Теорема

Определение. Назовем *обобщенно-линейной моделью с натуральной функцией связи* и априорным распределением на вектор параметров $p(\mathbf{w}|\mathbf{A})$ вероятностную модель, совместное правдоподобие которой имеет вид

$$p(\mathbf{y},\mathbf{w}|\mathbf{X},\mathbf{A}) = p(\mathbf{y}|\mathbf{X},\mathbf{w})p(\mathbf{w}|\mathbf{A}),\ \text{где}\ p(\mathbf{y}|\mathbf{X},\mathbf{w}) = \prod_{i=1}^{m} p(y_i|\mathbf{x}_i,\mathbf{w})$$

$$p(y_i|\mathbf{x}_i,\mathbf{w}) = c(y_i)\exp(\theta_i y_i - b(\theta_i)),\ \text{где}\ \theta_i = \mathbf{w}^T\mathbf{x}_i$$

Замечание. Отметим, что логистическая модель является обобщенно-линейной моделью с натуральной функцией связи для бинарной целевой переменной $y_i \in \{-1, 1\}$, так как

$$p(yi|\mathbf{x}_i,\mathbf{w}) = \sigma(y_i\mathbf{x}_i^T\mathbf{w}) = \frac{\exp(y_i\mathbf{x}_i^T\mathbf{w})}{\exp(\mathbf{w}^T\mathbf{x}_i) + \exp(-\mathbf{w}^T\mathbf{x}_i)}.$$

Тогда, взяв $(y_i) = 1, b(\theta_i) = \exp(\theta_i) + \exp(-\theta_i)$, получим требуемое.

Введем матрицу вторых производных $\mathbf{H}_m(\mathbf{w})$:

$$\mathbf{H}_m(\mathbf{w}) = \sum_{i=1}^{m} \mathbf{x}_i \frac{\partial^2 b}{\partial\theta^2}(\mathbf{w}^T\mathbf{x}_i)\mathbf{x}_i^T.$$

$\widetilde{\mathbf{H}}(\widetilde{\mathbf{v}}) = \mathbf{H}(\widetilde{\mathbf{v}}) + \mathbf{A},$ где $\mathbf{A}$ - матрица корелляции для априорного распределения весов.

Теорема 14 (Адуенко (2016))**.** Пусть имеется обобщенно-линейная модель с натуральной функцией связи с истинным вектором параметров $\mathbf{w}$ и нормальным априорным распределением на вектор параметров $\mathbf{w} \sim \mathcal{N}(\mathbf{w}|\mathbf{0},\ \mathbf{A}_m^{-1})$, причем $\exists\, c_0 < \infty :\ \|\mathbf{A}_m\| < c_0\ \forall\, m$. Пусть также $\sum_{i=1}^{m}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}$ имеет полный ранг для $m \geq m_0$, $\lambda_{\min}(\mathbf{H}(\mathbf{w})) \to \infty$ при $m \to \infty$ и

$$\forall\,\delta > 0\ \max_{\mathbf{v}\in O_m^{\delta}(\mathbf{w})} \|\mathbf{H}(\mathbf{w})^{-1/2}\mathbf{H}(\mathbf{v})\mathbf{H}(\mathbf{w})^{-T/2} - \mathbf{I}\| \to 0 \text{ при } m \to \infty.$$

Тогда $\tilde{\mathbf{H}}(\hat{\mathbf{w}})^{1/2}(\hat{\mathbf{w}} - \mathbf{w}) \xrightarrow{d} \mathcal{N}(\mathbf{0},\ \mathbf{I}_n)$, где $\hat{\mathbf{w}} = \arg\max_{\mathbf{v}} p(\mathbf{y},\ \mathbf{v}|\mathbf{X},\ \mathbf{A})$ есть оценка максимума апостериорной вероятности для вектора параметров модели, а $\mathbf{H}(\hat{\mathbf{w}})^{1/2}$ есть корень из матрицы в терминах разложениях Холецкого [3].

