Обобщенные линейные модели

Малов Сергей Васильевич

Санкт-Петербургский государственный электротехнический университет

5 декабря 2020 г.

План

Введение

Мотивы появления обобщенных регрессионнных моделей

- Классическая регрессионная модель подразумевает выполнение основных предположений
  - Наблюдаемая характеристика имеет нормальное распределение
  - Наблюдения независимы (некоррелированы) и имеют одинаковую дисперсию
- Далеко не всегда классические предположения выполненяются
- Отказ от предположения независимости и одинаковой дисперсии наблюдений возможен, но подразумевает сохранение некоторой концепции зависимости
- Нарушение предположений о нормальности наблюдений – основная причина появления обобщенных моделей
- Это особенно актуально для дискретных наблюдаемых переменных, в частности, бинарных $Y \in \{0, 1\}$

Введение

Для бинарной наблюдаемой характеристики $Y \in [0, 1]$

$$\mathbb{E}_\theta(Y|z) = p_z = \mathbb{P}(Y = 1|z) \in [0, 1]$$

- В частности, использование классической простой регрессии требует ограничений на значения параметров
$$\mathbb{E}_\theta(Y|z) = \beta_1 + \beta_2 z \in [0, 1]$$
- Ошибки $\epsilon = Y - \beta_1 - \beta_2 z$ не могут быть корректно интерпретированы
- Дисперсии $\mathbb{D}(Y|z) = p_z(1 - p_z)$ различны и выражаются через регрессию p_z

Введение

Обобщенный подход

- Пусть $g : [0, 1] \to (-\infty, \infty)$
  - Логистическая регрессия
$$g(u) = \mathrm{logit}(u) = \log(u/(1-u))$$
  - Использовать функцию распределения, например, функцию нормального распределения $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$
$$g(u) = \mathrm{probit}(u) = \Phi^{-1}(u)$$
- Обобщенная (простая) регрессия
$$g(p_z) = \beta_1 + \beta_2 z \in (-\infty, \infty)$$
- Для модели логистической регрессии
$$p_z = \frac{1}{1 + e^{-\beta_1 - \beta_2 z}}$$

План

Обобщенная линейная модель

Регрессионное соотношение

$$g(\mathbb{E}_\theta(Y|\boldsymbol{X})) = \boldsymbol{X}^T\boldsymbol{\beta}$$

- g – функция связи (link)
- g – строго монотонная функция с областью значений $\mathbb{R}$
- Область определения функции связи совпадает с множеством допустимых значений $\mathbb{E}(Y|z)$
- Целесообразно выбирать функцию связи таким образом, чтобы каждому вещественному значению g соответствовало некоторое значение из области определения

Подразумевается, что распределение Y при условии z принадлежит некоторому параметрическому семейству распределений

- Обычно используют экспоненциальные семейства с плотностями (дискретными плотностями)

$$f(y;\theta,\phi) = \exp\Big((y\theta - b(\theta))/a(\phi) + c(y;\phi)\Big)$$

Статистические данные: $(Y_1, z_1), \ldots, (Y_n, z_n)$.

- Наиболее часто предполагается, что наблюдения независимы.

Экспоненциальные семейства

Абсолютно непрерывные экспоненциальные семейства

- Семейство нормальных распределений $\mathcal{N}(a, \sigma^2)$ $(a \in \mathbb{R},\ \sigma > 0)$

$$f(y; a, b) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(y-a)^2}{2\sigma^2}\right)$$

- Семейство Гамма распределений $\Gamma(p, b)$ $(p > 0, b > 0)$

$$f(y; p, b) = \frac{y^{p-1} e^{-y/b}}{b^p \Gamma(p)} \mathbb{I}_{\{y>0\}}$$

- Семейство Бета-распределений $\mathcal{B}(\alpha, \beta)$ $(p > 0, r > 0)$

$$f(y; p, r) = \frac{\Gamma(p+r)}{\Gamma(p)\Gamma(r)} y^{r-1} (1-y)^{r-1} \mathbb{I}_{\{y \in (0,1)\}}$$

