

# Методы машинного обучения. Нелинейная регрессия, обобщённые линейные модели, нестандартные функции потерь

Воронцов Константин Вячеславович  
[www.MachineLearning.ru/wiki?title=User:Vokov](http://www.MachineLearning.ru/wiki?title=User:Vokov)  
вопросы к лектору: [voron@forecsys.ru](mailto:voron@forecsys.ru)

материалы курса:  
[github.com/MSU-ML-COURSE/ML-COURSE-21-22](https://github.com/MSU-ML-COURSE/ML-COURSE-21-22)  
орг.вопросы по курсу: [ml.cmc@mail.ru](mailto:ml.cmc@mail.ru)

## 1 Нелинейная регрессия

- Нелинейная модель регрессии
- Логистическая регрессия
- Обобщённая аддитивная модель

## 2 Обобщённая линейная модель

- В каких случаях нельзя использовать МНК
- Экспоненциальное семейство распределений
- Максимизация правдоподобия для GLM

## 3 Неквадратичные функции потерь

- Квантильная регрессия
- Робастная регрессия
- SVM-регрессия

## Нелинейная модель регрессии

$X^\ell = (x_i, y_i)_{i=1}^\ell$  — обучающая выборка,  $x_i \in \mathbb{R}^n$ ,  $y_i \in \mathbb{R}$   
 $y_i = y(x_i)$ ,  $y: X \rightarrow Y$  — неизвестная регрессионная зависимость  
 $f(x, \alpha)$  — нелинейная модель регрессии,  $\alpha \in \mathbb{R}^p$

**Метод наименьших квадратов (МНК):**

$$Q(\alpha, X^\ell) = \sum_{i=1}^{\ell} (f(x_i, \alpha) - y_i)^2 \rightarrow \min_{\alpha}.$$

**Метод Ньютона–Рафсона:**

1. Начальное приближение  $\alpha^0 = (\alpha_1^0, \dots, \alpha_p^0)$ .
2. Итерационный процесс

$$\alpha^{t+1} := \alpha^t - h_t (Q''(\alpha^t))^{-1} Q'(\alpha^t),$$

$Q'(\alpha^t)$  — градиент функционала  $Q$  в точке  $\alpha^t$ , вектор из  $\mathbb{R}^p$   
 $Q''(\alpha^t)$  — гессиан функционала  $Q$  в точке  $\alpha^t$ , матрица из  $\mathbb{R}^{p \times p}$   
 $h_t$  — величина шага (можно полагать  $h_t = 1$ ).

## Метод Ньютона-Рафсона

Компоненты градиента:

$$\frac{\partial Q(\alpha)}{\partial \alpha_j} = 2 \sum_{i=1}^{\ell} (f(x_i, \alpha) - y_i) \frac{\partial f(x_i, \alpha)}{\partial \alpha_j}.$$

Компоненты гессиана:

$$\frac{\partial^2 Q(\alpha)}{\partial \alpha_j \partial \alpha_k} = 2 \sum_{i=1}^{\ell} \frac{\partial f(x_i, \alpha)}{\partial \alpha_j} \frac{\partial f(x_i, \alpha)}{\partial \alpha_k} - \underbrace{2 \sum_{i=1}^{\ell} (f(x_i, \alpha) - y_i) \frac{\partial^2 f(x_i, \alpha)}{\partial \alpha_j \partial \alpha_k}}_{\text{при линейризации полагается} = 0}.$$

Не хотелось бы обращаться гессиан на каждой итерации...

**Линеаризация**  $f(x_i, \alpha)$  в окрестности текущего  $\alpha^t$ :

$$f(x_i, \alpha) = f(x_i, \alpha^t) + \sum_{j=1}^p \frac{\partial f(x_i, \alpha_j^t)}{\partial \alpha_j} (\alpha_j - \alpha_j^t) + o(\alpha_j - \alpha_j^t).$$

## Метод Ньютона-Гаусса

Матричные обозначения:

$F_t = \left( \frac{\partial f}{\partial \alpha_j}(x_i, \alpha^t) \right)_{\ell \times p}$  — матрица первых производных;

$f_t = (f(x_i, \alpha^t))_{\ell \times 1}$  — вектор значений  $f$ .

