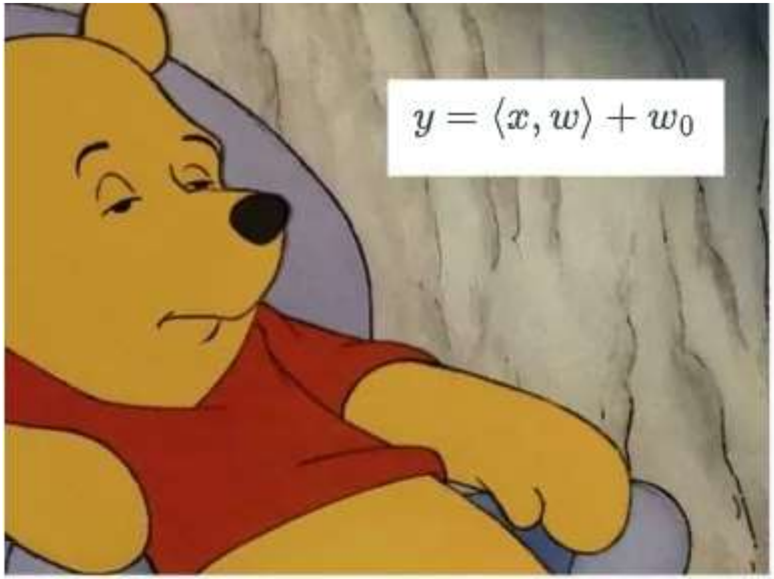
Тверской
государственный
технический
университет

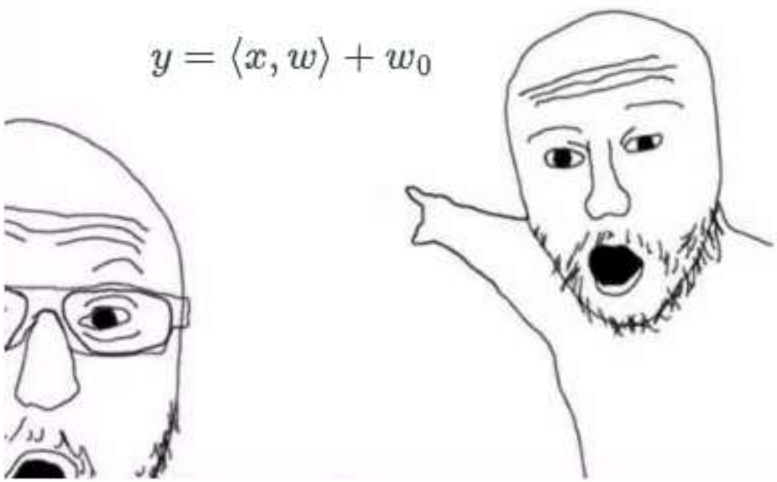
Интеллектуальные информационные системы

Линейные модели

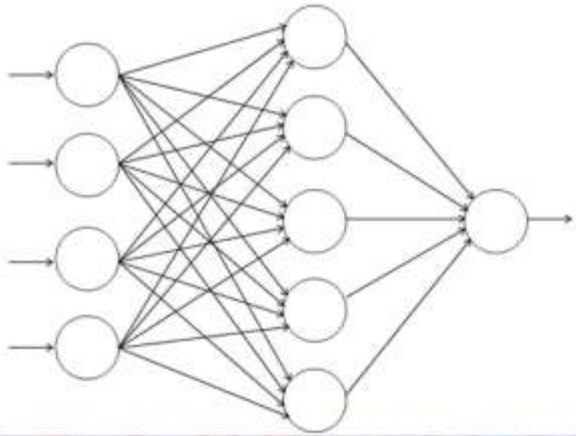
2025 г.