Экспоненциальные семейства

Экспоненциальные семейства дискретных распределений

- Семейство Биномиальных распределений $\mathrm{Bi}_m(p)$ $(p \in (0,1))$

$$f^*(y;a,b) = \frac{m!}{y!(m-y)!}p^y(1-p)^{m-y},\ y = 0,\dots,m$$

- Семейство распределений Пуассона $\mathrm{P}(\lambda)$ $(\lambda > 0)$

$$f^*(y;a,b) = \frac{\lambda^y}{y!}e^{-\lambda},\ y = 0,1,\dots$$

- Семейство отрицательных биномиальных распределений $\mathrm{Nb}_m(p)$ $(p \in (0,1))$

$$f^*(y;p) = \frac{(m+y-1)!}{y!(m-1)!}p^y(1-p)^m,\ y = 0,1,\dots$$

- Семейство Мультиномиальных распределений $\mathrm{Mult}_{m,k}(p_1,\dots,p_k)$ $(p_i \in [0,1], \sum_i p_i = 1)$

$$f^*(y_1,\dots,y_k;p,b) = \frac{m!}{y_1!\cdots y_m!}p_1^{y_1}\cdots p_k^{y_k},\ y_i = 0,1,\dots;\sum_i y_i = m.$$

Экспоненциальные семейства

Логарифм функции правдоподобия: $LL(\boldsymbol{Y};\boldsymbol{\theta},\phi)=\sum_{i=1}^{n}\ln L(Y_i;\theta_i;\phi)$

- Исследуем

$$\ln L_i=\ln L(Y_i;\theta_i;\phi)=(Y_i\theta_i-b(\theta_i))/a(\phi)+c(Y_i;\phi)$$

- Производные по θ_i равны:

$$\frac{\partial\ln L_i}{\partial\theta_i}=(Y_i-b'(\theta_i))/a(\phi),\qquad \frac{\partial^2\ln L_i}{\partial\theta_i^2}=-b''(\theta_i)/a(\phi)$$

- Исходя из $\mathbb{E}\left(\frac{\partial\ln L_i}{\partial\theta_i}\right)=0$ заключаем, что

$$\mu_i=\mathbb{E}(Y_i)=\mathbb{E}(Y|\theta_i,\phi)=b'(\theta_i)$$

- С использованием равенства

$$\mathbb{E}\left(\frac{\partial^2\ln L_i}{\partial\theta_i^2}\right)=-\mathbb{E}\left(\frac{\partial\ln L_i}{\partial\theta_i}\right)^2,$$

получаем, что

$$\mathbb{D}(Y_i)=\mathbb{D}(Y|\theta_i,\phi)=b''(\theta_i)\,a(\phi)$$

Экспоненциальные семейства

Логарифм функции правдоподобия: $LL(\boldsymbol{Y};\boldsymbol{\theta},\phi) = \sum_{i=1}^{n} \ln L(Y_i;\theta_i;\phi)$

- Исследуем

$$\ln L_i = \ln L(Y_i;\theta_i;\phi) = (Y_i\theta_i - b(\theta_i))/a(\phi) + c(Y_i;\phi)$$

- Производные по θ_i равны:

$$\frac{\partial \ln L_i}{\partial\theta_i} = (Y_i - b'(\theta_i))/a(\phi), \qquad \frac{\partial^2 \ln L_i}{\partial\theta_i^2} = -b''(\theta_i)/a(\phi)$$

- Исходя из $\mathbb{E}\left(\frac{\partial \ln L_i}{\partial\theta_i}\right) = 0$ заключаем, что

$$\mu_i = \mathbb{E}(Y_i) = \mathbb{E}(Y|\theta_i,\phi) = b'(\theta_i)$$

- С использованием равенства

$$\mathbb{E}\left(\frac{\partial^2 \ln L_i}{\partial\theta_i^2}\right) = -\mathbb{E}\left(\frac{\partial \ln L_i}{\partial\theta_i}\right)^2$$

получаем, что

$$\mathbb{D}(Y_i) = \mathbb{D}(Y|\theta_i,\phi) = b''(\theta_i)\,a(\phi)$$

Экспоненциальные семейства и GLM

Оценивание

- Модель: $g(\mu(\boldsymbol{X}_i)) = g(\mu_i) = \eta_i = \sum_{s=1}^{m} x_{si}\beta_s$
  - $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)^T$ – матрица плана
  - $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ – параметр экспоненциального семейства
  - $\mu(\boldsymbol{X}) = (\mu_1, \ldots, \mu_n)^T = \mathbb{E}(Y|\boldsymbol{X})$ – регрессия
- Функцию связи g будем называть *канонической*, если $\eta_i = \theta_i$.
- Оценка – решение системы уравнений

$$\frac{\partial LL(\boldsymbol{\beta})}{\partial \beta_l} = \sum_{i=1}^{n} \frac{\partial LL_i(\theta_i, \phi)}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_l} = 0,$$

