

ML task

1) Какая вероятность того, что мы выявили брак, если прибор выдал положительный результат - "продукция бракованная"?

2) Почему же в жизни все-таки используют такие приборы?
Что можно было бы изменить в процедуре поиска брака, не меняя вероятность ошибки прибора, так, чтобы вероятность из первого вопроса $P(\text{брак} | "+")$ выросла?

3) Какое соотношение можно вывести между процентом брака $P(\text{брак})$ и ошибкой прибора, если мы хотим, чтобы прибор работал лучше честной монетки, хуже или также?

Как оформить: `ivan_petrov.pdf` (подставьте ваше имя и фамилию) файл

#1:

Пусть

- "+" - событие, при котором прибор показывает "брак"
- "—" - прибор показывает, что брака нет
- f - брака нет на самом деле
- t - брак на самом деле есть

Тогда по формуле Байеса, вероятность, что прибор верно классифицировал брак:

$$P(f|+) = \frac{P(f) * P(+|f)}{P(+)} = \frac{P(f) * P(+|f)}{P(t) * P(+|t) + P(f) * P(+|f)} = \frac{0.05 * 0.95}{0.95 * 0.5 + 0.05 * 0.95} = 0.5$$

#2:

Такие приборы могут быть использованы как "ансамбль" слабых классификаторов, если точность таких приборов будет больше 50% (точность классификации в целом). Чем больше таких классификаторов, тем с большей вероятностью они большинством проголосуют за верный результат

#3

$let\ fault := P(f)$

$let\ err := P(\text{ошибка прибора})$. Тогда:

$$relation := \frac{P(f) * P(+|f)}{P(t) * P(+|t) + P(f) * P(+|f)} = \frac{fault * (1 - err)}{(1 - fault) * err + fault * (1 - err)} =$$

$$= \frac{fault - err * fault}{err + fault - 2err * fault} ? 0.5$$

$$2 * err - 2 * err * fault ? err + fault - 2 * err * fault$$

$$fault ? err$$

$$P(\text{брак}) ? P(\text{ошибка прибора})$$

=> если $P(\text{брак}) > P(\text{ошибка прибора})$, то прибор работает лучше монетки, если $P(\text{брак}) < P(\text{ошибка прибора})$, то хуже, а если $P(\text{брак}) = P(\text{ошибка прибора})$, то так же