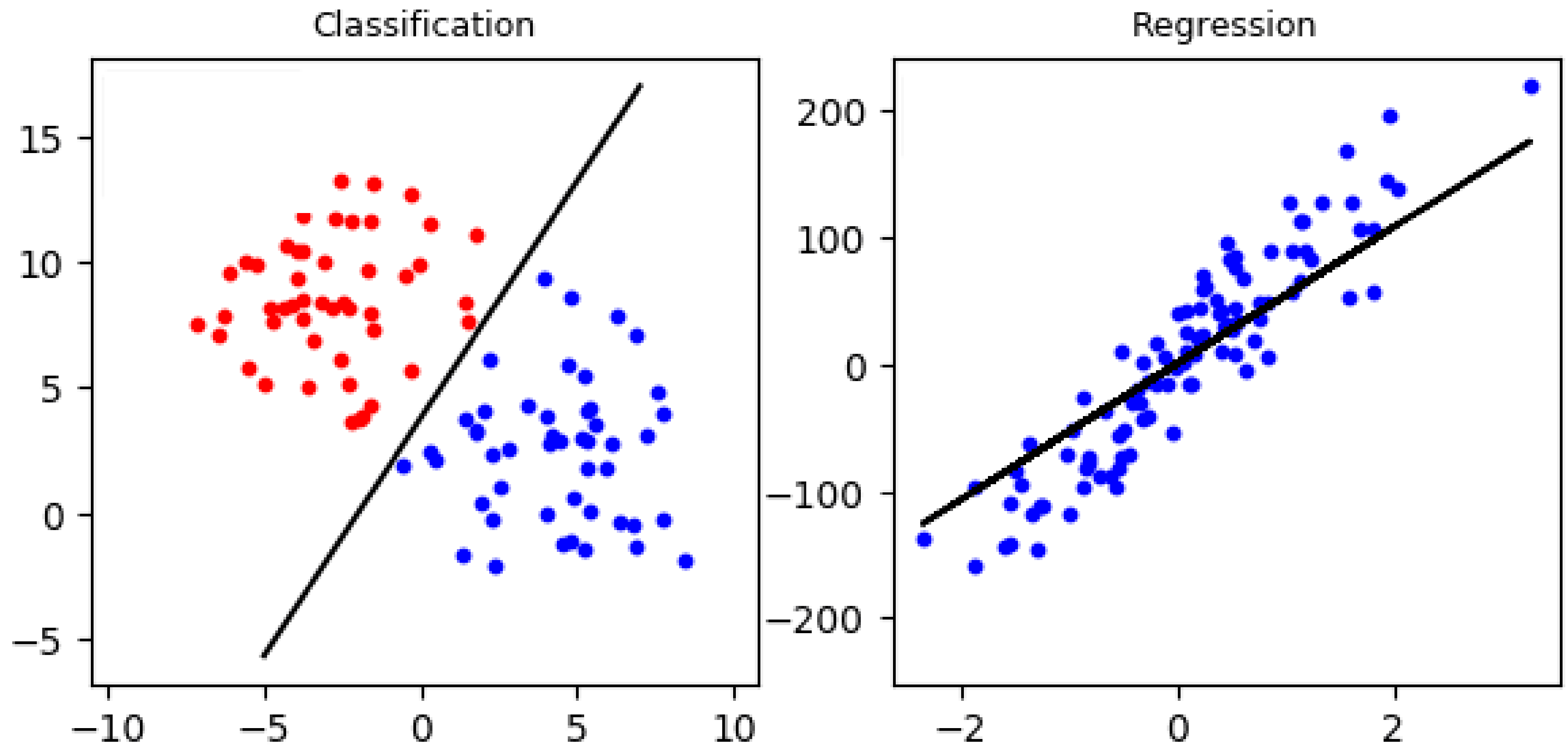
Employers when you tell them your app uses linear regression



Employers when you tell them your app uses “machine learning and A.I.”



$$y = \langle x, w \rangle + w_0$$


Classification
Regression
15
10
5
0
−5
−10
−5
0
5
10
200
100
0
−100
−200
−2
0
2

Линейная регрессия

Линейная регрессия – метод предсказания вещественного выходного значения (целевой переменной) $\boldsymbol{y} \in \mathbb{R}$ по вектору вещественных входных значений (признаков) $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{D}}$, при предположении, что ожидаемое выходное значение описывается линейной функцией входных значений

$$\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{w_0} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{d}} \boldsymbol{w_i x_i}$$

$\boldsymbol{w_0}$ - свободный коэффициент (сдвиг, bias)

$\boldsymbol{w_i}$ - параметры модели (коэффициенты, веса)

