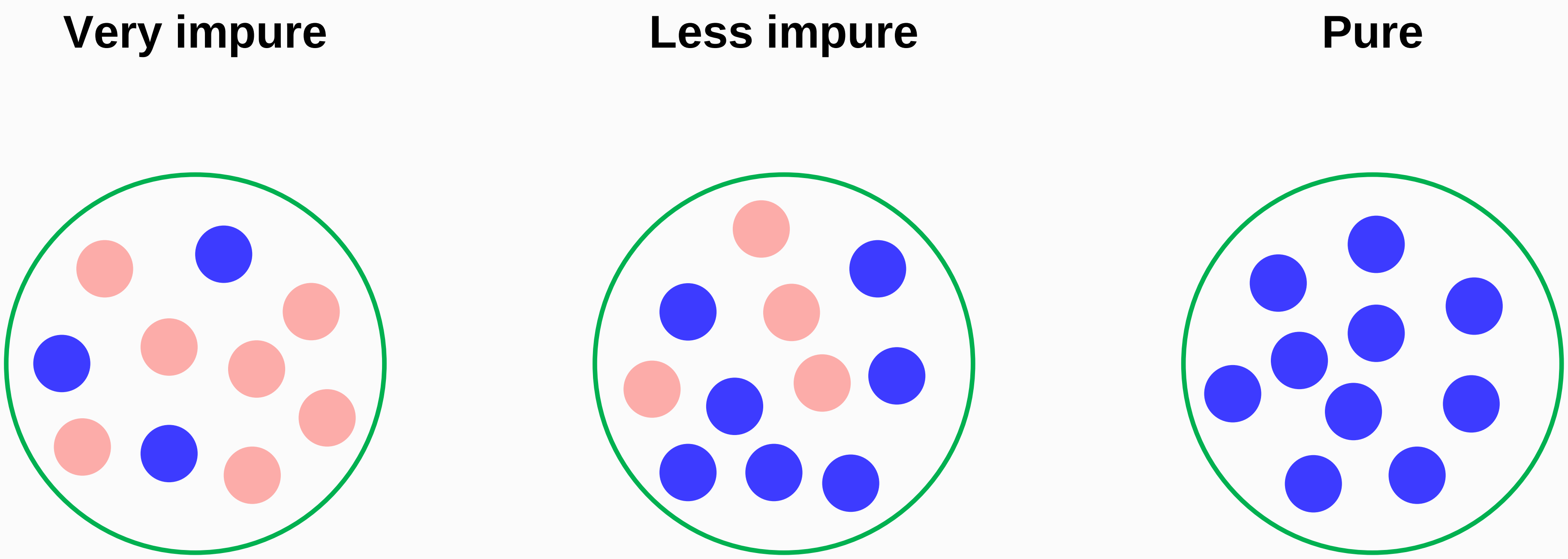
Пронумеруйте классы котиков: рыжий котик — 1, белый — 2, чёрный — 3.



$y1 = (1, 0, 0)$ $y2 = (0, 2/3, 1/3)$

Энтропия

Энтропия — мера хаоса.



Автор: Кантонистова Елена

H(R) для мягкой классификации

- Первый способ — **энтропия**

$$H(R) = -\sum_{k=1}^{K} p_k \cdot \log p_k \,,$$

где p_k — доля объектов *k*-го класса в вершине

H(R) для мягкой классификации

- Первый способ — **энтропия**

$$H(R) = -\sum_{k=1}^{K} p_k \cdot \log p_k\,,$$

где p_k — доля объектов *k*-го класса в вершине

- Левая вершина:

$$H(R_l) = -1 \cdot log1 - 0 - 0 = 0$$

- Правая вершина:

$$H(R_r) = -0 - \frac{1}{3}log\frac{1}{3} - \frac{2}{3}log\frac{2}{3} \approx 0.92$$

Схема: пример

Энтропия

- График энтропии для двухклассовой классификации:

