

# Функция ошибки в задачах восстановления регрессии\*

В.В. Стрижов

*Вычислительный центр РАН*

17 мая 2012 г.

## Аннотация

В работе описаны методы назначения функции ошибки при постановке задач регрессионного анализа. Рассматриваются различные гипотезы распределения зависимой переменной, задающие вид функции ошибки согласно байесовскому выводу. Для нормального распределения показана функция ошибки общего вида для различных предположений о статистической связи между элементами зависимой переменной. Также приведены примеры функций ошибок, используемых в прикладных задачах восстановления регрессии.

Ключевые слова: *восстановление регрессии, функция ошибки, регрессионная модель, ковариационная матрица, регрессионные остатки*

## 1 Введение

При моделировании измеряемых данных одной из важных проблем является оценка точности модели. Для оценки точности вводится функция ошибки, обсуждаемая в данной работе. Предполагая, что данные измеряются с некоторой погрешностью, будем рассматривать моделирование данных как задачу восстановления регрессии [1, 2, 3, 4]. Для этого введем следующие определения.

Измеряемые данные представляют собой пары значений зависимой переменной  $y$  и независимой переменной  $x$ . Считается [5], что эта зависимость является статистической и имеет вид

$$E(y|x) = f(\hat{w}, x). \quad (1)$$

---

\*Работа выполнена при поддержке Министерства образования и науки РФ в рамках Государственного контракта 07.524.11.4002 и гранта РФФИ 10-07-00422.

Для журнала «Заводская лаборатория. Диагностика материалов», 16 мая 2012 г.

*Регрессия* — математическое ожидание случайной величины  $y$ , зависящей от другой величины или от нескольких величин  $\mathbf{x}$ . Зависимость  $f$  называется *функцией регрессии* от независимой переменной  $\mathbf{x}$  при некоторых фиксированных параметрах  $\hat{\mathbf{w}}$ . Переменная  $\mathbf{x}$  также называется регрессором. Точность, с которой функция  $f$  передает изменение в среднем при изменении  $\mathbf{x}$ , измеряется дисперсией  $y$ , вычисляемой для каждого  $\mathbf{x}$ :  $D(y|\mathbf{x}) = \sigma_y^2(\mathbf{x})$ . Если  $D(y|\mathbf{x}) = 0$  при всех значениях  $\mathbf{x}$ , то с вероятностью равной единице эти величины связаны функциональной зависимостью. Если  $D(y|\mathbf{x}) \neq 0$  ни при каком значении  $\mathbf{x}$  и  $f(\hat{\mathbf{w}}, \mathbf{x})$  не зависит от  $\mathbf{x}$ , то регрессия  $y$  по  $\mathbf{x}$  отсутствует.

*Гипотезой порождения данных* называется предположение о виде распределения случайной величины  $y$  и значениях параметров этого распределения, если распределение параметрическое.

*Регрессионная выборка*  $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^m$  — множество  $m$  пар, состоящих из вектора  $\mathbf{x}_i = [x_{ij}]_{j=1}^n$  значений  $n$  свободных переменных и соответствующего этому вектору значения зависимой переменной  $y_i$ .

Предполагается, что переменные принадлежат множеству действительных чисел, либо его подмножеству:  $\mathbf{x} \in \mathbb{X} \subseteq \mathbb{R}^n$  и  $y \in \mathbb{Y} \subseteq \mathbb{R}^1$ . Индекс  $i$  элемента выборки и индекс  $j$  свободной переменной рассматриваются как элементы конечных неупорядоченных множеств  $i \in \mathcal{I} = \{1, \dots, m\}$  и  $j \in \mathcal{J} = \{1, \dots, n\}$ , причем  $\mathcal{I}, \mathcal{J} \neq \emptyset$ . В дальнейшем будет использоваться обозначение  $D = (\mathbf{X}, \mathbf{y})$ , где  $\mathbf{y} = [y_1, \dots, y_m]^\top$  — вектор значений зависимой переменной и  $\mathbf{X} = [\mathbf{x}_1^\top, \dots, \mathbf{x}_m^\top]^\top$  — *матрица плана*.

Предполагается, что элементы выборки связаны соотношением

$$y_i = f(\mathbf{w}, \mathbf{x}_i) + \varepsilon(\mathbf{x}_i), \quad (2)$$

которое аддитивно включает случайную величину  $\varepsilon = \varepsilon(\mathbf{x})$ .

*Ошибкой* или *регрессионным остатком*  $\varepsilon_i$  называется разность между значением функции регрессии  $f(\hat{\mathbf{w}}, \mathbf{x}_i)$  и значением зависимой переменной, соответствующей некоторой свободной переменной  $\mathbf{x}_i$ :  $\varepsilon_i = f(\hat{\mathbf{w}}, \mathbf{x}_i) - y_i$ . Вектор регрессионных остатков  $\boldsymbol{\varepsilon} = \mathbf{f} - \mathbf{y}$ , где вектор-функция  $\mathbf{f} = \mathbf{f}(\mathbf{w}, \mathbf{X}) = [f(\mathbf{w}, \mathbf{x}_1), \dots, f(\mathbf{w}, \mathbf{x}_m)]^\top \in \mathbb{Y}^m$ .

Для нахождения функции регрессии  $f$  используются регрессионные модели. *Регрессионная модель* — параметрическое семейство функций, отображение  $f : \mathbb{W} \times \mathbb{X} \rightarrow \mathbb{Y}$  декартова произведения области допустимых значений  $\mathbb{W}$  параметров модели и области допустимых значений  $\mathbb{X}$  свободных переменных в область значения  $\mathbb{Y}$  зависимой переменной. Иначе, регрессионная модель есть поэлементное отображение  $f : (\mathbf{w}, \mathbf{x}) \mapsto y$ , в котором вектор параметров  $\mathbf{w} \in \mathbb{W}$ , свободная переменная  $\mathbf{x} \in \mathbb{X}$  и зависимая переменная  $y \in \mathbb{Y}$ .

*Функция регрессии*  $f^*$  — функция

$$f^*(\hat{\mathbf{w}}, \mathbf{x}) = f|_{\mathbb{W} \ni \mathbf{w} = \hat{\mathbf{w}}} : \mathbb{X} \rightarrow \mathbb{Y}$$

полученная путем сужения области определения регрессионной модели  $f$  на заданное значение вектора параметров  $\hat{\mathbf{w}}$ .