Формула  $t$ -й итерации метода Ньютона-Гаусса:

$$\alpha^{t+1} := \alpha^t - \underbrace{h_t (F_t^T F_t)^{-1} F_t^T}_{\beta} (f_t - y).$$

$\beta$  — это решение задачи многомерной линейной регрессии

$$\|F_t \beta - (f_t - y)\|^2 \rightarrow \min_{\beta}.$$

Нелинейная регрессия сведена к серии линейных регрессий.

Скорость сходимости — как и у метода Ньютона-Рафсона, но для вычислений можно применять линейные методы.

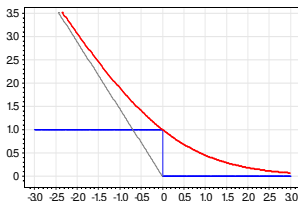
## Задача классификации. Логистическая регрессия

$Y = \{-1, +1\}$  — два класса,  $a(x, w) = \text{sign}(w^T x)$ ,  $x, w \in \mathbb{R}^n$ .

Функционал аппроксимированного эмпирического риска:

$$Q(w) = \sum_{i=1}^{\ell} [M_i(w) < 0] \leq \sum_{i=1}^{\ell} \mathcal{L}(w^T x_i y_i) \rightarrow \min_w,$$

где  $\mathcal{L}(M) = \log(1 + e^{-M})$  — логарифмическая функция потерь



$$M_i = w^T x_i y_i$$

## Метода Ньютона-Рафсона

Метода Ньютона-Рафсона для минимизации функционала  $Q(w)$ :

$$w^{t+1} := w^t - h_t(Q''(w^t))^{-1} Q'(w^t),$$

Элементы градиента — вектора первых производных  $Q'(w^t)$ :

$$\frac{\partial Q(w)}{\partial w_j} = - \sum_{i=1}^{\ell} (1 - \sigma_i) y_i f_j(x_i), \quad j = 1, \dots, n.$$

Элементы гессиана — матрицы вторых производных  $Q''(w^t)$ :

$$\frac{\partial^2 Q(w)}{\partial w_j \partial w_k} = \sum_{i=1}^{\ell} (1 - \sigma_i) \sigma_i f_j(x_i) f_k(x_i), \quad j, k = 1, \dots, n,$$

где  $\sigma_i = \sigma(y_i w^T x_i)$ ,  $\sigma(z) = \frac{1}{1+e^{-z}}$  — сигмоидная функция.

## Снова сведение к задаче линейной регрессии

В матричных обозначениях  $F = (f_j(x_i))_{\ell \times n}$ ,  $D = \text{diag}((1 - \sigma_i)\sigma_i)$

$$(Q''(w))^{-1} Q'(w) = -(F^T D F)^{-1} F^T \left( \frac{y_i}{\sigma_i} \right).$$

Это совпадает с МНК-решением линейной задачи регрессии со взвешенными объектами и модифицированными ответами:

$$Q(w) = \sum_{i=1}^{\ell} (1 - \sigma_i) \sigma_i \left( w^T x_i - \frac{y_i}{\sigma_i} \right)^2 \rightarrow \min_w.$$

Интерпретация:

- $\sigma_i = P(y_i | x_i)$  — вероятность правильной классификации  $x_i$
- чем ближе  $x_i$  к границе, тем больше вес  $(1 - \sigma_i)\sigma_i$
- чем выше вероятность ошибки, тем больше  $\frac{1}{\sigma_i}$

