

# Три сюжета напоследок

Эконометрика. Лекция 10

- Квантильная регрессия
- Алгоритм случайного леса
- Байесовский подход

Моделировать можно не только среднее, но и медиану или другой определённый квантиль.

# Классическая регрессия — модель для среднего

Предпосылки классической модели:

- $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$
- экзогенность,  $E(\varepsilon_i | x_i) = 0$
- другие предпосылки

Следствие:

$$E(y_i | x_i) = \beta_1 + \beta_2 x_i$$

# Минимизация суммы квадратов

Модель:  $E(y_i|x_i) = \beta_1 + \beta_2 x_i$

- Сумма квадратов остатков,  $Q(\hat{\beta}_1, \hat{\beta}_2) = \sum_i (y_i - \hat{y}_i)^2$
- Минимизируя  $Q(\hat{\beta}_1, \hat{\beta}_2)$  получаем состоятельные оценки  $\hat{\beta}_1, \hat{\beta}_2$

# Медианная регрессия

Модель:  $Med(y_i|x_i) = \beta_1 + \beta_2 x_i$

На большой выборке:

Математическое ожидание — среднее арифметическое значение объясняемой переменной  $y_i$  при заданном  $x_i$

Медиана,  $Med(y_i|x_i)$  — число, больше которого оказывается ровно половина  $y_i$  при заданном  $x_i$

# Алгоритм получения оценок

- Сумма модулей остатков,  $M(\hat{\beta}_1, \hat{\beta}_2) = \sum_i |y_i - \hat{y}_i|$
- Минимизируя  $M(\hat{\beta}_1, \hat{\beta}_2)$  получаем состоятельные оценки  $\hat{\beta}_1, \hat{\beta}_2$

# Пример у неоновой доски

Найдите оценку  $\hat{\beta}$  медианной регрессии:

$$Med(y_i|x_i) = \beta x_i$$

Набор данных:

y	x
1	1
2	5
6	5



# Медианная и классическая регрессия

- Классическая: от каких факторов зависит  $E(y_i|x_i)$ ?
- Медианная: от каких факторов зависит  $Med(y_i|x_i)$ ?
- Оценки  $\hat{\beta}_j$  и  $se(\hat{\beta}_j)$  считаются по разным формулам
- Если распределение  $\varepsilon_i$  симметрично, то оба подхода дают асимптотически одинаковые оценки
- Сходная проверка гипотез:  $t = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \rightarrow N(0, 1)$

# Медианная регрессия: минусы

- Нет явных формул для оценок коэффициентов и стандартных ошибок
- Только асимптотические свойства оценок коэффициентов

# Медианная регрессия: плюсы

- Взгляд на данные с другой стороны
- Более устойчивые оценки в случае “выбросов” в  $\varepsilon_i$

- Медиана,  $Med(y_i)$ , — квантиль 50%

$$P(y_i \leq Med(y_i)) = 0.5$$

- Квантиль порядка  $\tau$ ,  $q_\tau$ :

$$P(y_i \leq q_\tau) = \tau$$

- Например:

Квантиль порядка 10% для  $y_i$  — такое число  $q_{0.1}$ , что вероятность того, что  $y_i$  окажется меньше этого числа, равна 10%.

# Квантильная регрессия

Модель:  $q_\tau(y_i|x_i) = \beta_1^\tau + \beta_2^\tau x_i$

- Зависимость для разных квантилей может быть разная!

# Асимметричная сумма модулей остатков:

$$M(\hat{\beta}_1, \hat{\beta}_2) = \sum_i w_i \cdot |y_i - \hat{y}_i|$$

где веса  $w_i$  равны:

$$w_i = \begin{cases} (1 - \tau), & y_i < \hat{y}_i \\ \tau, & y_i \geq \hat{y}_i \end{cases}$$

- Минимизируя  $M(\hat{\beta}_1, \hat{\beta}_2)$  получаем состоятельные оценки  $\hat{\beta}_1, \hat{\beta}_2$

# Квантильная регрессия стоимости квартир

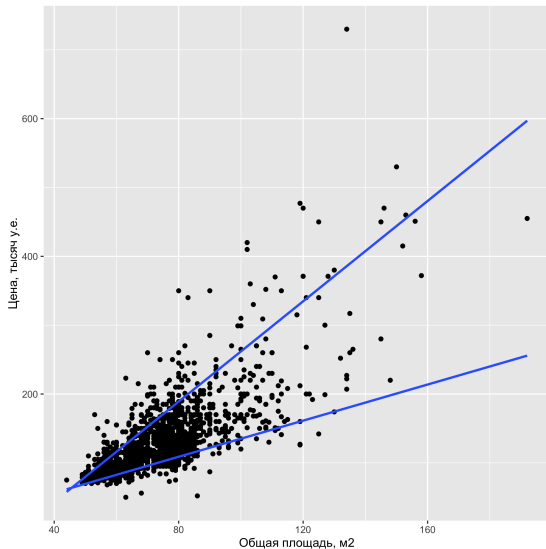
недорогое жильё (10%-ый квантиль):