Теорема 15 (Адуенко (2016)). Пусть имеется две обобщенно-линейные модели с натуральной функцией связи с одинаковым истинным вектором параметров $\mathbf{w}$ и нормальным априорным распределением на вектор параметров $\mathbf{w} \sim \mathcal{N}(\mathbf{w}|\mathbf{0},\ \mathbf{A}_{m^k}^{-1})$, $k = 1,\ 2$, причем $\exists\ c_0^k < \infty:\ \ \|\mathbf{A}_{m^k}^k\| < c_0^k\ \ \forall\ m^k,\ \ k = 1,\ 2$. Пусть также значение вектора параметров $\mathbf{w}$ для второй модели известно точно. Пусть для первой модели выполнено, что $\sum_{i=1}^{m^1} \mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}$ имеет полный ранг для $m^1 \geq m_0$, $\lambda_{\min}(\mathbf{H}_{m^1}(\mathbf{w})) \to \infty$ при $m^1 \to \infty$ и

$$\forall\,\delta > 0\ \max_{\mathbf{v}\in O^{\delta}_{m^1}(\mathbf{w})} \|\mathbf{H}_{m^1}(\mathbf{w})^{-1/2}\mathbf{H}_{m^1}(\mathbf{v})\mathbf{H}_{m^1}(\mathbf{w})^{-T/2} - \mathbf{I}\| \to 0 \ \text{ при }\ m^1 \to \infty.$$

Тогда $-2\log \text{s-score} = (\hat{\mathbf{w}} - \mathbf{w})^{\mathsf{T}}\tilde{\mathbf{H}}_{m^1}(\hat{\mathbf{w}})(\hat{\mathbf{w}} - \mathbf{w}) \xrightarrow{d} \chi^2(n)$, где $\hat{\mathbf{w}}$ есть оценка максимума апостериорной вероятности для вектора параметров модели(см. (3.1) для $\mathbf{A} = \mathbf{A}_{m^1}^1$), $\mathbf{H}_{m^1}(\hat{\mathbf{w}})^{1/2}$ есть корень из матрицы в терминах разложениях Холецкого [3], а m^1 – число объектов в выборке для первой модели.

1 Теорема

(Адуенко (2016)). Пусть имеется две обобщенно-линейные модели с натуральной функцией связи с одинаковым истинным вектором параметров $\mathbf{w}$ и нормальным априорным распределением на вектор параметров $\mathbf{w} \sim \mathcal{N}(\mathbf{w}|\mathbf{0}, \mathbf{A}_{m^k}^{-1})$, $k = 1, 2$, причем $\exists\ c_0^k < \infty$: $\|A_{m^k}^k\| < c_0^k\ \forall\ m^k, k = 1, 2$. Пусть для обеих моделей выполнено, что $\sum_{i=1}^{m^k} \mathbf{x}_i\mathbf{x}_i^T$ имеет полный ранг для $m^k \geqslant m_0^k, k = 1, 2$, $\lambda_{\min}(\mathbf{H}_{m^k}(\mathbf{w})) \to \infty$ при $m^k \to \infty, k = 1, 2$ и

$$\forall \delta > 0 \max_{\mathbf{v}\in O^{\delta}_{m^k}(\mathbf{w})} \|\mathbf{H}_{m^k}(w)^{-1/2}\mathbf{H}_{m^k}(v)\mathbf{H}_{m^k}(w)^{-T/2} - \mathbf{I}\| \to 0 \text{ при } m^k \to \infty, k = 1, 2.$$

Пусть также

$$\|\widetilde{\mathbf{H}}_{m^1}(\hat{\mathbf{w}}_1)\|\|\widetilde{\mathbf{H}}_{m^2}^{-1}(\hat{\mathbf{w}}_2)\| \to 0 \text{ при } m = \min(m^1, m^2) \to \infty.$$

Тогда $-2\log \mathrm{s} - \mathrm{score} = (\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}_1)^T(\widetilde{\mathbf{H}}_{m^1}^{-1}(\hat{\mathbf{w}}_1) + \widetilde{\mathbf{H}}_{m^2}^{-1}(\hat{\mathbf{w}}_2))^{-1}(\hat{w}_2 - \hat{w}_1) \xrightarrow{d} \chi^2(n)$, где $\hat{\mathbf{w}}_k, k = 1, 2$ есть оценки максимума апостериорной вероятности для вектора параметров модели $k = 1, 2$, $\widetilde{\mathbf{H}}_{m^k}(\hat{\mathbf{w}}_k)^{1/2}$ есть корень из матрицы в терминах разложениях Холецкого, а m^k число объектов в выборке для модели $k = 1, 2$.