**ВЫВОД:** на каждой итерации происходит более точная настройка на «наиболее трудных» объектах.



## МНК с итерационным перевзвешиванием объектов IRLS — Iteratively Reweighted Least Squares

**Вход:**  $F, y$  — матрица «объекты–признаки» и вектор ответов;

**Выход:**  $w$  — вектор коэффициентов линейной комбинации.

---

- 1:  $w := (F^T F)^{-1} F^T y$  — нулевое приближение, обычный МНК;
- 2: **для**  $t := 1, 2, 3, \dots$
- 3:    $\sigma_i = \sigma(y_i w^T x_i)$  для всех  $i = 1, \dots, \ell$ ;
- 4:    $\gamma_i := \sqrt{(1 - \sigma_i) \sigma_i}$  для всех  $i = 1, \dots, \ell$ ;
- 5:    $\tilde{F} := \text{diag}(\gamma_1, \dots, \gamma_\ell) F$ ;
- 6:    $\tilde{y}_i := y_i \sqrt{(1 - \sigma_i) / \sigma_i}$  для всех  $i = 1, \dots, \ell$ ;
- 7:   выбрать градиентный шаг  $h_t$ ;
- 8:    $w := w + h_t (\tilde{F}^T \tilde{F})^{-1} \tilde{F}^T \tilde{y}$ ;
- 9:   **если**  $\{\sigma_i\}$  мало изменились **то** выйти из цикла;

## Обобщённая аддитивная модель (Generalized Additive Model)

Регрессия с нелинейными преобразованиями признаков  $\varphi_j$ :

$$f(x, \alpha) = \sum_{j=1}^n \varphi_j(f_j(x), \alpha_j)$$

В частности, при  $\varphi_j(f_j(x), \alpha_j) = \alpha_j f_j(x)$  это линейная модель

**Идея 1:** поочерёдно уточнять  $\varphi_j$  по выборке  $(f_j(x_i), z_i)_{i=1}^{\ell}$ :

$$\sum_{i=1}^{\ell} \left( \varphi_j(f_j(x_i), \alpha_j) - \underbrace{\left( y_i - \sum_{k \neq j} \varphi_k(f_k(x_i), \alpha_k) \right)}_{z_i} \right)^2 + \tau R(\alpha_j) \rightarrow \min_{\alpha_j}$$

**Идея 2:** постепенно уменьшать  $\tau$  у регуляризатора гладкости

$$R(\alpha_j) = \int (\varphi_j''(\zeta, \alpha_j))^2 d\zeta$$

В качестве  $\varphi_j$  использовать сплайны или ядерное сглаживание

## Метод backfitting [Хасты, Тибширани, 1986]

**Вход:**  $F, y$  — матрица «объекты–признаки» и вектор ответов;

**Выход:**  $\varphi_j(f_j, \alpha_j)$  — обучаемые преобразования признаков.

1: начальное приближение:

$\alpha :=$  решение задачи МЛР с признаками  $f_j(x)$ ;

$\varphi_j(f_j, \alpha_j) := \alpha_j f_j(x), \quad j = 1, \dots, n$ ;

2: **повторять**

3: **для**  $j = 1, \dots, n$

4: 
$$z_i := y_i - \sum_{k=1, k \neq j}^n \varphi_k(f_k(x_i), \alpha_k), \quad i = 1, \dots, \ell;$$

5: 
$$\alpha_j := \arg \min_{\alpha} \sum_{i=1}^{\ell} (\varphi(f_j(x_i), \alpha) - z_i)^2 + \tau R(\alpha);$$

6: уменьшить коэффициент регуляризации  $\tau$ ;

7: **пока**  $Q(\alpha, X^{\ell})$  и/или  $Q(\alpha, X^k)$  заметно уменьшаются;

## Связь МНК с методом максимума правдоподобия

Модель данных с некоррелированным гауссовским шумом:

$$y_i = f(x_i, \alpha) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_i^2), \quad i = 1, \dots, \ell.$$

Эквивалентная запись:  $y_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ ,  $\mu_i = \mathbb{E}y_i = f(x_i, \alpha)$ .