  где
  - $LL_i(\theta, \phi) = \ln L(Y_i; \theta, \phi)$
  - $\frac{\partial LL_i(\theta_i,\phi)}{\partial \theta_i} = \frac{Y_i - b'(\theta_i)}{a(\phi)} = \frac{Y_i - \mu_i}{a(\phi)}$,
  - $\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) = \mathbb{D}(Y_i)/a(\phi)$,
  - $\frac{\partial \eta_i}{\partial \beta_l} = x_{il}$
  - $\frac{\partial \mu_i}{\partial \eta_i} = (g^{-1}(\eta_i))'$
- Для решения системы используют численный метод Ньютона–Рафсона

Оценивание в каноническом случае

- В каноническом случае параметр $\boldsymbol{\beta}$ не входит в левую часть системы, а зависимость от параметра $\boldsymbol{\beta}$ реализуется через $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$
- Система уравнений в каноническом случае

$$\sum_{i=1}^{n} \frac{\partial \ln L(Y_i, \theta_i, \phi)}{\partial \theta_i} \frac{\partial \eta_i}{\partial \beta_l} = \sum_{i=1}^{n} \frac{(Y_i - \mu_i) x_{il}}{a(\phi)} = 0$$

- В матричной форме получаем

$$\mathbf{X}\boldsymbol{Y} = \mathbf{X}\boldsymbol{\mu}$$

- Если строки матрицы $\mathbf{X}$ линейно независимы, то $\hat{\boldsymbol{\beta}}$ – коэффициенты разложения $\hat{\boldsymbol{\eta}} = g(\hat{\boldsymbol{\mu}})$ по базису строк $\mathbf{X}$
- В частном случае семейства нормальных распределений получаем сразу систему нормальных уравнений

$$\mathbf{X}\boldsymbol{Y} = \mathbf{X}\mathbf{X}'\boldsymbol{\beta}$$

- В общем случае следует избегать матриц плана неполного ранга

План

Асимптотические свойства оценок

При выполнении условий регулярности оценка максимального правдоподобия $\hat{\boldsymbol{\beta}}$ асимптотически нормальна

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow \mathcal{N}\left(0, \bar{\mathbb{I}}^{-1}(\boldsymbol{\beta})\right)$$

- $\bar{\mathbb{I}}(\boldsymbol{\beta}) = \lim_{n\to\infty} \mathbb{I}(\boldsymbol{\beta})/n$, где $\mathbb{I}(\boldsymbol{\beta})$ – матрица информации Фишера
- $\mathbb{I}(\boldsymbol{\beta}) = -\mathbb{E}(LL''(\boldsymbol{\beta})) = -\left\|\mathbb{E}\left(\frac{\partial^2 LL(\boldsymbol{\beta})}{\partial\beta_i\partial\beta_j}\right)\right\|_{i,j=1}^m = \left\|\mathbb{E}\left(\frac{\partial LL(\boldsymbol{\beta})}{\partial\beta_i}\frac{\partial LL(\boldsymbol{\beta})}{\partial\beta_j}\right)\right\|_{i,j=1}^m$
- Для экспоненциального семейства $\mathbb{I}(\boldsymbol{\beta}) = \mathbf{X}^T\mathbf{W}\mathbf{X}$
  - $\mathbf{W} = \mathbf{W}(\boldsymbol{\beta}, \phi)$ — диагональная матрица
  - диагональные элементы $1/(\mathbb{D}(Y_r)g'(\mu_r)^2)$, $r = 1, \ldots, m$
- Оценка $\mathbb{I}$: $\hat{\mathbb{I}} = \mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}$, где $\hat{\mathbf{W}} = \mathbf{W}(\hat{\boldsymbol{\beta}}, \hat{\phi})$
- В случае канонической функции связи

$$\mathbb{I}(\beta) = \left\|\sum_{r=1}^n \frac{x_{ri}x_{rj}}{a(\phi)g'(\mu_r)}\right\|_{i,j=1}^m = \left\|\sum_{r=1}^n \frac{b''(\theta_r)x_{ri}x_{rj}}{a(\phi)}\right\|_{i,j=1}^m$$