Различают следующие виды регрессионных моделей:

- 1) *линейные модели* [6, 7] — модели, которые могут быть представлены в виде скалярного произведения вектора свободных переменных и вектора параметров модели

$$f = \mathbf{w}^\top \mathbf{x} \quad (3)$$

в частности, линейными являются полиномиальные и криволинейные модели;

- 2) *обобщенно-линейные модели* [8, 9, 10] — модели вида

$$f = \mu^{-1}(\mathbf{w}^\top \mathbf{x}), \quad (4)$$

где функция  $\mu$ , называемая функцией связи, принадлежит множеству функций, заданному гипотезой о том, что распределение зависимой переменной принадлежит экспоненциальному семейству;

- 3) *нелинейные модели* [11] — модели вида

$$f = f(\mathbf{w}, \mathbf{x}), \quad (5)$$

которые не могут быть представлены как линейные (3) или обобщенно-линейные (4).

Основной задачей регрессионного анализа является нахождение функции регрессии  $f$  и оценка параметров  $\mathbf{w}$ . Параметры модели назначаются таким образом, что модель наилучшим образом приближает данные. Для оценки качества решения этой задачи вводится *функция ошибки*, исходя из значения которой делается вывод о том, насколько хорошо модель приближает данные, а также насколько адекватна гипотеза порождения данных [1, 12, 13, 14].

*Функция ошибки*  $S(\mathbf{w})$  — функция, значение которой требуется минимизировать для получения оценок параметров  $\mathbf{w}$  модели  $f$ , удовлетворяющих заданному требованию. Как правило это требование:

- максимизации правдоподобия данных,
- максимизации вероятности параметров модели,
- максимизации правдоподобия самой регрессионной модели,
- состоятельности, несмещенности оценки параметров,

либо другое требование, определяемое решаемой задачей восстановления регрессии.

Функция ошибки назначается одним из двух способов:

- 1) статистическим путем; как правило путем байесовского вывода, тогда она определяется гипотезой порождения данных [15, 16, 17];

- 2) эмпирическим путем — путем, который учитывает особенности постановки решаемой прикладной задачи. В этом случае, как правило, несмещенность и другие статистические свойства полученных оценок параметров не исследуются. В качестве примера приведем функции ошибки, обеспечивающие выполнение требований промышленных стандартов [18] и требований к минимизации потерь при совершении торговых операций [19].

Далее в работе будут рассмотрены, во-первых, статистический вывод функций ошибок для нормального и биномиального распределения, и, во-вторых, несколько эмпирических функций ошибок, используемых при решении прикладных задач.

## 1.1 Функция ошибки и гипотеза порождения данных

Гипотеза порождения данных играет центральную роль в выборе функции ошибки и, как следствие, в методе оценки параметров модели. Для подтверждения или опровержения этой гипотезы выполняются статистические тесты, называемые анализом регрессионных остатков [20, 15, 16, 17]. При этом считается, что независимая переменная  $\mathbf{x}$  не является случайной величиной, не содержит ошибок и не нуждается в дополнительных статистических гипотезах.

**Гипотеза нормального распределения зависимой переменной.** Так как дисперсия и математическое ожидание зависимой переменной  $y$  зависит от значений  $\mathbf{x}_1, \dots, \mathbf{x}_m$  свободной переменной  $\mathbf{x}$ , будем считать реализации зависимой переменной  $y$  многомерной случайной величиной  $\mathbf{y} = [y_1, \dots, y_m]^T$ . В регрессионных задачах рассматриваются две основные многомерные случайные величины: зависимая переменная  $\mathbf{y}$  и вектор параметров  $\mathbf{w}$ . Пусть элементы этих многомерных случайных величин имеют конечные дисперсии.

Для зависимой переменной  $\mathbf{y}$  выполняется один из трех вариантов гипотезы порождения данных:

- 1) элементы зависимой переменной, случайной величины  $\mathbf{y}$ , имеют одинаковую дисперсию  $\sigma^2(\mathbf{y})$  и независимы,  $\text{cov}(y_i, y_j) = 0, i, j \in \mathcal{I}, i \neq j$ ,

$$\mathbf{y} \sim \mathcal{N}(\mathbf{f}, \sigma^2(\mathbf{y})\mathbf{I}) \stackrel{\text{def}}{=} \mathcal{N}(\mathbf{f}, \beta^{-1}\mathbf{I}); \quad (6)$$

- 2) элементы переменной  $\mathbf{y}$  имеют различную дисперсию и независимы, то есть,

$$\mathbf{y} \sim \mathcal{N}(\mathbf{f}, \text{diag}^{-1}(\beta_1, \dots, \beta_m)\mathbf{I}); \quad (7)$$

- 3) элементы переменной  $\mathbf{y}$  описываются ковариационной матрицей общего вида,

$$\mathbf{y} \sim \mathcal{N}(\mathbf{f}, \mathbf{B}^{-1}). \quad (8)$$

В выражениях (3), (4) и (5) переменные  $\mathbf{y}$  и  $\mathbf{w}$  связаны функциональной зависимостью. Поэтому параметры модели  $f$  также являются случайными величинами. Распределение этих параметров зависит от гипотезы порождения данных. При заданной линейной модели

$$\mathbb{E}(\mathbf{y}|\mathbf{X}) = \mathbf{f}(\mathbf{w}, \mathbf{X}) = \mathbf{X}\mathbf{w},$$

либо при заданном линейризованном виде

$$\mathbb{E}(\mathbf{y} - \mathbf{f}|\mathbf{X}) = \mathbf{f}(\mathbf{w}, \mathbf{X}) = \mathbf{J}\Delta\mathbf{w},$$

в котором матрица  $\mathbf{J}$  есть матрица Якоби функции  $f$ , линейное отображение задаваемое матрицей  $\mathbf{X}$ , переводит распределение многомерной случайной величины  $\mathbf{y}$  в распределение многомерной случайной величины  $\mathbf{w}$ . В случае нормального распределения случайной величины  $\mathbf{y}$  распределение величины  $\mathbf{w}$  также является нормальным. Обозначим  $\mathbf{A}^{-1}$  ковариационную матрицу параметров  $\mathbf{w}$ .