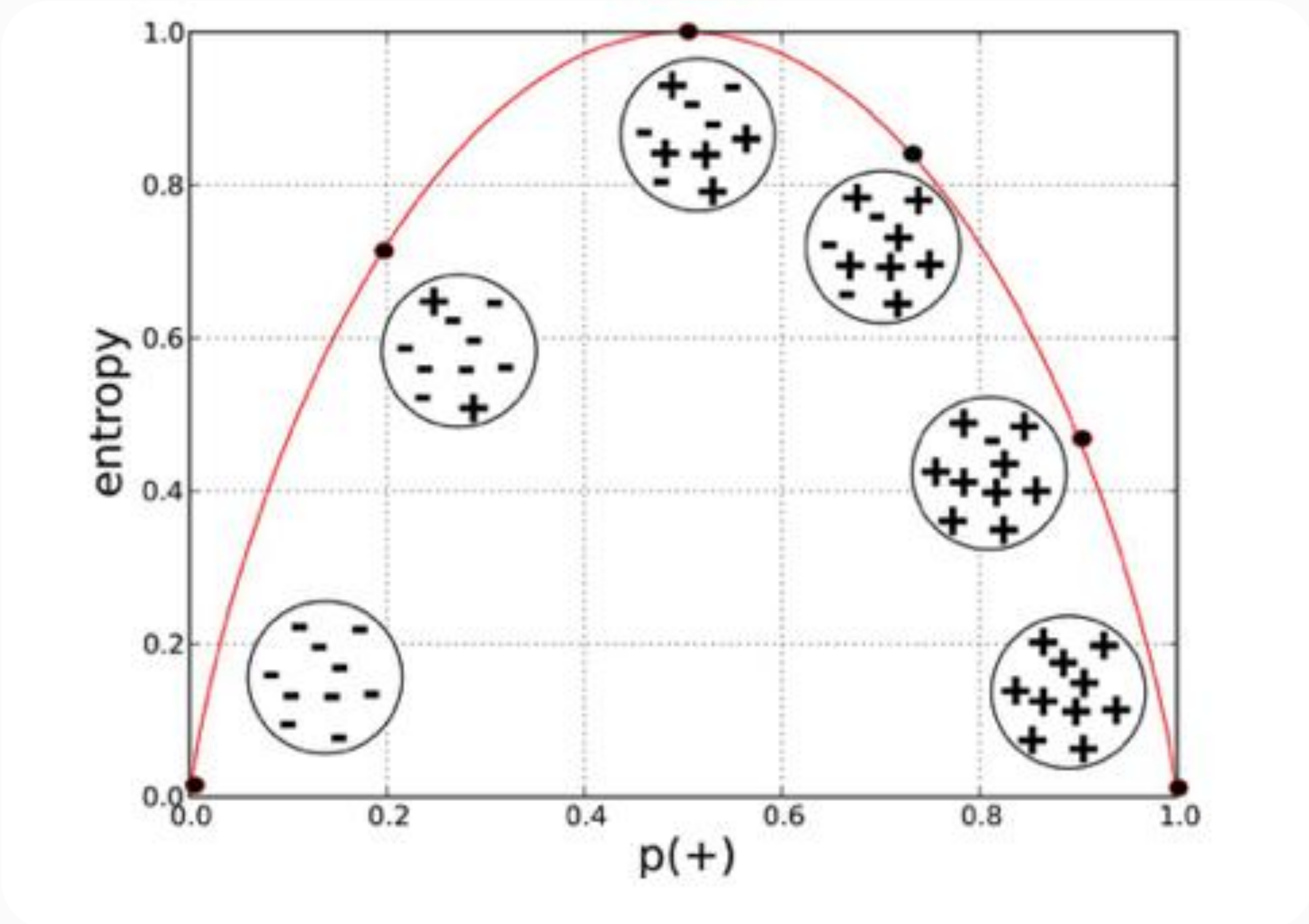


Схема: энтропия

H(R) для мягкой классификации

- Первый способ — **критерий Джини**: $$H(R) = \sum_{k=1}^{K} p_k \cdot (1 - p_k),$$