Доверительное оценивание

Пусть $\boldsymbol{\psi} = \mathbf{C}'\boldsymbol{\beta}$ – функция параметра, $\mathbf{C}$ – $m \times q$ матрица ранга q

- В качестве оценки $\boldsymbol{\psi}$ используем $\hat{\boldsymbol{\psi}} = \mathbf{C}'\hat{\boldsymbol{\beta}}$
- Асимптотическая нормальность $\hat{\boldsymbol{\beta}}$ влечет асимптотическую нормальность
$$\sqrt{n}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \Rightarrow \mathcal{N}(0, \Gamma_{\hat{\psi}})$$
    - $\Gamma_{\hat{\psi}} = \mathbf{C}'\bar{\mathbb{I}}^{-1}\mathbf{C}$ – предельная матрица ковариации $\hat{\boldsymbol{\psi}}$
    - $\hat{\Gamma}_{\hat{\psi}} = n\,\mathbf{C}'(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{C}$ – оценка $\Gamma_{\hat{\psi}}$
- Таким образом,
$$(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi})'\hat{\mathbf{B}}^{-1}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \Rightarrow \chi^2_q$$
    - $\hat{\mathbf{B}} = \mathbf{C}'(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{C}$
- Доверительный (асимптотически) эллипсоид
$$\left\{\boldsymbol{\psi} : (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi})'\hat{\mathbf{B}}^{-1}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \le x_\alpha\right\}$$
    - x_α – квантиль χ^2_q-распределения порядка $1 - \alpha$
- При $q = 1$ получаем асимптотический доверительный интервал

Проверка гипотез

Пусть $\boldsymbol{\psi} = \mathbf{C}'\boldsymbol{\beta}$ – функция параметра, $\mathbf{C}$ – $m \times q$ матрица ранга q

- Статистическая гипотеза: $H_0 : \boldsymbol{\psi} = 0$
- Критерий типа Вальда
    - Статистика критерия:
    $$Z = \hat{\boldsymbol{\psi}}'\hat{\mathbf{B}}^{-1}\hat{\boldsymbol{\psi}}$$
    - Асимптотическое распределение при нулевой гипотезе: χ^2_q
- Критерий отношения правдоподобия
    - Статистика критерия:
    $$G = 2LL(Y; \hat{\boldsymbol{\theta}}, \hat{\phi}) - 2LL(Y; \hat{\boldsymbol{\theta}}_H, \hat{\phi}_H),$$
    где $\hat{\boldsymbol{\theta}}_H, \hat{\phi}_H$ – ОМП при ограничении $\mathbf{C}'\boldsymbol{\beta} = 0$
    - Асимптотическое распределение при нулевой гипотезе: χ^2_q

План

Абсолютно-непрерывные модели

- Нормальное распределение
  - Функция правдоподобия:
    $L(y;a,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}}\exp\Big(-\frac{(y-a)^2}{2\sigma^2}\Big) = \exp\Big(\frac{ay-a^2/2}{\sigma^2} - \big(\frac{y^2}{2\sigma^2} + \frac{\log(2\pi\sigma)}{2}\big)\Big)$,
  - Моменты: $\mathbb{E}_\theta Y = a$; $\mathbb{D}_\theta Y = \sigma^2$
  - Каноническая функция связи: $g(\mu) = \mu$
  - Компонент информационной матрицы (канонический)
    $\mathbf{W} = \|w_{ij}\|_{i,j}$: $w_{ij} = \sigma^{-2}\mathbb{1}_{\{i=j\}}$
- Гамма распределение:
  - Функция правдоподобия:
    $L(y;a,\sigma^2) = \frac{x^{p-1}a^p\exp(-ay)}{\Gamma(p)} = \exp\Big(\frac{ay-p\log a}{-1} + (p-1)\log x - \log\Gamma(p)\Big)$,
  - Моменты: $\mathbb{E}_\theta Y = p/a$; $\mathbb{D}_\theta Y = p/a^2$
  - Каноническая функция связи: $g(\mu) = p/\mu$ (непригодна)
  - Допустимая функция связи: $g(\mu) = \log\mu$
  - Компонент информационной матрицы $\mathbf{W} = \|w_{ij}\|_{i,j}$:
    $w_{ij} = p\mathbb{1}_{\{i=j\}}$