$$\widehat{price}_i = 3.9 + 1.3 totsp_i$$

дорогое жильё (90%-ый квантиль):

$$\widehat{price}_i = -102.4 + 3.6 totsp_i$$

# Квантильная регрессия стоимости на графике





# Алгоритм случайного леса

- Очень хорошо прогнозирует
- Не объясняет, как устроены данные

# Две версии алгоритма

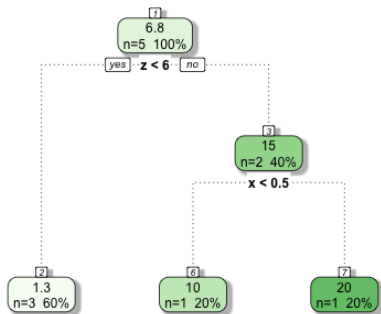
- Для непрерывной  $y_i$
- Для качественной  $y_i$

# Каждый мужчина должен посадить дерево

Набор данных

y	x	z
1	1	-2
1	0	3
2	0	-4
10	0	9
20	1	9

# Каждый мужчина должен посадить дерево



# Как посадить дерево?

- Из имеющихся  $k$  переменных случайно отбираем  $k' = \lceil k/3 \rceil$  переменных
- Из отобранных  $k'$  переменных выбираем ту, которая даёт наилучшее деление ветви дерева на две
- Повторяем до тех пор, пока в каждом терминальном узле остаётся больше *nodesize* = 5 наблюдений

# Наилучшее деление

До деления:  $RSS = 274.8$

$\{1, 1, 2, 10, 20\}$ ,  $\hat{y} = \bar{y} = 6.8$ ,

После разбиения:  $RSS = RSS_1 + RSS_2 = 50.67$

Слева:  $\{1, 1, 2\}$ ,  $\hat{y} = \bar{y} = 1.33$ ,  $RSS_1 = 0.67$

Справа:  $\{10, 20\}$ ,  $\hat{y} = \bar{y} = 15$ ,  $RSS_2 = 50$

Повторное применение алгоритма к тому же набору данных даст слегка другие оценки

# Мужчина, владеющий R, может посадить целый лес!

- Случайным образом отбираем (с повторениями)  $n$  наблюдений из исходных  $n$  наблюдений
- Сажаем дерево по случайной подвыборке
- Повторяем до получения  $n_{tree} = 500$  деревьев



# Прогноз случайного леса:

- Каждое из  $n_{tree} = 500$  деревьев даёт свой прогноз  $\hat{y}_i$
- Усредняем и получаем финальный прогноз

# Неоновая доска. Пример построения регрессионного дерева

y	x
1	1
2	2
9	3
10	4
10	5

Опишем наше незнание параметра  $\theta$  в виде априорного закона распределения!

# Пример. Неизвестная вероятность

- $p \in [0; 1]$

Априорная плотность:

$$f(p) = \begin{cases} 1, & p \in [0; 1] \\ 0, & \text{иначе} \end{cases}$$

# Пример. Неизвестный положительный коэффициент

- $\beta \in [0; +\infty)$

Априорная плотность:

$$f(\beta) = \begin{cases} \exp(-\beta), & \beta \in [0; \infty) \\ 0, & \text{иначе} \end{cases}$$

Модель задаёт закон распределения наблюдений,  $y_i$ , при фиксированном значении параметров

Например,

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

Определяем:

- Априорное распределение,  $f(\theta)$
- Модель для данных,  $f(y|\theta)$

По формуле условной вероятности получаем:

- Апостериорное распределение,  $f(\theta|y)$

# Формула условной вероятности

$$f(\theta|y) = \frac{f(y|\theta) \cdot f(\theta)}{f(y)} \sim f(y|\theta) \cdot f(\theta)$$



# Пример у неоновой доски

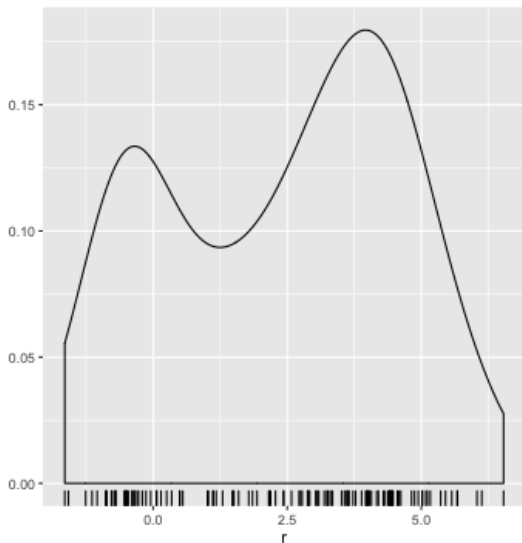
Наблюдения: пойманы 2 карася и щука.