МНК эквивалентен методу максимума правдоподобия (ММП):

$$L(\varepsilon_1, \dots, \varepsilon_\ell | \alpha) = \prod_{i=1}^{\ell} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_i^2} \varepsilon_i^2\right) \rightarrow \max_{\alpha};$$

$$-\ln L(\varepsilon_1, \dots, \varepsilon_\ell | \alpha) = \text{const}(\alpha) + \frac{1}{2} \sum_{i=1}^{\ell} \frac{1}{\sigma_i^2} (f(x_i, \alpha) - y_i)^2 \rightarrow \min_{\alpha};$$

Как использовать линейные модели, если  $y_i$  не гауссовские, в частности, если  $y_i$  дискретнозначные?

## Обобщённая линейная модель (Generalized Linear Model, GLM)

*Нормальная линейная модель для математического ожидания:*

$$y_i \sim \mathcal{N}(\mu_i, \sigma_i^2), \quad \mu_i = x_i^\top \alpha = \mathbb{E}y_i$$

*Обобщённая линейная модель для математического ожидания:*

$$y_i \sim \text{Exp}(\theta_i, \phi_i), \quad \theta_i = x_i^\top \alpha = g(\mathbb{E}y_i)$$

**Exp** — экспоненциальное семейство распределений

с параметрами  $\theta_i$ ,  $\phi_i$  и параметрами-функциями  $c(\theta)$ ,  $h(y, \phi)$ :

$$p(y_i | \theta_i, \phi_i) = \exp\left(\frac{y_i \theta_i - c(\theta_i)}{\phi_i} + h(y_i, \phi_i)\right)$$

Математическое ожидание и дисперсия с.в.  $y_i \sim \text{Exp}(\theta_i, \phi_i)$ :

$$\mu_i = \mathbb{E}y_i = c'(\theta_i) \Rightarrow \theta_i = g(\mu_i) = [c']^{-1}(\mu_i)$$

$$\text{D}y_i = \phi_i c''(\theta_i)$$

**$g(\mu)$**  — монотонная функция связи (link function)

## Примеры распределений из экспоненциального семейства

Нормальное (гауссовское) распределение,  $y_i \in \mathbb{R}$ :

$$\begin{aligned} p(y_i | \mu_i, \sigma_i^2) &= \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2\sigma_i^2}(y_i - \mu_i)^2\right) = \\ &= \exp\left(\frac{y_i\mu_i - \frac{1}{2}\mu_i^2}{\sigma_i^2} - \frac{y_i^2}{2\sigma_i^2} - \frac{1}{2}\ln(2\pi\sigma_i^2)\right); \end{aligned}$$

$$\theta_i = g(\mu_i) = \mu_i, \quad c(\theta_i) = \frac{1}{2}\mu_i^2 = \frac{1}{2}\theta_i^2, \quad \phi_i = \sigma_i^2.$$

Распределение Бернулли,  $y_i \in \{0, 1\}$ :

$$p(y_i | \mu_i) = \mu_i^{y_i} (1 - \mu_i)^{1-y_i} = \exp\left(y_i \ln \frac{\mu_i}{1-\mu_i} + \ln(1 - \mu_i)\right);$$

$$\theta_i = g(\mu_i) = \ln \frac{\mu_i}{1-\mu_i}, \quad c(\theta_i) = -\ln(1 - \mu_i) = \ln(1 + e^{\theta_i}).$$

## Примеры распределений из экспоненциального семейства

Биномиальное распределение,  $y_i \in \{0, 1, \dots, n_i\}$ :

$$\begin{aligned} p(y_i | \mu_i, n_i) &= C_{n_i}^{y_i} \left( \frac{\mu_i}{n_i} \right)^{y_i} \left( 1 - \frac{\mu_i}{n_i} \right)^{n_i - y_i} = \\ &= \exp \left( y_i \ln \frac{\mu_i}{n_i - \mu_i} + n_i \ln(n_i - \mu_i) + \ln C_{n_i}^{y_i} - n_i \ln n_i \right); \end{aligned}$$

$$\theta_i = g(\mu_i) = \ln \frac{\mu_i}{n_i - \mu_i}, \quad c(\theta_i) = -n_i \ln(n_i - \mu_i) = n_i \ln \frac{1 + e^{\theta_i}}{n_i}.$$

