

Логистическая регрессия

① Обучение логистической регрессии

Пусть есть n объектов, на которых мы наблюдаем значения признаков x_1, \dots, x_k и отклика y .

$$\begin{pmatrix} x_{11} & \dots & x_{1k} \\ \vdots & & \\ x_{n1} & \dots & x_{nk} \end{pmatrix} \text{ и } \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

При этом $y_i \in \{0, 1\}$.

Предполагаем, что y_i — это реализации случайной величины Y , которая принимает значение 1 с вероятностью $p(x)$ и значение 0 с вероятностью $1 - p(x)$.

Как выглядит функция

$$p_+(x_i) = p_+(x_{i1}, \dots, x_{ik})?$$

Будем считать, что $p_+(x_i) = f(\underbrace{\theta_0 + \theta_1 \cdot x_{i1} + \dots + \theta_k \cdot x_{ik}}_{\langle \theta, x_i \rangle})$

$\theta_0 + \theta_1 x_{i1} + \dots + \theta_k x_{ik} \in \mathbb{R}$
 $p_+(x_i) \in [0, 1]$

$$\Rightarrow f: \mathbb{R} \rightarrow [0, 1].$$

В модели логистической регрессии выбирается

$$f(z) = \frac{1}{1 + e^{-z}} \leftarrow \text{логистическая функция.}$$

Но,

$$p_+(x_i, \theta) = \frac{1}{1 + e^{-\langle \theta, x_i \rangle}}$$

Как обучить модель логистической регрессии?

Воспользуемся методом максимальной правдоподобия.

$$L(\theta) = P(Y=y_1) \cdot \dots \cdot P(Y=y_n).$$

↑
θ-ия
правдоподобия

$$P(Y=y_i) = \begin{cases} p_+(x_i, \theta), & \text{если } y_i=1 \\ 1 - p_+(x_i, \theta), & \text{если } y_i=0 \end{cases}$$

Можем записать $P(Y=y_i)$ в виде

$$P(Y=y_i) = p_+(x_i, \theta)^{y_i} \cdot (1 - p_+(x_i, \theta))^{1-y_i}$$

Тогда

$$L(\theta) = \prod_{i=1}^n p_+(x_i, \theta)^{y_i} \cdot (1 - p_+(x_i, \theta))^{1-y_i}.$$

Вместо $L(\theta) \rightarrow \max$ будем решать задачу

- $\ln L(\theta) \rightarrow \min$:

$$\begin{aligned} -\ln L(\theta) &= -\ln \prod_{i=1}^n p_+(x_i, \theta)^{y_i} \cdot (1 - p_+(x_i))^{\cdot 1-y_i} = \\ &= -\sum_{i=1}^n (y_i \cdot \ln p_+(x_i, \theta) + (1-y_i) \cdot \ln (1 - p_+(x_i, \theta))) = \\ &= \boxed{-\sum_{i=1}^n \left(y_i \cdot \ln \frac{1}{1+e^{-\langle \theta, x_i \rangle}} + (1-y_i) \ln \left(1 - \frac{1}{1+e^{-\langle \theta, x_i \rangle}} \right) \right)} \rightarrow \min_{\theta} \end{aligned}$$

↑
Функция Logistic Loss

Классификация с помощью логистической регрессии

Обученная логистическая регрессия может использоваться не только для оценивания вероятности того, что объект принадлежит классу +1, но и для задач классификации.

Рассмотрим объект со значениями признаков x_{j1}, \dots, x_{jk} .

Если $p_+(x_j, \theta) > t$ \Rightarrow относим данный объект к классу +1.

порог
 $t \in [0, 1]$

Выбор порога t — это отдельная задача (решение которой часто зависит от области применения логистической регрессии).