Как и для зависимой переменной  $\mathbf{y}$ , для параметров  $\mathbf{w}$  выполняется один из трех вариантов:

- 1) параметры имеют одинаковую дисперсию  $\sigma^2(\mathbf{w})$  и независимы,  $\text{cov}(w_j, w_k) = 0, j, k \in \mathcal{J}, j \neq k$ ,

$$\mathbf{w} \sim \mathcal{N}(\mathbf{w}_{\text{ML}}, \sigma^2(\mathbf{w})\mathbf{I}) \stackrel{\text{def}}{=} \mathcal{N}(\mathbf{0}, \alpha^{-1}\mathbf{I}); \quad (9)$$

- 2) параметры модели имеют различную дисперсию и независимы,

$$\mathbf{w} \sim \mathcal{N}(\mathbf{w}_{\text{ML}}, \text{diag}^{-1}(\alpha_1, \dots, \alpha_n)\mathbf{I}); \quad (10)$$

- 3) параметры модели описываются ковариационной матрицей общего вида

$$\mathbf{w} \sim \mathcal{N}(\mathbf{w}_{\text{ML}}, \mathbf{A}^{-1}). \quad (11)$$

Параметры  $\alpha_j, j \in \mathcal{J}$  распределения параметров  $\mathbf{w}$  модели  $f(\mathbf{w}, \mathbf{x})$  называются *гиперпараметрами*. Ковариационная матрица  $\mathbf{A} = [\alpha_{kj}]$  называется матрицей гиперпараметров. Для гипотез (9), (10) и (11) матрица гиперпараметров имеет, соответственно, вид

- 1)  $\mathbf{A} = \alpha\mathbf{I}$ ,
- 2)  $\mathbf{A} = \text{diag}(\alpha_1, \dots, \alpha_n)\mathbf{I}$ ,
- 3)  $\mathbf{A}$  — симметричная, неотрицательно определенная,  $\mathbf{A} = \mathbf{A}^\top$ , и  $\mathbf{w}^\top \mathbf{A} \mathbf{w} \geq 0$  для любого  $\mathbf{w} \in \mathbb{R}^n$ .

Аналогично, назовем гиперпараметрами и обозначим  $\beta_i, i \in \mathcal{I}$  параметры распределения зависимой переменной  $\mathbf{y}$ .

Рассмотрим нормальное распределение зависимой переменной в качестве гипотезы порождения данных при восстановлении линейной или существенно-нелинейной регрессии:

$$\mathbf{E}(y|\mathbf{x}) = f(\mathbf{w}, \mathbf{x}).$$

Для нахождения *наиболее правдоподобных параметров* модели используем метод наибольшего правдоподобия [21, 22, 23]. Пусть многомерная случайная величина  $\mathbf{y} \sim \mathcal{N}(\mathbf{f}, \mathbf{B})$  имеет нормальное распределение

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{m}{2}} \det^{\frac{1}{2}}(\mathbf{B}^{-1})} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{f})^{\top} \mathbf{B}(\mathbf{y} - \mathbf{f})\right). \quad (12)$$

В выражении (12) часть под экспонентой  $(\mathbf{y} - \mathbf{f})^{\top} \mathbf{B}(\mathbf{y} - \mathbf{f})$  является квадратом расстояния Махаланобиса [24]. Матрицу можно представить в виде разложения Холецкого [25],  $\mathbf{B} = \mathbf{L}\mathbf{L}^{\top}$ , где  $\mathbf{L}$  — диагональная нижнетреугольная матрица, либо в виде произведения  $\mathbf{B} = \mathbf{U}^{\top}\mathbf{U}$ , см. [26]. Оба разложения единственны. В общем случае не предполагается, что её элементы независимы и не коррелируют. Ковариационная матрица  $\mathbf{B}$  является симметричной неотрицательно определенной матрицей. Диагональные элементы  $\beta_i$  этой матрицы являются обратны значениям дисперсии элементов случайной величины  $\mathbf{y}$ :

$$\sigma_i^2 = \frac{1}{\beta_i}, \quad i \in \mathcal{I}.$$

Так как правая часть выражения (12) зависит от вида регрессионной модели  $f$ , вектора параметров  $\mathbf{w}$ , независимой переменной  $\mathbf{x}$  и от ковариационной матрицы  $\mathbf{B}$ , перепишем его в виде

$$p(\mathbf{y}|\mathbf{x}, \mathbf{w}, \mathbf{B}, f) \stackrel{\text{def}}{=} p(D|\mathbf{w}, \beta, f) = \frac{\exp(-E_D)}{Z_D(\beta)}, \quad (13)$$

где  $Z_D$  — нормирующий коэффициент для плотности нормального распределения

$$Z_D = (2\pi)^{\frac{m}{2}} \det^{\frac{1}{2}}(\mathbf{B}^{-1}). \quad (14)$$

Функция ошибки, соответствующая матожиданию регрессионной модели при данной гипотезе, определена как

$$E_D = \frac{1}{2}(\mathbf{y} - \mathbf{f})^{\top} \mathbf{B}(\mathbf{y} - \mathbf{f}). \quad (15)$$

Рассмотрим частный случай гипотезы порождения данных: элементы вектора  $\mathbf{y}$  не коррелируют и имеют одинаковую дисперсию, то есть обратная ковариационная матрица  $\mathbf{B} = \beta \mathbf{I}_m$ . В этом случае вид функции правдоподобия (13) упрощается: коэффициент  $Z_D$  имеет вид

$$Z_D(\beta) = \left(\frac{2\pi}{\beta}\right)^{\frac{m}{2}},$$

так как

$$Z_D(\beta) = \det^{\frac{1}{2}}(\sigma^2 \mathbf{I})(2\pi)^{\frac{m}{2}} = \sigma^{\frac{2m}{2}} \det^{\frac{1}{2}}(\mathbf{I})(2\pi)^{\frac{m}{2}} = \sigma^m (2\pi)^{\frac{1}{2}} = \left(\frac{2\pi}{\beta}\right)^{\frac{m}{2}},$$

а функция ошибки равна

$$E_D = \frac{1}{2} \beta \sum_{i \in \mathcal{I}} (y_i - f(\mathbf{w}, \mathbf{x}_i))^2,$$

так как при

$$(\sigma^2 \mathbf{I})^{-1} = \frac{\mathbf{I}}{\sigma^2} = \beta \mathbf{I}$$

справедливо выражение

$$E_D = \frac{1}{2} (\mathbf{y} - \mathbf{f})^\top (\beta \mathbf{I}) (\mathbf{y} - \mathbf{f}) = \frac{1}{2} \beta \|\mathbf{y} - \mathbf{f}\|^2.$$