Пуассоновское распределение,  $y_i \in \{0, 1, 2, \dots\}$ :

$$p(y_i | \mu_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} = \exp \left( \frac{y_i \ln(\mu_i) - \mu_i}{1} - \ln y_i! \right);$$

$$\theta_i = g(\mu_i) = \ln(\mu_i), \quad c(\theta_i) = \mu_i = e^{\theta_i}, \quad \phi_i = 1.$$

## Примеры распределений из экспоненциального семейства

- нормальное (гауссовское)
- распределение Пуассона
- биномиальное и мультиномиальное
- геометрическое
- $\chi^2$ -распределение
- бета-распределение
- гамма-распределение
- распределение Дирихле
- распределение Лапласа с фиксированным матожиданием

**Контр-примеры** не экспоненциальных распределений:

- $t$ -распределение Стьюдента, Коши, гипергеометрическое



## Максимизация правдоподобия для GLM

Принцип максимума правдоподобия:

$$L(\alpha) = \ln \prod_{i=1}^{\ell} p(y_i | \theta_i, \phi_i) = \sum_{i=1}^{\ell} \frac{y_i \theta_i - c(\theta_i)}{\phi_i} \rightarrow \max_{\alpha},$$

где  $\theta_i$  линейно зависит от  $\alpha$ :  $\theta_i = x_i^T \alpha = \sum_{j=1}^n \alpha_j f_j(x_i)$ .

Метод Ньютона-Рафсона:  $\alpha^{t+1} := \alpha^t + h_t (L''(\alpha^t))^{-1} L'(\alpha^t)$ .

Компоненты вектора градиента  $L'(\alpha)$ :

$$\frac{\partial L(\alpha)}{\partial \alpha_j} = \sum_{i=1}^{\ell} \frac{y_i - c'(x_i^T \alpha)}{\phi_i} f_j(x_i).$$

Компоненты матрицы Гессе  $L''(\alpha)$ :

$$\frac{\partial^2 L(\alpha)}{\partial \alpha_j \partial \alpha_k} = - \sum_{i=1}^{\ell} \frac{c''(x_i^T \alpha)}{\phi_i} f_j(x_i) f_k(x_i).$$

## Матричные обозначения

$F = (f_j(x_i))_{\ell \times n}$  — матрица «объекты–признаки»;

$\tilde{F} = W_t F$ ,  $W_t = \text{diag}\left(\sqrt{\frac{1}{\phi_i} c''(\theta_i)}\right)$  — веса объектов,  $\theta_i = x_i^T \alpha^t$ ;

$\tilde{y} = (\tilde{y}_i)_{\ell \times 1}$ ,  $\tilde{y}_i = \frac{y_i - c'(\theta_i)}{\sqrt{\phi_i c''(\theta_i)}}$  — модифицированный вектор ответов.

Тогда метод Ньютона-Рафсона снова приводит к IRLS:

$$\alpha^{t+1} := \alpha^t - h_t \underbrace{(F^T W_t W_t F)^{-1} F^T W_t}_{(\tilde{F}^T \tilde{F})^{-1} \tilde{F}^T} \underbrace{\left( \sqrt{\frac{\phi_i}{c''(\theta_i)}} \frac{y_i - c'(\theta_i)}{\phi_i} \right)}_{\tilde{y}_i}_{\ell \times 1}.$$