Метрики качества классификации

① Метрики при фиксированном значении порога

Пусть $a(x_j) = \begin{cases} 1, & \text{если } p_+(x_j, \theta) > t \\ 0, & \text{иначе} \end{cases}$
классификатор

$$\text{accuracy} = \frac{1}{m} \sum_{j=1}^m [a(x_j) = y_j]$$

доля правильных ответов классификатора

+: легко интерпретируема

-: выдает некорректные результаты при несбалансированных данных

Матрица ошибок

		$y=1$	$y=0$
		True Positive (TP)	False Positive (FP)
$a(x)=1$	True Negative (TN)	False Negative (FN)	

$$\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{precision} = \frac{TP}{TP + FP}$$

точность (доля правильно распознан. "1" из всех объектов, распознанных как "1")

Пример: судебная система ($1 = \text{виновен}$)

$$\text{recall} = \frac{TP}{TP + FN}$$

полнота (доля правильно распознан. "1" из всех "1")

Пример: поиск поимен. клиентов

Пример:

		1 не вернул (бы) кредит	0 вернул (бы) кредит
не вернет кредит	1	TP	FP
	0 вернет кредит	FN	TN

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

доля правильно распознан.
"1" из всех "1".

Если precision будет маленькой, то многие хорошие потенциальные клиенты не получат кредит.

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

доля правильно распознан.
"1" из всех объектов,
распознан. как "1".

Если precision будет маленькой, то многие хорошие потенциальные клиенты не получат кредит

Что ваннее?
Надо учитывать финансовые потери при тех или иных ошибках.

F-мера:

$$\underbrace{F_1}_{\substack{\uparrow \\ \text{ближка к нулю, если precision или recall}}} = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

ближка к нулю, если precision или recall
ближки к нулю

$$\underbrace{F_B}_{\substack{\beta < 1, \text{ если важнее точность} \\ \beta > 1, \text{ если важнее полнота}}} = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{\beta \cdot \text{precision} + \text{recall}}$$

$\beta < 1$, если важнее точность

$\beta > 1$, если важнее полнота

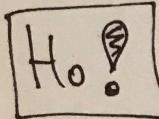
Пример несбалансированных данных, на которых
accuracy дает плохой результат

		1 (spam)	0 (не-spam)
		1 (spam)	0 (не-spam)
1 (spam)	TP=5	FP=10	
0 (не-spam)	FN=5	TN=90	

Есть 100 не-spam писем

10 spam писем

$$\Rightarrow \text{accuracy} = \frac{5+90}{110} = 0,86$$



Если все письма без разбора отправить

вне-spam, то

		1	0
		1	0
1	TP=0	FP=0	
0	10	100	

$\Rightarrow \text{accuracy} = \frac{100}{110} = 0,91$

[Здесь выше FPR!]

II ROC-кривая и AUC-ROC

$$FPR = \frac{FP}{FP + TN}$$

False Positive Rate

$$TPR = \frac{TP}{TP + FN} (= \text{recall})$$

True Positive Rate

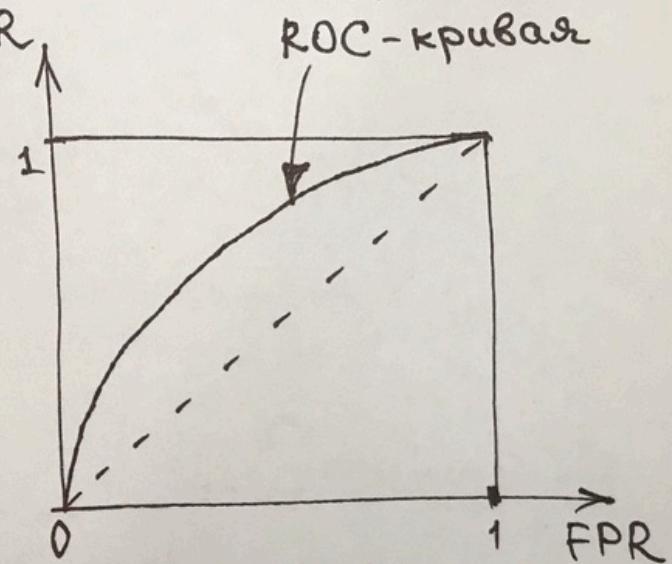
AUC-ROC = площадь
под
ROC-кривой
(Area Under ROC Curve)

Если классификатор $a(x)$ не допускает ошибок \Rightarrow
AUC-ROC = 1.

Если $a(x)$ классифицирует случайным образом \Rightarrow AUC-ROC ≈ 0.5 .

Receiver Operator Characteristic

ROC-кривая



Каждая точка на
графике соответствует
некоторому значению
порога $t \in [0,1]$