Рассмотрим вектор параметров  $\mathbf{w}$  модели  $f$  — многомерную случайную величину. Согласно принятой гипотезе распределения (12) зависимой переменной и теореме о функциях связи распределений [27, 28], распределение параметров  $\mathbf{w} \sim \mathcal{N}(\mathbf{w}_{\text{ML}}, \mathbf{A}^{-1})$  является нормальным с матожиданием  $\mathbf{w}_{\text{ML}}$ , ковариационной матрицей  $\mathbf{A}^{-1}$  и имеет вид

$$p(\mathbf{w} | \mathbf{A}, f) = \frac{\exp(-E_{\mathbf{w}})}{Z_{\mathbf{w}}(\mathbf{A})}. \quad (16)$$

Выражение (16) справедливо для линейных моделей, поскольку многомерные случайные величины  $\mathbf{y}$  и  $\mathbf{w}$  связаны линейным отображением  $\mathbf{X}$ . Для существенно нелинейных моделей предполагается, что это выражение будет справедливо в окрестности  $\Delta \mathbf{w}$  некоторой точки  $\mathbf{w}_0$  при линеаризации

$$\mathbf{y} - \mathbf{f}(\hat{\mathbf{w}}, \mathbf{X}) = \mathbf{J} \Delta \mathbf{w}.$$

Матрица Якоби  $\mathbf{J}$  — это матрица частных производных модели  $f$  по элементам  $w_j$  вектора параметров  $\mathbf{w}$ :

$$\mathbf{J} = \left[ \frac{\partial f(\mathbf{w}, \mathbf{x}_i)}{\partial w_j} \right], \quad \text{где } \mathbf{w} = [w_1, \dots, w_j, \dots, w_n]^\top \text{ и } i \in \mathcal{I}, j \in \mathcal{J}.$$

Это предположение используется, например, в аппроксимации Лапласа [29, 30], согласно которой функция распределения параметров существенно нелинейной модели может быть приближена функций нормального распределения.

Нормирующий коэффициент  $Z_{\mathbf{w}}(\mathbf{A})$  равен

$$Z_{\mathbf{w}}(\mathbf{A}) = (2\pi)^{\frac{n}{2}} \det^{\frac{1}{2}}(\mathbf{A}^{-1}), \quad (17)$$

где  $n$  — число параметров модели  $f$ . Функция-штраф за большое значение параметров модели для принятого распределения определена как

$$E_{\mathbf{w}} = \frac{1}{2}(\mathbf{w} - \mathbf{w}_0)^\top \mathbf{A}(\mathbf{w} - \mathbf{w}_0). \quad (18)$$

Рассмотрим частный случай: дисперсии элементов  $w_j$  вектора параметров  $\mathbf{w}$  равны, обратная ковариационная матрица имеет вид  $\mathbf{A} = \alpha \mathbf{I}_n$ . В этом случае выражения (17) и (18) будут иметь вид

$$Z_{\mathbf{w}}(\alpha) = \left(\frac{2\pi}{\alpha}\right)^{\frac{n}{2}} \quad \text{и} \quad E_{\mathbf{w}} = \frac{1}{2}\alpha \|\hat{\mathbf{w}} - \mathbf{w}\|^2.$$

Для нахождения *наиболее вероятных параметров* модели  $f(\mathbf{w}, \mathbf{x})$  используем Байесовский вывод [31, 32, 33]. При заданной модели  $f$  и заданных значениях  $\mathbf{A}$  и  $\mathbf{B}$  выражение (13) принимает вид

$$p(\mathbf{w}|D, \mathbf{A}, \mathbf{B}, f) = \frac{p(D|\mathbf{w}, \mathbf{B}, f)p(\mathbf{w}|\mathbf{A}, f)}{p(D|\mathbf{A}, \mathbf{B}, f)}. \quad (19)$$

Элементы этого выражения и соответствующие им параметры:

$p(\mathbf{w}|D, \mathbf{A}, \mathbf{B}, f)$  — апостериорное распределение параметров,

$\mathbf{w}_{\text{MP}} = \arg \max p(\mathbf{w}|D, \mathbf{A}, \mathbf{B}, f)$  — наиболее вероятные параметры,

$\mathbf{w}_{\text{ML}} = \arg \max p(D|\mathbf{w}, \mathbf{B}, f)$  — наиболее правдоподобные параметры,

$p(D|\mathbf{w}, \mathbf{B}, f)$  — функция правдоподобия данных,

$p(\mathbf{w}|\mathbf{A}, f)$  — априорное распределение параметров,

$p(D|\mathbf{A}, \mathbf{B}, f)$  — функция правдоподобия модели  $f$ .

Запишем функцию ошибки  $S = E_{\mathbf{w}} + E_D$  в общем виде:

$$S(\mathbf{w}) = \frac{1}{2}(\mathbf{w} - \mathbf{w}_{\text{ML}})^\top \mathbf{A}(\mathbf{w} - \mathbf{w}_{\text{ML}}) + \frac{1}{2}(\mathbf{y} - \mathbf{f})^\top \mathbf{B}(\mathbf{y} - \mathbf{f}). \quad (20)$$

Получим вместо (19) выражение  $p(\mathbf{w}|D, \mathbf{A}, \mathbf{B}, f) = \frac{1}{Z_S} \exp(-S(\mathbf{w}))$ , где  $Z_S$  — нормирующий коэффициент.

Заметим, что матожидание вектора параметров в некоторых случаях может быть гипотетически принято равным нулю,  $\mathbf{w}_{\text{ML}} = 0$ . Такое предположение явно или неявно принимается при решении задач выбора признаков линейных регрессионных моделей, см. Лассо [34], Stagewise [35], LARS [36], а также при прореживании некоторых нелинейных моделей [37, 38].



Рис. 1: Пример регрессионной выборки и функции регрессии.

При рассмотрении частных случаев ковариационных матриц  $\mathbf{A} = \alpha \mathbf{I}_n$  и  $\mathbf{B} = \beta \mathbf{I}_m$ , параметров модели  $\mathbf{w}$  и гомоскедастичной зависимой переменной  $\mathbf{y}$  выражение (19) принимает вид

$$p(\mathbf{w}|D, \alpha, \beta, f) = \frac{p(D|\mathbf{w}, \beta, f)p(\mathbf{w}|\alpha, f)}{p(D|\alpha, \beta, f)}.$$

а функция ошибки —

$$S(\mathbf{w}) = \frac{1}{2}\alpha\|\mathbf{w}\|^2 + \frac{1}{2}\beta\|\mathbf{y} - \mathbf{f}\|^2.$$

Гиперпараметры  $\alpha$  и  $\beta$  в последнем выражении играют роль регуляризирующих множителей [39, 40].

Общепринятая функция ошибки — сумма квадратов регрессионных остатков

$$\text{SSE} = \|\mathbf{y} - \mathbf{f}\|^2$$

является частным случаем функции ошибки общего вида (20). При этом гиперпараметры  $\alpha = 0$  и  $\beta = \frac{1}{2}$ .