Доказательство. Обозначим $\widetilde{\mathbf{H}}_1 = \widetilde{\mathbf{H}}_{m^1}(\hat{\mathbf{w}}_1), \widetilde{\mathbf{H}}_2 = \widetilde{\mathbf{H}}_{m^2}(\hat{\mathbf{w}}_2)$. Зафиксируем произвольные $1 > \eta > 0$ и $1 > \varepsilon n > 0$. Из условия теоремы имеем, что

$$\exists\ m' :\ \forall\ m \geqslant m'\ \mathbb{P}(\|\widetilde{\mathbf{H}}_1\|\|\widetilde{\mathbf{H}}_2\| \leqslant \varepsilon) \geqslant 1 - \eta.$$

Тогда с учетом $\varepsilon < 1$ с вероятностью $\geqslant 1 - \eta$ имеем

$$-2\log \mathrm{s} - \mathrm{score} = (\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}_1)^T\widetilde{\mathbf{H}}_1(\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}_1) + (\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}_1)^T\widetilde{\mathbf{H}}_1\sum_{l=1}^{\infty}(-1)^l(\widetilde{\mathbf{H}}_2^{-1}\widetilde{\mathbf{H}}_1)^l(\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}_1).$$

С учетом
$$\left|(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)^T\widetilde{\mathbf{H}}_1\sum_{l=1}^{\infty}(-1)^l(\widetilde{\mathbf{H}}_2^{-1}\widetilde{\mathbf{H}}_1)^l(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)\right|\leqslant$$

$$\sum_{l=1}^{\infty}(-1)^l(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)^T\widetilde{\mathbf{H}}_1(\widetilde{\mathbf{H}}_2^{-1}\widetilde{\mathbf{H}}_1)^l(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)=$$

$$\sum_{l=1}^{\infty}(-1)^l tr\Big((\widetilde{\mathbf{H}}_2^{-1}\widetilde{\mathbf{H}}_1)^l(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)^T\widetilde{\mathbf{H}}_1\Big)\leqslant$$

$$\sum_{l=1}^{\infty}(-1)^l tr\Big((\widetilde{\mathbf{H}}_2^{-1}\widetilde{\mathbf{H}}_1)^l\Big)(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)^T\widetilde{\mathbf{H}}_1(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)\leqslant\sum_{l=1}^{\infty}(-1)^l(n\varepsilon)^l(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)^T\widetilde{\mathbf{H}}_1(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)=$$
$$\tfrac{n\varepsilon}{1-n\varepsilon}(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)^T\widetilde{\mathbf{H}}_1(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1).$$

Получаем, что

$$\mathbb{P}\left(\Big|-2\log\mathrm{s}-\mathrm{score}/(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)^T\widetilde{\mathbf{H}}_1(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)-1\Big|\leqslant\frac{n\varepsilon}{1-n\varepsilon}\right)\geqslant 1-\eta,$$

откуда в силу произвольности ε и η получим

$$-2\log\mathrm{s}-\mathrm{score}/(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)^T\widetilde{\mathbf{H}}_1(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)\xrightarrow{\mathbb{P}}1. \tag{1}$$

$$(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)^T\widetilde{\mathbf{H}}_1(\hat{\mathbf{w}}_2-\hat{\mathbf{w}}_1)=(\hat{\mathbf{w}}_2-\hat{\mathbf{w}})^T\widetilde{\mathbf{H}}_1(\hat{\mathbf{w}}_2-\hat{\mathbf{w}})+(\hat{\mathbf{w}}_1-\hat{\mathbf{w}})^T\widetilde{\mathbf{H}}_1(\hat{\mathbf{w}}_1-\hat{\mathbf{w}})+2(\hat{\mathbf{w}}_2-\hat{\mathbf{w}})^T\widetilde{\mathbf{H}}_1(\hat{\mathbf{w}}_1-\hat{\mathbf{w}}) \tag{2}$$

В условиях теоремы из теоремы 15 получаем, что

$$(\hat{\mathbf{w}}_k-\hat{\mathbf{w}})^T\widetilde{\mathbf{H}}_k(\hat{\mathbf{w}}_k-\hat{\mathbf{w}})\xrightarrow{d}\chi^2(n)\text{ при }m\to\infty,\ k=1,2 \tag{3}$$

Покажем, что

$$\|\widetilde{\mathbf{H}}_1\|\|\hat{\mathbf{w}}_2-\hat{\mathbf{w}}\|^2\ \xrightarrow{\mathbb{P}}0, \tag{4}$$

Из (3) заключаем, что

$$\forall\ \varepsilon>0\ \exists\ m_\varepsilon,\ \exists\ c_\varepsilon:\ \forall\ m^k\geqslant m_\varepsilon\mathbb{P}((\hat{\mathbf{w}}_k-\hat{\mathbf{w}})^T\widetilde{\mathbf{H}}_k(\hat{\mathbf{w}}_k-\hat{\mathbf{w}})>c_\varepsilon)<\varepsilon.$$