Это совпадает с МНК-решением линейной задачи регрессии со взвешенными объектами и модифицированными ответами:

$$Q(\alpha) = \|\tilde{F}\alpha - \tilde{y}\|^2 \rightarrow \min_{\alpha}.$$

## МНК с итерационным перевзвешиванием объектов IRLS — Iteratively Reweighted Least Squares

**Вход:**  $F, y$  — матрица «объекты–признаки» и вектор ответов;

**Выход:**  $w$  — вектор коэффициентов линейной комбинации.

- 
- 1:  $w := (F^T F)^{-1} F^T y$  — нулевое приближение, обычный МНК;
  - 2: **для**  $t := 1, 2, 3, \dots$
  - 3:  $\theta_i = x_i^T \alpha^t$  для всех  $i = 1, \dots, \ell$ ;
  - 4:  $\gamma_i := \sqrt{\frac{1}{\phi_i} c''(\theta_i)}$  для всех  $i = 1, \dots, \ell$ ;
  - 5:  $\tilde{F} := \text{diag}(\gamma_1, \dots, \gamma_\ell) F$ ;
  - 6:  $\tilde{y}_i := \frac{y_i - c'(\theta_i)}{\phi_i \gamma_i}$  для всех  $i = 1, \dots, \ell$ ;
  - 7: выбрать градиентный шаг  $h_t$ ;
  - 8:  $w := w + h_t (\tilde{F}^T \tilde{F})^{-1} \tilde{F}^T \tilde{y}$ ;
  - 9: **если**  $\{\theta_i\}$  мало изменились **то** выйти из цикла;

## Логистическая регрессия как частный случай GLM

Распределение Бернулли,  $y_i \in \{0, 1\}$ :  $p(y_i|\mu_i) = \mu_i^{y_i}(1 - \mu_i)^{1-y_i}$

$$\theta_i = g(\mu_i) = \ln \frac{\mu_i}{1-\mu_i} \quad \mu_i = g^{-1}(\theta_i) = \frac{1}{1+\exp(-\theta_i)} \equiv \sigma(\theta_i)$$

Апостериорная вероятность классов,  $\tilde{y}_i = 2y_i - 1 \in \{-1, +1\}$ :

$$\left. \begin{aligned} P(y_i=1|x_i) &= E y_i = \mu_i = \sigma(\theta_i) \\ P(y_i=0|x_i) &= 1 - \mu_i = \sigma(-\theta_i) \end{aligned} \right\} \quad p(y_i|x_i) = \sigma(\underbrace{\tilde{y}_i x_i^T \alpha}_{\text{margin}})$$

Максимум правдоподобия  $\iff$  минимум критерия log-loss:

$$\begin{aligned} \sum_{i=1}^{\ell} \ln p(y_i|x_i) &= \sum_{i=1}^{\ell} y_i \ln \mu_i + (1 - y_i) \ln(1 - \mu_i) \rightarrow \max_{\alpha} \\ &\iff \sum_{i=1}^{\ell} \ln(1 + \exp(-\tilde{y}_i x_i^T \alpha)) \rightarrow \min_{\alpha} \end{aligned}$$

## Логистическая регрессия как частный случай GLM

Распределение Бернулли,  $y_i \in \{0, 1\}$ :  $p(y_i|\mu_i) = \mu_i^{y_i}(1 - \mu_i)^{1-y_i}$

$$\theta_i = g(\mu_i) = \ln \frac{\mu_i}{1-\mu_i} \quad \mu_i = g^{-1}(\theta_i) = \frac{1}{1+\exp(-\theta_i)} \equiv \sigma(\theta_i)$$

Всего лишь из двух предположений:

- $y_i$  — бернуллиевские случайные величины с  $E y_i = \mu_i$ ;
- модель линейна и  $\mu_i$  монотонно зависит от  $\theta_i = x_i^T \alpha$ ;

вытекают все основные свойства логистической регрессии:

- логарифмическая функция потерь  $\ln(1 + \exp(-\tilde{y}_i x_i^T \alpha))$ ;
- сигмоидная функция связи  $p(y_i|x_i) = \sigma(\tilde{y}_i x_i^T \alpha)$ ;
- связь линейной модели с *отношением шансов* (odds ratio):

$$x_i^T \alpha = \theta_i = \ln \frac{\mu_i}{1 - \mu_i} = \ln \frac{P(y_i = 1|x_i)}{P(y_i = 0|x_i)}.$$

## Метод наименьших модулей

Функция потерь  $\mathcal{L}(\varepsilon_i)$ , где  $\varepsilon_i = (a(x_i) - y_i)$  — ошибка;

$Q = \sum_{i=1}^{\ell} \mathcal{L}(\varepsilon_i) \rightarrow \min_a$  — критерий обучения модели по выборке.

