

Семинар 6-7: Удержание клиентов

Задача 1

Бандерлог начинает все определения со слов «это доля правильных ответов»:

- а) ассигасу — это доля правильных ответов...
- б) точность (precision) — это доля правильных ответов...
- в) полнота (recall) — это доля правильных ответов...
- г) TPR — это доля правильных ответов...

Решение:

Закончите определения Бандерлога так, чтобы они были, хм, правильными.

- а) $\text{accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$
- б) $\text{precision} = \frac{TP}{TP+FP}$
- в) $\text{recall} = \frac{TP}{TP+FN}$
- г) $\text{TPR} = \frac{TP}{TP+FN}$

Задача 2

Бандерлог оценил три модели: нейросеть, случайный лес и KNN. Он построил на тестовой выборке прогнозы и получил три матрицы ошибок:

| | $y = 1$ | $y = 0$ |
|---------------|---------|---------|
| $\hat{y} = 1$ | 80 | 20 |
| $\hat{y} = 0$ | 20 | 80 |

| | $y = 1$ | $y = 0$ |
|---------------|---------|---------|
| $\hat{y} = 1$ | 48 | 2 |
| $\hat{y} = 0$ | 52 | 98 |

| | $y = 1$ | $y = 0$ |
|---------------|---------|---------|
| $\hat{y} = 1$ | 10 | 20 |
| $\hat{y} = 0$ | 90 | 10000 |

- а) Найдите для всех трёх моделей долю правильных ответов.
- б) Найдите для всех трёх моделей точность (precision) и полноту (recall)
- в) Несколько ситуаций, когда чё выбирать — ?
- г)

Решение:

- а)
- б)
- в)
- г)

Задача 3

Бандерлог из Лога¹ ведёт блог, любит считать логарифмы и оценивать логистические регрессии. С помощью нового алгоритма Бандерлог решил задачу классификации по трём наблюдениям и получил $b_i = \hat{p}(y_i = 1|x_i)$.

| y_i | b_i |
|-------|-------|
| 1 | 0.7 |
| 0 | 0.2 |
| 0 | 0.3 |

1. Постройте ROC-кривую.
2. Найдите площадь под ROC-кривой и индекс Джини.
3. Постройте PR-кривую (кривая точность-полнота).
4. Найдите площадь под PR-кривой.
5. Как по-английски будет «бревно»?

Решение:

Задача 4

1. Ещё задачи

Задача 5

Маша обучила логистическую регрессию и получила вектор предсказанных вероятностей принадлежности классу +1

$$p = [0.9 \text{ amp}; 0.1 \text{ amp}; 0.75 \text{ amp}; 0.56 \text{ amp}; 0.2 \text{ amp}; 0.37 \text{ amp}; 0.25],$$

а вектор правильный ответов равен

$$y = [+1 \text{ amp}; -1 \text{ amp}; -1 \text{ amp}; +1 \text{ amp}; +1 \text{ amp}; -1 \text{ amp}; -1]$$

.

Бинаризируйте ответ по порогу t и посчитайте точность, полноту и F1-меру.

$$t = 0.3 \quad t = 0.8$$

¹деревня в Кадуйском районе Вологодской области

Задача 6

Задача 7