

Задачи по темам

- Введение в статистическое обучение (лекция 1)
- Параметрические классификаторы для бинарного Y (лекция 2)

(1) Задача

i	y_i	\hat{y}_i
1	1	1
2	0	0
3	0	1
4	1	1
5	1	1
6	1	1
7	0	1
8	1	1
9	0	0
10	1	0
11	1	1
12	0	0
Σ	7	8

В таблице слева приводятся фактически значения Y и значения прогноза по модели бинарной классификации. Значение 1 означает наличие признака, 0 – отсутствие признака. Верны ли следующие утверждения:

- Точность модели (обобщающий показатель) составляет 0,25.
- Модель точнее прогнозирует наличие признака, чем его отсутствие.
- Модель точнее прогнозирует отсутствие признака, чем его наличие.
- Положительному прогнозу по модели стоит доверять больше, чем отрицательному.
- Положительному прогнозу по модели стоит доверять меньше, чем отрицательному.
- Можно утверждать, что предсказанные классы положительно коррелируют с истинными.

(2) Задача 4.6¹

Представьте, что мы собрали данные для группы студентов, посещающих курс по статистике, с переменными X_1 – количество часов, посвящённых изучению предмета, X_2 – средний балл в школе и Y – индикатор получения самой высокой отметки за экзамен. Мы подогнули логистическую регрессию и получили оценки коэффициентов $\hat{\beta}_0 = -6$, $\hat{\beta}_1 = 0.05$, $\hat{\beta}_2 = 1$.

- Рассчитайте вероятность того, что студент, затративший на обучение 40 часов и имеющий средний школьный балл 3.5, покажет наилучший результат на экзамене.
- Сколько часов студенту из вопроса а) необходимо затратить на обучение, чтобы с 50%-ной вероятностью получить наивысшую отметку за экзамен?

(3) Задача 4.7

Представим, что мы хотим предсказать выплату дивидендов по некоторым акциям в этом году («да» или «нет») на основе X – прибыли, выраженной в процентах от объёма продаж. Мы изучаем большое количество компаний и выясняем, что среднее значение X для компаний, выплативших дивиденды, составило $\bar{X} = 10$, тогда как для компаний, которые дивиденды не выплатили, оно составило $\bar{X} = 0$. Кроме того, дисперсия X у этих двух групп компаний составила $\hat{\sigma}^2 = 36$. Наконец, дивиденды были выплачены 80% компаний. Допустив, что X имеет нормальное распределение, рассчитайте вероятность выплаты дивидендов некоторой компанией в этом году при условии, что в прошлом году прибыль, выраженная в процентах от объёма продаж, составила $X = 4$.

Подсказка: вспомните, что функция плотности вероятности случайной нормально распределённой переменной выражается как $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$. Вам придётся воспользоваться теоремой Байеса.

¹ Задачи с нумерацией X.Y – из книги «Введение в статистическое обучение с примерами на языке R». X – номер главы, Y – номер задания из раздела теоретических упражнений.