Метод наименьших квадратов,  $\mathcal{L}(\varepsilon) = \varepsilon^2$ :

$$\sum_{i=1}^{\ell} (a - y_i)^2 \rightarrow \min_a \Rightarrow a = \frac{1}{\ell} \sum_{i=1}^{\ell} y_i.$$

Метод наименьших модулей,  $\mathcal{L}(\varepsilon) = |\varepsilon|$ :

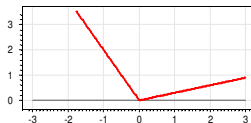
$$\sum_{i=1}^{\ell} |a - y_i| \rightarrow \min_a \Rightarrow a = \text{median}\{y_1, \dots, y_{\ell}\} = y^{(\ell/2)},$$

где  $y^{(1)}, \dots, y^{(\ell)}$  — вариационный ряд значений  $y_i$ .

Медиана более устойчива к редким большим выбросам  $y_i$ .

## Квантильная регрессия

$$\mathcal{L}(\varepsilon) = \begin{cases} C_+ |\varepsilon|, & \varepsilon > 0 \\ C_- |\varepsilon|, & \varepsilon < 0; \end{cases}$$



$$\sum_{i=1}^{\ell} \mathcal{L}(a - y_i) \rightarrow \min_a \Rightarrow a = y^{(q)}, \quad q = \frac{\ell C_-}{C_- + C_+}$$

где  $y^{(1)}, \dots, y^{(\ell)}$  — вариационный ряд значений  $y_i$

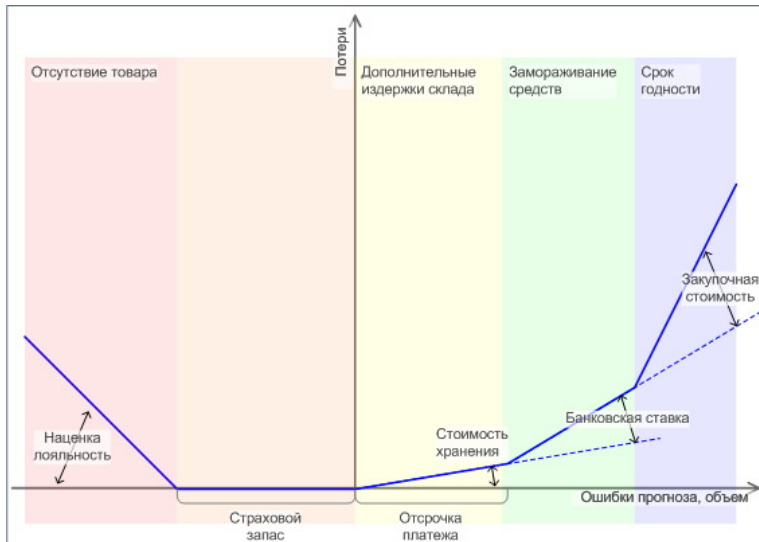
Линейная модель регрессии:  $a(x_i) = \langle x_i, w \rangle$ .

**Сведение к задаче линейного программирования:**

замена переменных  $\varepsilon_i^+ = (a(x_i) - y_i)_+$ ,  $\varepsilon_i^- = (y_i - a(x_i))_+$ ;

$$\begin{cases} Q = \sum_{i=1}^{\ell} C_+ \varepsilon_i^+ + C_- \varepsilon_i^- \rightarrow \min_w; \\ \langle x_i, w \rangle - y_i = \varepsilon_i^+ - \varepsilon_i^-; \quad \varepsilon_i^+ \geq 0; \quad \varepsilon_i^- \geq 0. \end{cases}$$