  где p_k — доля объектов *k*-го класса в вершине

H(R) для мягкой классификации

- Первый способ — **критерий Джини**: $$H(R) = \sum_{k=1}^{K} p_k \cdot (1 - p_k),$$

  где p_k — доля объектов *k*-го класса в вершине

- Левая вершина:

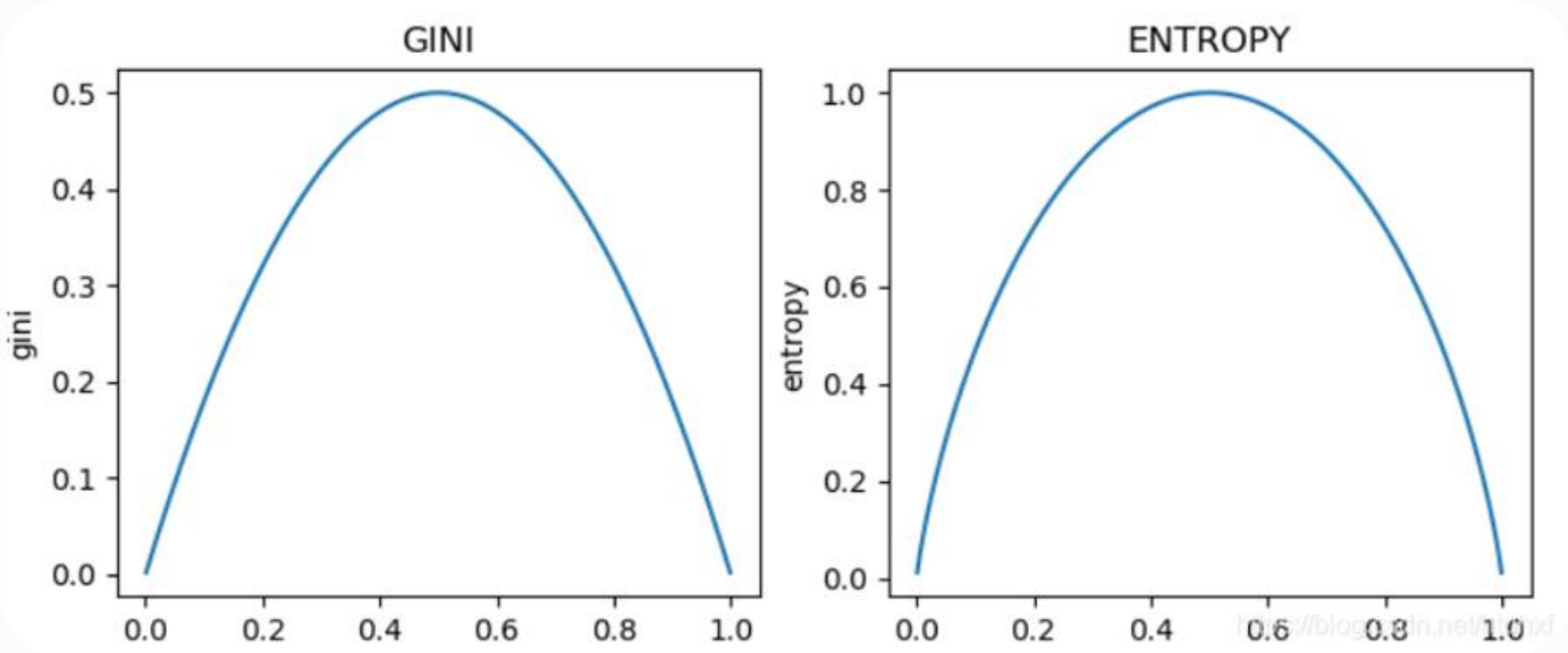
$$H(R) = 0$$

- Правая вершина:

$$H(R) = 0 + \frac{1}{3} \cdot \frac{2}{3} + \frac{2}{3} \cdot \frac{1}{3} = \frac{4}{9} \approx 0.44$$