Отдельные наблюдения независимы, вероятность поймать щуку и карася стабильна во времени.

Найдите апостериорную плотность вероятности поймать карася в пруду.

- нет информации
- Бабушка: караси встречаются чаще щук!

# Как описать сложную функцию плотности?



# Сколь угодно точное описание любой плотности!

- Большая выборка независимых значений случайной величины  $r$ :

$r_1, r_2, r_3, \dots, r_{10000}$

- Можно оценить всё:  $E(r)$ ,  $E(r^2)$ ,  $P(r > 0)$

# Монте-Карло по схеме Марковской цепи

MCMC (Markov Chain Monte Carlo)

Заменяет формулу условной вероятности

На входе:

- Априорное распределение,  $f(\theta)$
- Модель для данных,  $f(y|\theta)$

На выходе:

- Большая выборка из апостериорного распределения,  $f(\theta|y)$

Повторное применение алгоритма к тому же набору данных даст слегка другие оценки

# Плюсы байесовского подхода

- Можно задавать вопросы про неизвестные параметры:

$$P(\beta_3 > 0|y), P(\beta_3 = 0|y), E(\beta_3|y)?$$

- Апостериорное распределение есть всегда!

даже при жесткой мультиколлинеарности и полном отсутствии наблюдений

# Минусы байесовского подхода

- Его не все знают
- Может требовать больших объемов вычислений



“Идеальное прогнозирование” — ситуация, в которой ML оценки логит-модели не существуют

$$y_i \in \{0, 1\}.$$

$$y_i = \begin{cases} 1, & y_i^* \geq 0 \\ 0, & y_i^* < 0 \end{cases}$$

Скрытая переменная:  $y_i^* = \beta_1 + \beta_2 x_i + \varepsilon_i$ .

# Априорное распределение для логит модели

$$\beta \sim N(b_0, B_0^{-1})$$

Гиперпараметры:

$b_0$  — априорное среднее

$B_0$  — априорная матрица точности

$$B_0^{-1} = \text{Var}(\beta)$$

Традиционно:

$$b_0 = (0, 0, \dots, 0)'$$

$$B_0 = \begin{pmatrix} d & 0 & 0 & \dots \\ 0 & d & 0 & \dots \\ 0 & 0 & d & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Число  $d$  мало

То есть:  $\beta_1 \sim N(0, 1/d)$ ,  $\beta_2 \sim N(0, 1/d)$ ,  $\dots$

# Пример проблемной ситуации

y	x
0	1
0	2
1	3

Логит и пробит оценки не существуют

Априорно:  $\beta_1 \sim N(0, 10^2)$ ,  $\beta_2 \sim N(0, 10^2)$

Апостериорные средние:

$$\hat{y}_i^* = -10.8 + 4.5x_i$$

$$y_i = \begin{cases} 1, & y_i^* \geq 0 \\ 0, & y_i^* < 0 \end{cases}$$

Модель:  $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ ,  $\varepsilon_i \sim N(0, \sigma^2)$

# Вариант априорного распределения пик-плато

- $\beta_j | \gamma_j, \tau_j^2 \sim N(0, \gamma_j \cdot \tau_j^2)$
- $\gamma_j = \begin{cases} 1, & \text{с вероятностью } 1/2 \\ 0, & \text{с вероятностью } 1/2 \end{cases}$
- $\tau_j^2 \sim \Gamma^{-1}(a_1, a_2)$
- $\sigma^2 \sim \Gamma^{-1}(b_1, b_2)$

Гиперпараметры:  $a_1, a_2, b_1, b_2$



Позволяет напрямую отвечать на вопрос:  
Чему равна вероятность  $P(\beta_2 = 0|y)$ ?

# Пример с машинами

Апостериорные средние значения коэффициентов:

$$\widehat{dist}_i = 12.81 + 0.28speed_i + 0.01speed_i^2$$

Апостериорные вероятности:

$$P(\beta_{speed} = 0|y) = 0.15$$

$$P(\beta_{speed^2} = 0|y) = 0.05$$

Нам не удалось решить все наши задачи.

Решения, что мы находим, лишь ставят перед нами новые вопросы.

В каком-то смысле, мы также мало знаем, как и раньше. Но мы верим, что наше незнание стало глубже, а не знаем мы всё более важные вещи.

Большое спасибо тем, кто прошел вместе с нами этот курс до конца!