## Пример. Задача прогнозирования объёмов продаж

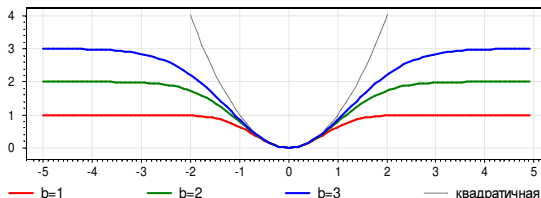




## Робастная регрессия — устойчивая к шумовым выбросам $y_i$

Модель регрессии:  $a(x) = f(x, \alpha)$

Функция Мешалкина:  $\mathcal{L}(\varepsilon) = b(1 - \exp(-\frac{1}{b}\varepsilon^2))$



Постановка задачи:

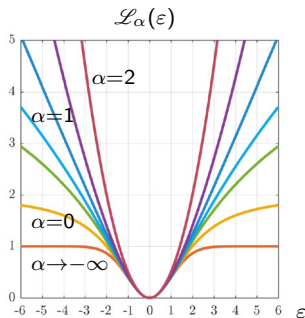
$$\sum_{i=1}^{\ell} \exp\left(-\frac{1}{b}(f(x_i, \alpha) - y_i)^2\right) \rightarrow \max_{\alpha}.$$

Эта задача также решается методом Ньютона-Рафсона.

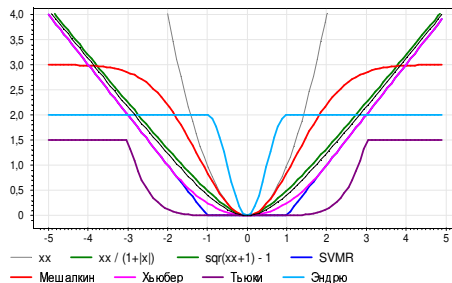
## Функции потерь для робастной регрессии

Семейство функций потерь Баррона с параметром  $\alpha$ :

$$\mathcal{L}_\alpha(\varepsilon) = \frac{|\alpha - 2|}{\alpha} \left( \left( \frac{\varepsilon^2}{|\alpha - 2|} + 1 \right)^{\alpha/2} - 1 \right)$$



другие примеры робастных  $\mathcal{L}(\varepsilon)$

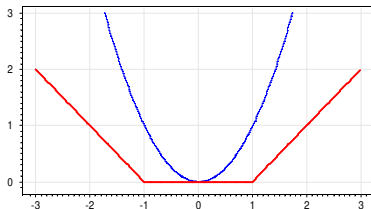


Jonathan T. Barron. A General and Adaptive Robust Loss Function. 2019.

## Регрессия опорных векторов SVMR (Support Vector Machine)

Модель регрессии:  $a(x) = \langle x, w \rangle - w_0$ ,  $w \in \mathbb{R}^n$ ,  $w_0 \in \mathbb{R}$ .

Функция потерь:  $\mathcal{L}(\varepsilon) = (|\varepsilon| - \delta)_+$



Постановка задачи:

$$\sum_{i=1}^{\ell} (|\langle w, x_i \rangle - w_0 - y_i| - \delta)_+ + \frac{1}{2C} \|w\|^2 \rightarrow \min_{w, w_0}.$$

Задача решается методами квадратичного программирования

- Нелинейная регрессия
  - сводится к последовательности линейных регрессий
  - метод Ньютона-Рафсона приводит к IRLS
- Логистическая регрессия
  - не регрессия, а классификация
  - метод Ньютона-Рафсона приводит к IRLS
- Обобщённая линейная модель (GLM)
  - мощно обобщает обычную и логистическую регрессию
  - метод Ньютона-Рафсона приводит к IRLS
- Обобщённая аддитивная регрессия (GAM, backfitting)
  - сводится к серии одномерных сглаживаний
- Неквадратичные функции потерь
  - проблемно-ориентированные (зависят от задачи)
  - в том числе робастная регрессия
  - приводят к разным методам, отличным от МНК
  - в некоторых случаях к методу Ньютона-Рафсона и IRLS