В качестве примера иллюстрации различных вариантов ковариационных матриц параметров  $\mathbf{w}$  рассмотрим выборку, представленную на рис. 1. В такой выборке одному значению переменной  $\mathbf{x}$  соответствует несколько значений переменной  $\mathbf{y}$ . Линией показана функция регрессии  $f(\hat{\mathbf{w}}, \xi) = w_1 + w_2\xi^2 + \varepsilon(\xi)$ . Ее оптимальные параметры  $\hat{\mathbf{w}} = [0.2839, 0.2412]$ . Соответствующая регрессионная модель имеет вид  $f = \mathbf{x}^T \mathbf{w}$ , где  $x_1 = \xi^0, x_2 = \xi^2$ . На рис. 2 показано значение функции ошибки в окрестности вектора параметров  $\hat{\mathbf{w}}$ . Верхняя строка графиков показывает теоретической функции плотности распределения параметров согласно полученным оценкам ковариационной матрицы. Нижняя строка графиков показывает проекцию теоретической функции плотности на пространство параметров и эмпирическую функцию плотности, восстановленную путем сэмплирования параметров. Эмпирическая плотность на графиках одинакова и соответствует функции плотности с ковариационной матрицей общего вида. Столбцы графиков слева на право соответствуют гипотезам о видах ковариационных матриц  $\mathbf{A} = \alpha \mathbf{I}$  (элементы многомерной случайной величины  $\mathbf{w}$  независимы и распределены с одинаковой дисперсией),  $\mathbf{A} = \text{diag}(\alpha_1, \dots, \alpha_n)$  (элементы независимы и распределены с различными дисперсиями) и последний столбец — матрица  $\mathbf{A}$  имеет общий вид, симметрична и неотрицательно определена.

**Биномиальное распределение зависимой переменной и функция ошибки.** Рассмотрим задачу восстановления логистической регрессии [41, 42, 28] как частный

Рис. 2: Вид функции  $\exp(-S(\mathbf{w}))$  в окрестности вектора параметров  $\hat{\mathbf{w}}$ .

случай задачи восстановления регрессии (1). Примем гипотезу о биномиальном распределении зависимой переменной  $y \in \{0, 1\}$ :

$$y \sim \mathcal{B}(f, 1 - f), \quad (21)$$

случайная величина  $y$  принимает значение 0 с вероятностью  $p$  и значение 1 с вероятностью  $1 - p$ . Таким образом, функция регрессии  $f(\hat{\mathbf{w}}, \mathbf{x})$  в случае биномиального распределения зависимой переменной восстанавливает регрессию, равную вероятности принадлежности вектора  $\mathbf{x}$  к одному из двух классов. Функция правдоподобия реализаций многомерной случайной величины  $\mathbf{y}$  имеет вид

$$p(\mathbf{y}|\mathbf{w}, \mathbf{X}, f) = \prod_{i=1}^m f(\mathbf{w}, \mathbf{x}_i)^{y_i} (1 - f(\mathbf{w}, \mathbf{x}_i))^{1-y_i}. \quad (22)$$

Модель логистической регрессии имеет вид

$$\mathbf{f}(\mathbf{w}, \mathbf{X}) = \boldsymbol{\sigma}(-\mathbf{X}\mathbf{w}) = \frac{1}{1 + \exp(-\mathbf{X}\mathbf{w})}. \quad (23)$$

Рис. 3: Логистическая регрессия от двух переменных.

В [27, 28] показано, что из гипотезы о принадлежности распределения  $p(\mathbf{y})$  зависимой переменной  $\mathbf{y}$  каноническому экспоненциальному семейству, при использовании соответствующей этому распределению функции связи  $\boldsymbol{\mu}$ , следует, что параметры  $\mathbf{w} \sim \mathcal{N}(\hat{\mathbf{w}}, \mathbf{A}^{-1})$  обобщенно-линейной модели распределены нормально. Следовательно, учитывая ранее введенную функцию ошибки параметров (18) и вышеприведенную функцию правдоподобия (22) получим, что при гипотезе (21) функция ошибки имеет вид

$$S(\mathbf{w}) = - \sum_{i=1}^m (y_i \ln f(\mathbf{w}, \mathbf{x}_i) + (1 - y_i) \ln(1 - f(\mathbf{w}, \mathbf{x}_i))). \quad (24)$$

На рисунке 3 синими и красными точками показана выборка, состоящая из двух классов. Восстановленная регрессия показана поверхностью  $f(\hat{\mathbf{w}}, \mathbf{x})$ , значение которой равно вероятности принадлежности элемента выборки  $\mathbf{x}$  к одному из двух классов. Прогнозируемая принадлежность показана окружностями. При несовпадении цветов точки и окружности следует говорить об ошибке классификации.

Как и ранее, найдем плотность распределения  $p(\mathbf{w}|D, \alpha, f)$  параметров модели, воспользовавшись формулой Байеса

$$p(\mathbf{w}|D, \alpha, f) = \frac{p(D|\mathbf{w}, f)p(\mathbf{w}|\alpha, f)}{p(D|\alpha, f)}, \quad (25)$$

где  $p(D|\mathbf{w}, f)$  — правдоподобие данных,  $p(\mathbf{w}|\alpha, f)$  — задаваемая априорно плотность распределения параметров модели. В выражении (25) нормировочный множитель  $p(D|\alpha, f)$  определяется выражением

$$p(D|\alpha, f) = \int p(D|\mathbf{w}, f)p(\mathbf{w}|\alpha, f)d\mathbf{w}.$$

Обозначив  $Z(\alpha) = p(D|\alpha, f)$ , получим

$$\begin{aligned} p(\mathbf{w}|D, f) &= \frac{p(y|\mathbf{x}, \mathbf{w}, f)p(\mathbf{w}|f, \alpha)}{Z(\alpha)} = \\ &= \frac{\alpha^{\frac{|I|}{2}}}{(2\pi)^{\frac{|I|}{2}} Z(\alpha)} \exp\left(-\frac{\alpha}{2}\|\mathbf{w} - \hat{\mathbf{w}}\|^2\right) \prod_{i=1}^m f(\mathbf{w}, \mathbf{x}_i)^{y_i} (1 - f(\mathbf{w}, \mathbf{x}_i))^{1-y_i}, \end{aligned}$$

где  $Z(\alpha)$  — нормировочный множитель.