Тогда для модели $k=2$ получаем с учетом

$$(\hat{\mathbf{w}}_2-\hat{\mathbf{w}})^T\widetilde{\mathbf{H}}_2(\hat{\mathbf{w}}_2-\hat{\mathbf{w}})\geqslant\lambda_{\min}(\widetilde{\mathbf{H}}_2)\|\hat{\mathbf{w}}_2-\hat{\mathbf{w}}\|^2$$

$$\forall\ \varepsilon>0\ \exists\ m_\varepsilon,\ \exists\ c_\varepsilon:\ \forall\ m^k\geqslant m_\varepsilon\mathbb{P}(\lambda_{\min}(\widetilde{\mathbf{H}}_2)\|\hat{\mathbf{w}}_2-\hat{\mathbf{w}}\|^2>c_\varepsilon)<\varepsilon.$$

В условиях теоремы из сходимости

$$\|\widetilde{\mathbf{H}}_1\|\|\widetilde{\mathbf{H}}_2^{-1}\|\xrightarrow{\mathbb{P}}0\text{ при }m\to 0$$

получаем, что для любого фиксированного c_ε

$$\forall\, \eta > 0\ \forall\, \delta\ \exists\, m_\delta:\ \forall\, m \geqslant m_\delta\ \mathbb{P}\left(\frac{\lambda_{\max}(\widetilde{\mathbf{H}}_1)}{\lambda_{\min}(\widetilde{\mathbf{H}}_2)} < \frac{\delta}{c_\varepsilon}\right) > 1-\eta.$$

Тогда получаем, что

$$\forall\, \varepsilon > 0\ \forall\, \eta > 0\ \forall\, \delta > 0\ \exists\, m'' = \max(m_\varepsilon, m_\delta):\ \forall\, m \geqslant m''\ \mathbb{P}(\|\widetilde{\mathbf{H}}_1\|\|\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}\|^2 < \delta) > 1-\varepsilon-\eta.$$

что как раз и означает требуемое (4). Покажем теперь, что

$$(\hat{\mathbf{w}}_2 - \hat{\mathbf{w}})^T \widetilde{\mathbf{H}}_1 (\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}) \xrightarrow{\mathbb{P}} 0 \text{ при } m \to \infty \tag{5}$$

$$(\hat{\mathbf{w}}_1 - \hat{\mathbf{w}})^T \widetilde{\mathbf{H}}_1 (\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}) \xrightarrow{\mathbb{P}} 0 \text{ при } m \to \infty \tag{6}$$

Отметим, что (5) непосредственно следует из (4), так как

$$\hat{\mathbf{w}}_2 - \hat{\mathbf{w}})^T \widetilde{\mathbf{H}}_1 (\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}) \leqslant \|\widetilde{\mathbf{H}}_1\| \|\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}\|^2.$$

Для доказательства (6) заметим, что

$$\hat{\mathbf{w}}_1 - \hat{\mathbf{w}})^T \widetilde{\mathbf{H}}_1 (\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}) = \hat{\mathbf{w}}_2 - \hat{\mathbf{w}})^T \widetilde{\mathbf{H}}_1^{1/2} \widetilde{\mathbf{H}}_1^{T/2} (\hat{\mathbf{w}}_1 - \hat{\mathbf{w}}).$$

Из условий теоремы в силу теоремы 14 имеем

$$\widetilde{\mathbf{H}}_1^{T/2} (\hat{\mathbf{w}}_1 - \hat{\mathbf{w}}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{I}_n) \text{ при } m \to \infty$$

и в силу (4)

$$(\hat{\mathbf{w}}_2 - \hat{\mathbf{w}})^T \widetilde{\mathbf{H}}_1^{1/2} \xrightarrow{\mathbb{P}} 0 \text{ при } m \to \infty$$

Отсюда из теоремы Слуцкого имеем (6)

Таким образом, из (1), (2) и сходимости к нулю по вероятности первого (5) и последнего (6) слагаемых в (2) по теореме Слуцкого в силу (3) для $k = 1$ получаем требуемое

$$-2\log \mathrm{s} - \mathrm{score} = (\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}_1)^T (\widetilde{\mathbf{H}}_{m^1}^{-1}(\hat{\mathbf{w}}_1) + \widetilde{\mathbf{H}}_{m^2}^{-1}(\hat{\mathbf{w}}_2))^{-1} (\hat{w}_2 - \hat{w}_1) \xrightarrow{d} \chi^2(n) \text{ при } m \to \infty.$$