Обучение линейной регрессии

Задача оптимизации:

$$\frac{1}{l}\sum_{i=1}^{l}(\langle w, x_i\rangle - y_i)^2 \to \min_{w}$$

В матричном виде:

$$\frac{1}{l}\|Xw - y\|^2 \to \min_{w}$$

Решение:

$$w = \left(X^T X\right)^{-1} X^T y$$

Измерение ошибки в задачах регрессии

- $\boldsymbol{MSE(y,\widehat{y}) = \frac{1}{l}\sum_{i=1}^{l}(\widehat{y}_i - y_i)^2}$

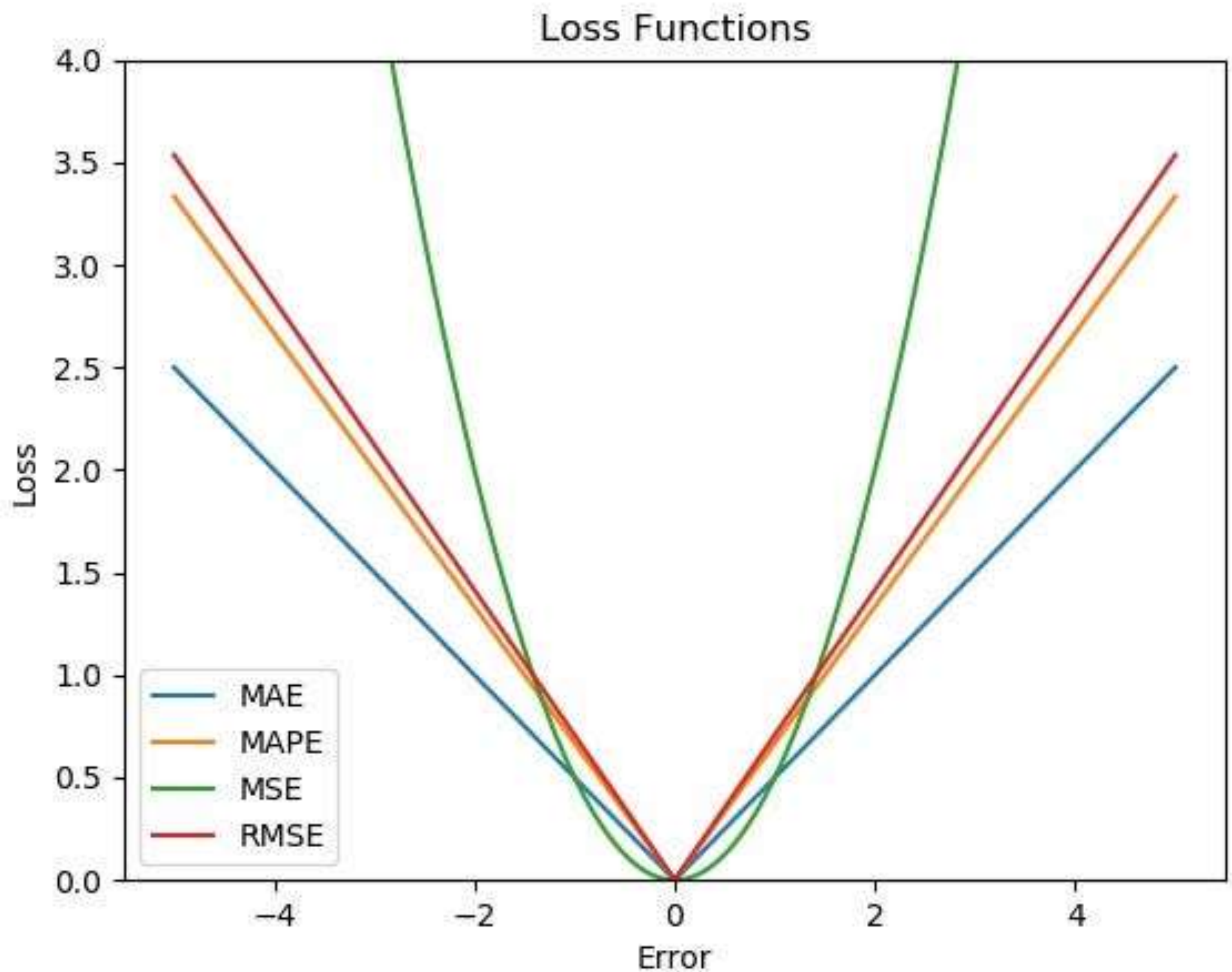
- $\boldsymbol{MAE(y,\widehat{y}) = \frac{1}{l}\sum_{i=1}^{l}|\widehat{y}_i - y_i|}$