**Экспоненциальное семейство.** Экспоненциальное семейство распределений [43] многомерной случайной величины  $\mathbf{y}$  с вектором параметров  $\boldsymbol{\eta}$  задается набором распределений вида

$$p(\mathbf{y}|\boldsymbol{\eta}) = h(\mathbf{y})g(\boldsymbol{\eta}) \exp(\boldsymbol{\eta}^T \mathbf{u}(\mathbf{y})). \quad (26)$$

Вектор  $\boldsymbol{\eta}$  называется вектором естественных параметров распределения. Сомножитель  $\mathbf{u}(\mathbf{y})$  — некоторая вектор-функция многомерной случайной величины  $\mathbf{y}$ .

Гипотеза принадлежности распределения зависимой переменной экспоненциальному семейству используется при построении обобщенных линейных моделей (4), то есть, моделей вида

$$\mathbf{E}(\mathbf{y}) = \boldsymbol{\mu}(\mathbf{X}\mathbf{w}), \quad (27)$$

для которых  $\boldsymbol{\mu}$  — функция связи, а  $\mathbf{X}\mathbf{w}$  — линейная комбинация признаков выборки. В связи с тем, что гипотезы экспоненциального семейства, за исключением нормального и биномиального распределения, на практике используются весьма редко, в данной работе соответствующие им функции ошибки опущены.

## 2 Эмпирическая функции ошибки

В ряде прикладных задач, например, в задачах, когда оценка погрешности измерений задана стандартами или другими специальными требованиями, гипотезы порождения данных не рассматриваются и функция ошибки назначается согласно этим требованиям.

Например, при восстановлении регрессии измерений некоторых физических величин используется метод наименьших модулей [44, 45, 46], согласно которому функция ошибки задана как сумма модулей регрессионных остатков. Рассмотрим симметричную функцию ошибки — абсолютную ошибку вида

$$S(\mathbf{w}) = \sum_{i \in I} \|f(\mathbf{w}, \mathbf{x}_i) - y_i\|_1. \quad (28)$$

и ее общий вид — несимметричную функцию ошибки, являющаяся кусочно-линейной функцией, вида

$$S(\mathbf{w}) = \sum_{i \in I} \sum_s \begin{cases} a_s + b_s(f(\mathbf{w}, \mathbf{x}_i) - y_i), & \text{при } f \in (z_{s-1}, z_s); \\ 0, & \text{в противном случае.} \end{cases} \quad (29)$$

Параметры  $a_s, b_s$  и концы отрезков  $z_s$  выбираются согласно условиям прикладной задачи. Функция ошибки  $S(\mathbf{w})$  оптимизируется, исходя из условий поставленной задачи. Задача нахождения минимального значения функций вида (28) или (29) решается методами линейного программирования. В таблице 1 приведен набор функций ошибок, часто используемых при решении задач прогнозирования.

Среднее арифметическое модулей остатков	$\text{MAE} = \frac{1}{m} \sum_{i=1}^m  \varepsilon_i $
Среднее арифметическое модулей относительных остатков	$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m \left  \frac{\varepsilon_i}{y_i} \right $
Среднее отклонение модулей остатков	$\text{PMAD} = \frac{1}{m} \sum_{i=1}^m  \varepsilon_i  \left( \sum_{i=1}^m  y_i  \right)^{-1}$
Среднеквадратичная ошибка	$\text{MSE} = \frac{1}{m} \sum_{i=1}^m \varepsilon_i^2$
Корень среднеквадратичной ошибки	$\text{RMSE} = \frac{1}{\sqrt{m}} \sqrt{\sum_{i=1}^m \varepsilon_i^2}$
Сила прогноза	$\text{SS} = 1 - \frac{\text{MSE}_{\text{forecast}}}{\text{MSE}_{\text{history}}}$

Таблица 1: Функции ошибок регрессионных моделей.

**Функция ошибки и разбиение выборки.** Выше не предполагалось, что для оценки наиболее вероятных параметров модели, либо для выбора наиболее правдоподобной модели из некоторого множества требуется разбиение множества индексов  $\mathcal{I}$  элементов выборки  $D = \{(\mathbf{x}_i, y_i)\}, i \in \mathcal{I}$  на обучающую и контрольную:  $\mathcal{I} = \mathcal{L} \sqcup \mathcal{C}$ . Тем не менее, следует отметить, что при выборе моделей такое разбиение является одним из наиболее эффективных способов избежать переобучения, см. [47, 48, 49]. Поэтому ниже приведен ряд примеров эвристических функций ошибок, предложенных авторами метода группового учета аргументов. Данные функции были названы критериями. Значительная их часть опубликована в [50].

Используемые в этом подразделе обозначения  $\mathbf{X}_{\mathcal{C}}, \mathbf{y}_{\mathcal{C}}, \mathbf{w}_{\mathcal{L}}$  означают, что значения переменных  $\mathbf{X}, \mathbf{y}, \mathbf{w}$  фиксированы, в выборку  $(\mathbf{X}_{\mathcal{C}}, \mathbf{y}_{\mathcal{C}})$  вошли только объекты с индексами из множества  $\mathcal{C} \in \mathcal{I} \neq \emptyset$ , а оценка вектора параметров  $\mathbf{w}_{\mathcal{L}}$  получена с использованием выборки, состоящей из элементов с индексами из множества  $\mathcal{L} \subset \mathcal{I} \neq \emptyset$ :

$$\mathbf{w}_{\mathcal{L}} = \arg \min_{i \in \mathcal{L} \subset \mathcal{I}} S(\mathbf{w} | \mathbf{X}_{\mathcal{L}}, \mathbf{y}_{\mathcal{L}}, f).$$

При этом считается, что множество индексов  $\mathcal{I}$  элементов выборки разбито на подмножества

$$\mathcal{I} = \mathcal{L} \sqcup \mathcal{C} \sqcup \mathcal{V},$$

в котором  $\mathcal{L}$  — обучающая выборка,  $\mathcal{C}$  — контрольная выборка,  $\mathcal{V}$  — валидационная выборка. Последняя в ряде задач может быть пустой.

Метод группового учета аргументов [51, 52, 53] использует внутренний и внешний критерий, так как при оценке параметров моделей и при выборе моделей используются разные элементы выборки. *Внутренний критерий* используется для оценки параметров: их значения оцениваются на подвыборке элементов с индексами из  $\mathcal{L}$ . Выбор моделей производится с помощью *внешнего критерия*, значение которого вычисляется на множестве  $\mathcal{C}$ . При выборе минимум внешнего критерия означает, что модель, доставляющая такой минимум, является искомой.