Схема: пример

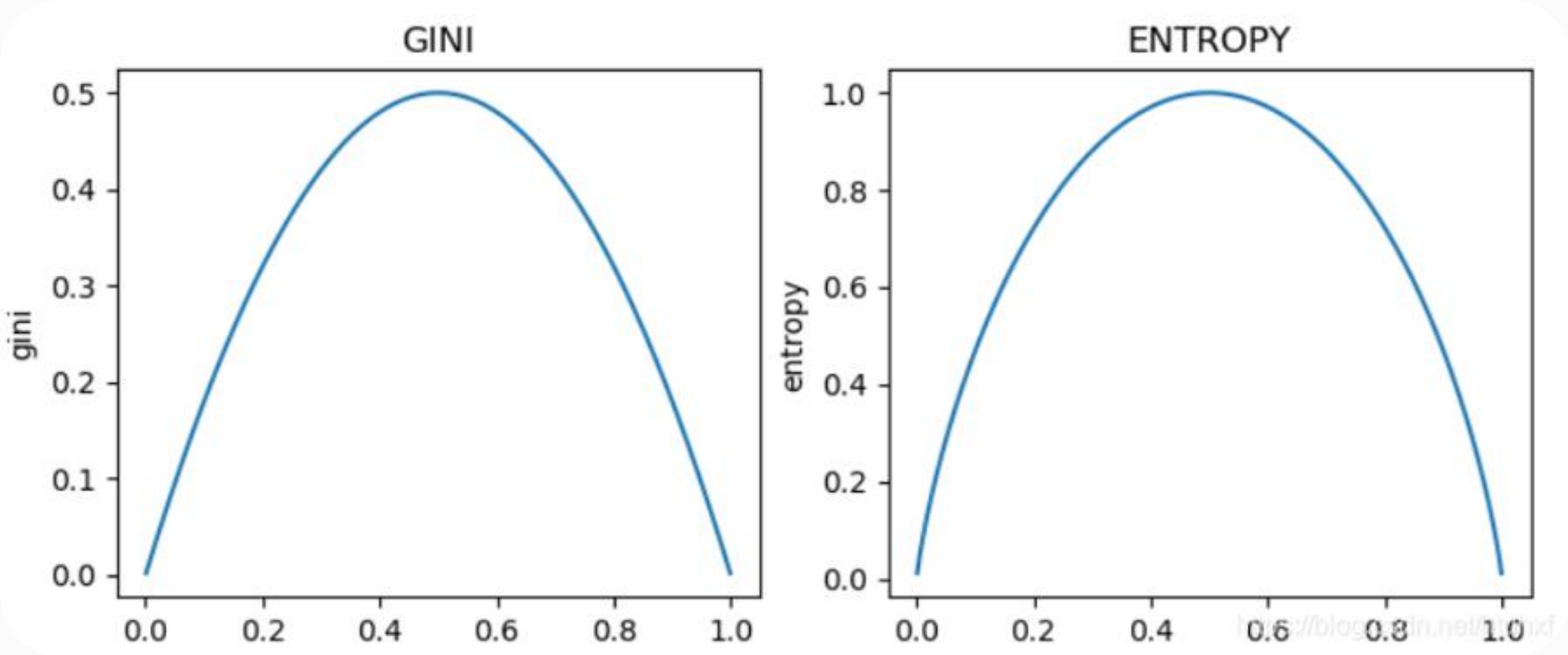
Энтропия и критерий Джини



- Если при построении решающего дерева вы используете энтропию, то такой алгоритм в литературе называется ID3 (Iterative Dichotomiser 3)
- Если же вы используете критерий Джини, то такой алгоритм называется CART (Classification And Regression Tree)

Автор: Кантонистова Елена

Итоги



- Вы узнали, что для построения решающего дерева в задаче мягкой классификации используются энтропия и индекс Джини

Автор: Кантонистова Елена