- $\boldsymbol{MAPE(y,\widehat{y}) = \frac{1}{l}\sum_{i=1}^{l}\frac{|\widehat{y}_i - y_i|}{y_i} \times 100\%}$

- $\boldsymbol{RMSE(y,\widehat{y}) = \sqrt{\frac{1}{l}\sum_{i=1}^{l}(\widehat{y}_i - y_i)^2}}$

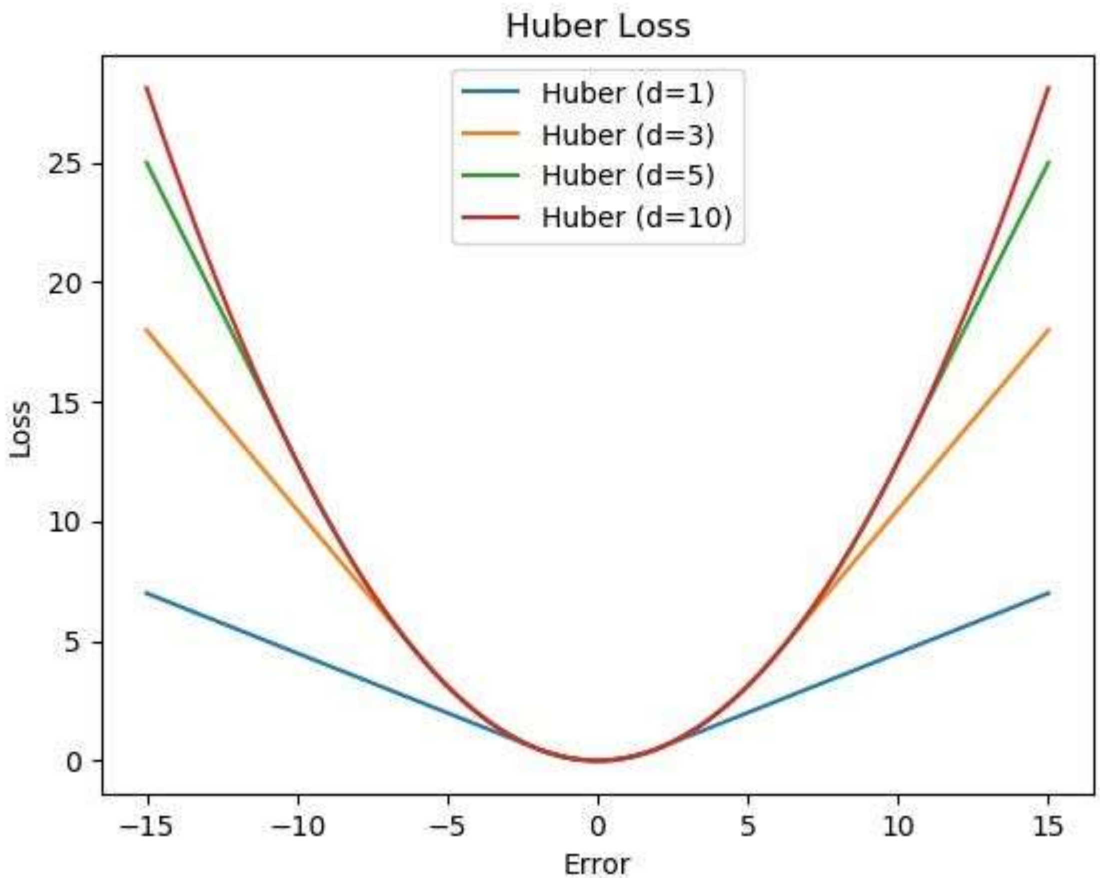
- $\boldsymbol{MSLE(y,\widehat{y}) = \frac{1}{l}\sum_{i=1}^{l}(log(1+\widehat{y}_i) - log(1+y_i))^2}$