Дискретные модели

- Распределение Бернулли
  - Функция правдоподобия:
    $L(y;p) = p^y(1-p)^{1-y} = \exp(y\log(p/(1-p)) - \log(1-p)),\ y = 0,1$
  - Моменты: $\mathbb{E}_p Y = p$; $\mathbb{D}_p Y = p(1-p)$
  - Канонический параметр: $\theta = \log(p/(1-p))$
  - Каноническая функция связи: $g(\mu) = \log(\mu/(1-\mu))$
  - Компонент информационной матрицы (канонический)
    $\mathbf{W} = \|w_{ij}\|_{i,j}$: $w_{ij} = p_i(1-p_i)\mathbb{1}_{\{i=j\}}$
- Распределение Пуассона
  - Функция правдоподобия:
    $L(y;\lambda) = \lambda^y e^{-\lambda}/y! = \exp(y\log(\lambda) - \lambda - \log(y!)),\ y = 0,1,\ldots$
  - Моменты: $\mathbb{E}_\lambda Y = \lambda$; $\mathbb{D}_\lambda Y = \lambda$
  - Канонический параметр: $\theta = \log(\lambda)$
  - Каноническая функция связи: $g(\mu) = \log(\mu)$
  - Компонент информационной матрицы (канонический)
    $\mathbf{W} = \|w_{ij}\|_{i,j}$: $w_{ij} = \lambda_i\mathbb{1}_{\{i=j\}}$

Обобщенные линейные модели

Дискретные модели: обобщенное распределение Пуассона (Conway–Maxwell-Poisson)

- Применяется при наличии избыточной или недостточной дисперсии (overdispersion,underdispersion)
- Функция правдоподобия:
  $L(y;\lambda,\nu) = \frac{\lambda^y}{C(\lambda,\nu)(y!)^\nu} = \exp\big(y\log\lambda - \log C(\lambda,\nu) - \nu\log(y!)\big)$
  - $y = 0, 1, \dots$
  - $C(\lambda,\nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$
  - при $\nu = 1$ – распределение Пуассона $\mathrm{Pois}(\lambda)$
  - при $\nu \to \infty$ – распределение Бернулли с параметром $p = \lambda/(1+\lambda)$
  - при $\nu \to 0_+ (\lambda < 1)$ – геометрическое распределение: $p = 1 - \lambda$
- Моменты: $\mathbb{E}_p Y = \lambda \frac{C'_\lambda(\lambda,\nu)}{C(\lambda,\nu)}$; $\mathbb{D}_p Y = \lambda \frac{C'_\lambda(\lambda,\nu)}{C(\lambda,\nu)} + \lambda^2 \left(\frac{C''_{\lambda\lambda}(\lambda,\nu)}{C(\lambda,\nu)} - \frac{C'_\lambda(\lambda,\nu)^2}{C(\lambda,\nu)^2} \right)$
  - $C'_\lambda(\lambda,\nu) = \sum_{j=1}^{\infty} \frac{j\lambda^{j-1}}{(j!)^\nu}$
  - $C''_{\lambda\lambda}(\lambda,\nu) = \sum_{j=2}^{\infty} \frac{j(j-1)\lambda^{j-2}}{(j!)^\nu}$
- функция связи: $g(\mu) = \log\mu$
- Компонент информационной матрицы $\mathbf{W} = \|w_{ij}\|_{i,j}$:

$$w_{ij} = \frac{\lambda C'_\lambda(\lambda,\nu)}{C(\lambda,\nu)C'_\lambda(\lambda,\nu) + \lambda(C''_{\lambda\lambda}(\lambda,\nu)C(\lambda,\nu) + C'_\lambda(\lambda,\nu)^2)} \mathbb{I}_{\{i=j\}}$$