*Критерий регулярности*  $S_{\Delta^2}$  равен норме разности вектора значений зависимой переменной и вектора значений функции регрессии на тестовой подвыборке  $\mathcal{C}$  при параметрах, оцененных на обучающей подвыборке  $\mathcal{L}$ .

$$S_{\Delta^2} = \|\mathbf{y}_{\mathcal{C}} - \mathbf{X}_{\mathcal{C}}\mathbf{w}_{\mathcal{L}}\|^2,$$

где

$$\mathbf{w}_{\mathcal{L}} = (\mathbf{X}_{\mathcal{L}}^T \mathbf{X}_{\mathcal{L}})^{-1} (\mathbf{X}_{\mathcal{L}}^T \mathbf{y}_{\mathcal{L}}).$$

Этот критерий может быть нормирован выражениями  $\|\mathbf{y}_{\mathcal{L}}\|^2$  или  $\|\mathbf{y}_{\mathcal{L}} - \text{mean}(\mathbf{y}_{\mathcal{L}})\|^2$ .

*Критерий предсказательной способности* — модификация критерия регулярности для задач прогнозирования. Этот критерий включает среднеквадратичную ошибку для валидационной выборки  $\mathcal{V}$ , которая не была использована ни при оценке параметров, ни при выборе модели. В этом случае выборка делится на три части. Критерий предсказательной способности имеет вид

$$S_{\Delta^2_{\mathcal{V}}} = \frac{\|\mathbf{y}_{\mathcal{I}} - \mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{V}}\|^2}{\|\mathbf{y}_{\mathcal{I}} - \text{mean}(\mathbf{y}_{\mathcal{I}})\|^2}.$$

*Критерий минимального смещения* или *критерий непротиворечивости*: модель, которая имеет на обучающей и на контрольной выборках различные векторы невязок, называется противоречивой. Критерий задан разностью между значениями функции регрессии, вычисленными на двух различных выборках, заданных множествами  $\mathcal{L}$  и  $\mathcal{C}$  и требует, чтобы оценки параметров, вычисленные на этих выборках, различались минимально. Он имеет вид:

$$S_{\eta_{\text{bs}}^2} = \|\mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{C}} - \mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{L}}\|^2,$$

модификация:

$$S_{\eta_{\text{a}}^2} = \|\mathbf{w}_{\mathcal{C}} - \mathbf{w}_{\mathcal{L}}\|^2,$$

где  $\mathbf{w}_{\mathcal{C}}$  и  $\mathbf{w}_{\mathcal{L}}$  — векторы параметров, полученные с использованием подвыборок  $\mathcal{C}$  и  $\mathcal{L}$ .

Критерий иммунитета к шуму имеет вид

$$S_{V^2} = (\mathbf{X}_{\mathcal{I}}\mathbf{w}_C - \mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{I}})^\top (\mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{I}} - \mathbf{X}_{\mathcal{I}}\mathbf{w}_{\mathcal{L}}) = (\mathbf{w}_C - \mathbf{w}_{\mathcal{I}})^\top \mathbf{X}_{\mathcal{I}}^\top \mathbf{X}_{\mathcal{I}} (\mathbf{w}_{\mathcal{I}} - \mathbf{w}_{\mathcal{L}}),$$

где  $\mathbf{w}_{\mathcal{I}}$  — вектор параметров, полученный с использованием полной выборке  $\mathcal{I}$ . Утверждается [51], что с помощью этого критерия в сильно зашумленных данных можно найти скрытые закономерности.

Комбинированный критерий позволяет использовать при выборе моделей линейную комбинацию нескольких критериев. Комбинированный критерий

$$S_{\kappa^2} = \sum_{k=1}^K v_k S_k, \quad \text{при условии} \quad \sum_{k=1}^K v_k = 1.$$

Здесь  $S_k$  — принятые на рассмотрение критерии, а  $v_k$  — веса этих критериев, назначенные в начале вычислительного эксперимента.

### 3 Заключение

В работе рассматриваются связи гипотезы порождения данных — статистического предположения о характере распределения зависимой переменной и функции ошибки с точки зрения байесовского вывода. Показаны функции ошибки общего вида для нормального и биномиального распределения. Также приведены эмпирические функции ошибки, назначаемые исходя не из статистических гипотез, а из требований к качеству регрессионных моделей, предъявляемых прикладными задачами.

### Список литературы

- [1] *Draper N. R., Smith H.* Applied Regression Analysis. — John Wiley and Sons, 1998. — 706 pp.
- [2] *Демиденко Е. З.* Оптимизация и регрессия. — М.: Наука, 1989. — 296 с.
- [3] *Орлов А. И.* Прикладная статистика. — М.: Экзамен, 2006. — 671 с.
- [4] *Магнус Я. Р., Катъшев П. К., Персецкий А. А.* Эконометрика. — М.: Дело, 2004. — 576 с.
- [5] Вероятность и математическая статистика: Энциклопедия / Под ред. Ю. В. Прохоров. — М: Большая Российская энциклопедия, 1999. — 910 с.
- [6] *Kutner M. H., Nachtsheim C. J., Neter J.* Applied linear regression models. — New York: McGraw-Hill/Irwin, 2004. — 867 pp.

- [7] *Searle S. R.* Linear models. — New York: John Wiley & Sons, 1971. — 532 pp.
- [8] *McCullagh P., Nelder J.* Generalized Linear Models. — Boca Raton: Chapman and Hall/CRC, 1989. — 532 pp.
- [9] *Dobson A. J., Barnett A. G.* Introduction to Generalized Linear Models. — Boca Raton, FL: Chapman and Hall/CRC, 2008. — 391 pp.
- [10] *Hardin J., Hilbe J.* Generalized Linear Models and Extensions. — College Station: Stata Press, 2007. — 387 pp.
- [11] *Seber G. A. F., Wild C.* Nonlinear Regression. — Wiley-IEEE, 2003. — 768 pp.
- [12] *Айвазян С. А., Енюков И. С., Мешалкин Л. Д.* Прикладная статистика: исследование зависимостей. — М.: Финансы и статистика, 1985. — 471 с.
- [13] *Ванник В. Н.* Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979. — 448 с.
- [14] *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning. — Springer, 2001. — 533 pp.
- [15] *Lehmann E. L., Romano J. P.* Testing Statistical Hypothesis. — Springer, 2005. — 784 pp.
- [16] *Pagan A. R., Hall A. D.* Diagnostic tests as residual analysis. — Australian National University, 1983. — 55 pp.
- [17] *Montgomery D. C.* Design and analysis of experiments. — John Wiley and Sons, 2008. — 696 pp.
- [18] ГОСТ 8.207-76, государственная система обеспечения единства измерений. Прямые измерения с многократными наблюдениями. методы обработки результатов наблюдений. Основные положения.
- [19] *Каневский Д. Ю., Кудинов П. Ю., Воронцов К. В.* Прогнозирование с несимметричной функцией потерь при наличии стохастического тренда // Интеллектуализация обработки информации (ИОИ-2008): тезисы докладов. — 2008. — С. 113–115.
- [20] *Anscombe F. J., Tukey J. W.* The examination and analysis of residuals // *Technometrics*. — 1963. — Vol. 5. — Pp. 141–160.
- [21] *Le Cam L., Lo Yang G.* Asymptotics in statistics: some basic concepts. — Springer, 2000. — 285 pp.
- [22] *Le Cam L.* Maximum likelihood — an introduction // *ISI Review*. — 1990. — Vol. 58 (2). — Pp. 153–171.