Дискретные модели

- Геометрическое распределение
  - Функция правдоподобия: $L(y;\lambda) = p^y(1-p) = \exp(y\log p + \log(1-p)),\ y = 0,1,\ldots$
  - Моменты: $\mathbb{E}_p Y = p/(1-p)$; $\mathbb{D}_p Y = p/(1-p)^2$
  - Канонический параметр: $\theta = \log p \in (-\infty, 0)$
  - Функция связи: $g(\mu) = \log\mu$ $(= \mathrm{logit}(p))$
  - Компонент информационной матрицы $\mathbf{W} = \|w_{ij}\|_{i,j}$: $w_{ij} = p_i \mathbb{1}_{\{i=j\}}$

Не экспоненциальные семейства

- Отрицательное биномиальное распределение
  - Функция правдоподобия: $L(y;\alpha,\lambda) = \frac{\Gamma(\alpha+y)}{y!\Gamma(\alpha)} \frac{\alpha^\alpha \lambda^y}{(\lambda+\alpha)^{\alpha+y}}$
  - Моменты: $\mathbb{E}_\theta Y = \lambda$; $\mathbb{D}_\theta Y = \lambda(1+\lambda/\alpha)$ (overdispersion)
  - При $\alpha \to \infty$ получаем семейство распределений Пуассона с параметром λ
  - Функция связи: $g(\mu) = \log\mu$
  - Компонент информационной матрицы $\mathbf{W} = \|w_{ij}\|_{i,j}$: $w_{ij} = \frac{\lambda_i \alpha}{\alpha+\lambda_i} \mathbb{1}_{\{i=j\}}$

План

Выбор оптимальной модели

Выбор (обобщенной) линейной модели для проведения дальнейшего количественного исследования – одна из ключевых задач регрессионного анализа

- Излишнее упрощение модели может не позволить достигнуть целей исследования
  - модель должна соответствовать истинному положению дел
- В случае усложнения модели (при наличии большого числа сопутствующих факторов)
  - недостаток статистических данных не позволяет делать статистические выводы
  - возникают сложности с интерпретацией результатов анализа
- Каждая модель ассоциируется с некоторой гипотезой H
  - при наличии наиболее общей модели $H : \mathbf{C}'\boldsymbol{\beta} = 0$
    - $\boldsymbol{\beta}$ – параметр общей модели
  - наличие наиболее общей модели не является обязательным
- Из набора допустимых моделей $H_1, \ldots, H_k$ требуется выбрать оптимальную исходя из имеющегося набора статистических данных и целей исследования

Выбор оптимальной модели

Методы исключения и включения параметров в модель

- Метод исключения
  - исследование начинается с наиболее общей модели (saturated)
  - наименее значимые параметры последовавтельно исключаются из модели
  - порядок исключения определяет характер модели
    - в модели дисперсионного анализа в первую очередь исключаются взаимодействия высоких порядков
    - в полиномиальной модели в первую очередь исключаются более высокие степени
  - процедура исключения останавливается, если все входящие в модель параметры (классы парметров) являются значимыми
- Метод включения
  - исследование начинается с наиболее простой модели (intercept only)
  - последовательно добавляют параметры (группы параметров) в модель согласно выбранному алгоритму
  - признаки, для которых не выявляется значимое влияние на результат, в модель не включаются
  - может привести к потере признаков, значимо влияющих на результат на уровне взаимодействий

Выбор оптимальной модели

Информационные критерии

- Решение принимается с использованием логарифма функции правдоподобия LL, размерности параметра и количества статистической информации
- Целевая функция $IC_R(\boldsymbol{X}; H) = -2\sup_H LL(\boldsymbol{X}, \theta) + 2R(H; \boldsymbol{X})$,
  - $R(H; \boldsymbol{X})$ — неотрицательная функция пенализации, зависящая от размерности параметра в предположении H, от числа и характера наблюдений
  - оптимальной считается модель H, для которой IC_R принимает наименьшее значение

Наиболее часто используют

- Критерий Акайке (1973): $AIC(\boldsymbol{X}; H) = IC_R(\boldsymbol{X}; H)$
  - $R(H; \boldsymbol{X}) = \dim(\Theta_H)$ — размерность параметра θ в предположении H
- Байесовский информационный критерий (Шварц, 1978): $BIC(\boldsymbol{X}; H) = IC_R(\boldsymbol{X}; H)$
  - $R(H; \boldsymbol{X}) = \dim(\Theta_H) \log n/2$ — определяется размерностью параметра θ в предположении H и числом наблюдений
- Байесовский критерий дает большую пенализацию на размерность параметра при больших n, чем критерий Акайке