- [23] *Hald A.* On the history of maximum likelihood in relation to inverse probability and least squares // *Statistical Science*. — 1999. — Vol. 14 (2). — Pp. 214–222.
- [24] *Mahalanobis P. C.* On the generalised distance in statistics // *Proceedings of the National Institute of Sciences of India*. — 1936. — Vol. 2 (1). — Pp. 49–55.
- [25] *Golub G. H., Loan C. F. V.* Matrix computations. — Johns Hopkins University Press, 1996. — 728 pp.
- [26] *Садовничий В. А.* Теория операторов. — Дрофа, 2001. — 352 с.
- [27] *Муха В. С.* Статистические методы обработки данных. — Минск: Издательский центр БГУ, 2009. — 183 с.
- [28] *Lee Y., Nelder J. A., Pawitan Y.* Generalized linear models with random effects: unified analysis via h-likelihood. — Chapman & Hall/CRC, 2006. — 411 pp.
- [29] *MacKay D. J.* Choice of basis for laplace approximation: Tech. rep.: Machine Learning, 1998.
- [30] *Welling M., Parise S.* Bayesian random fields: The bethe-laplace approximation // UAI. — AUAI Press, 2006. — Pp. 512–519.
- [31] *Bickel P. J., Doksum K. A.* Mathematical Statistics, Volume 1: Basic and Selected Topics. — Pearson Prentice-Hall, 2007. — 556 pp.
- [32] *Jaynes E.* Probability Theory: The Logic of Science. — CUP, 2003. — 758 pp.
- [33] *Howson C., Urbach P.* Scientific Reasoning: the Bayesian Approach. — Open Court Publishing Company, 2005. — 327 pp.
- [34] *Tibshirani R.* Regression shrinkage and selection via the lasso // *Journal of the Royal Statistical Society*. — 1996. — Vol. 32, no. 1. — Pp. 267–288.
- [35] *Hastie T., Taylor J., Tibshirani R., Walther G.* Forward stagewise regression and the monotone lasso // *Electronic Journal of Statistics*. — 2007. — Vol. 1, no. 1. — Pp. 1–29.
- [36] *Efron B., Hastie T., Johnstone I., Tibshirani R.* Least angle regression // *The Annals of Statistics*. — 2004. — Vol. 32, no. 2. — P. 407–499.
- [37] *Hassibi B., Stork D. G.* Second order derivatives for network pruning: Optimal brain surgeon // Advances in Neural Information Processing Systems / edited by S. J. Hanson, J. D. Cowan, C. L. Giles. — Vol. 5. — Morgan Kaufmann, San Mateo, CA, 1993. — Pp. 164–171.
- [38] *LeCun Y., Denker J., Solla S., Howard R. E., Jackel L. D.* Optimal brain damage // Advances in Neural Information Processing Systems II / edited by D. S. Touretzky. — San Mateo, CA: Morgan Kauffman, 1990. — Pp. 598–605.



- [39] *Bjorkstrom A.* Ridge regression and inverse problems: Tech. rep.: Stockholm University, Sweden, 2001.
- [40] *Tarantola A.* Inverse Problem Theory and Methods for Model Parameter Estimation. — SIAM, 2005. — 358 pp.
- [41] *Komarek P.* Logistic regression for data mining and high-dimensional classification: Tech. rep. — Pittsburgh, PA: Robotics Institute, Carnegie Mellon University, May 2004.
- [42] *Hosmer D. W., Lemeshow S.* Applied Logistic Regression. — Wiley, 2000. — 373 pp.
- [43] *Bernardo J. M., Smith A. F. M.* Bayesian Theory. — Wiley, 1994. — 640 pp.
- [44] *Панюков А. В., Тырсин А. Н.* Взаимосвязь взвешенного и обобщенного вариантов метода наименьших модулей // *Известия Челябинского научного центра.* — 2007. — Т. 1, № 35. — С. 6–11.
- [45] *Тырсин А.* Исследование свойств обобщенного метода наименьших модулей (на примере оценки параметра сдвига) // *Заводская лаборатория. Диагностика материалов.* — 2007. — Т. 73, № 11. — С. 71–76.
- [46] *Тырсин А.* Об эквивалентности знакового и наименьших модулей методов построения линейных моделей // *Обозрение прикладной и промышленной математики.* — 2005. — Т. 12, № 4. — С. 879–880.
- [47] *Воронцов К. В.* Комбинаторная теория надёжности обучения по прецедентам: Дис. док. физ.-мат. наук.: Ph.D. thesis / Вычислительный центр РАН. — 2010.
- [48] *Воронцов К. В.* Комбинаторные обоснования обучаемых алгоритмов // *Журнал вычислительной математики и математической физики.* — 2004. — Т. 44 (11). — С. 2099–2112.
- [49] *Madala H., Ivakhnenko A.* Inductive learning algorithms for complex systems modelling. — Boca Raton, Florida, USA, CRC Press, 1994. — 368 pp.
- [50] Group method for data handling. — <http://www.gmdh.net>. — 2000.
- [51] *Ивахненко А. Г.* Индуктивный метод самоорганизации моделей сложных систем. — Киев: Наукова думка, 1981. — 296 с.
- [52] *Ивахненко А. Г., Степашко В. С.* Помехоустойчивость моделирования. — Киев: Наукова думка, 1985. — 216 с.
- [53] *Ивахненко А. Г., Юрачковский Ю. П.* Моделирование сложных систем по экспериментальным данным. — М.: Радио и связь, 1987. — 120 с.