Измерение ошибки в задачах регрессии



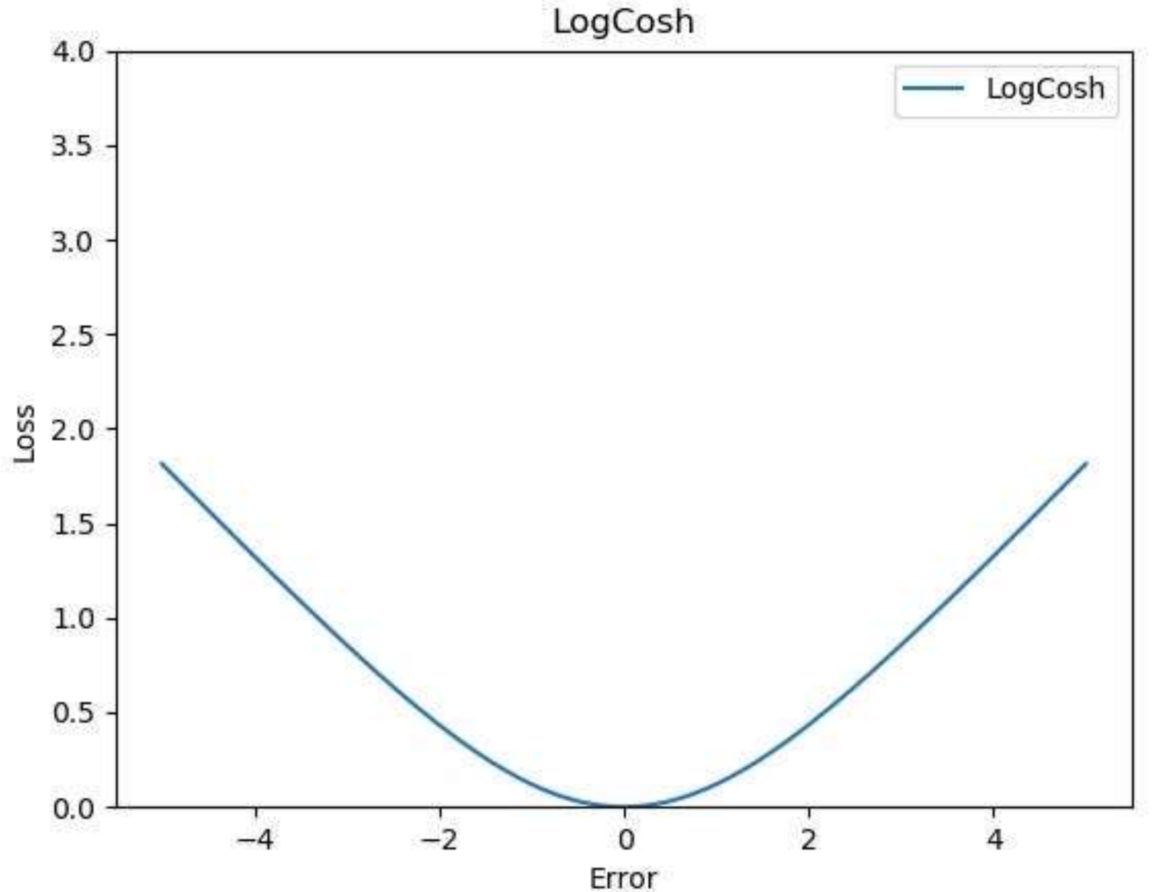
Измерение ошибки в задачах регрессии

$$\boldsymbol{Huber_{\delta}(y,\hat{y})} = \begin{cases} \frac{1}{2}(\hat{y}-y)^2, & |\hat{y} - y| < \delta \\ \delta\left(|\hat{y} - y| - \frac{1}{2}\delta\right), & |\hat{y} - y| \geq \delta \end{cases}$$



Измерение ошибки в задачах регрессии

$$\boldsymbol{LogCosh(y, \widehat{y}) = \sum_{i=1}^{l} log(cosh(\widehat{y}_i - y_i))}$$

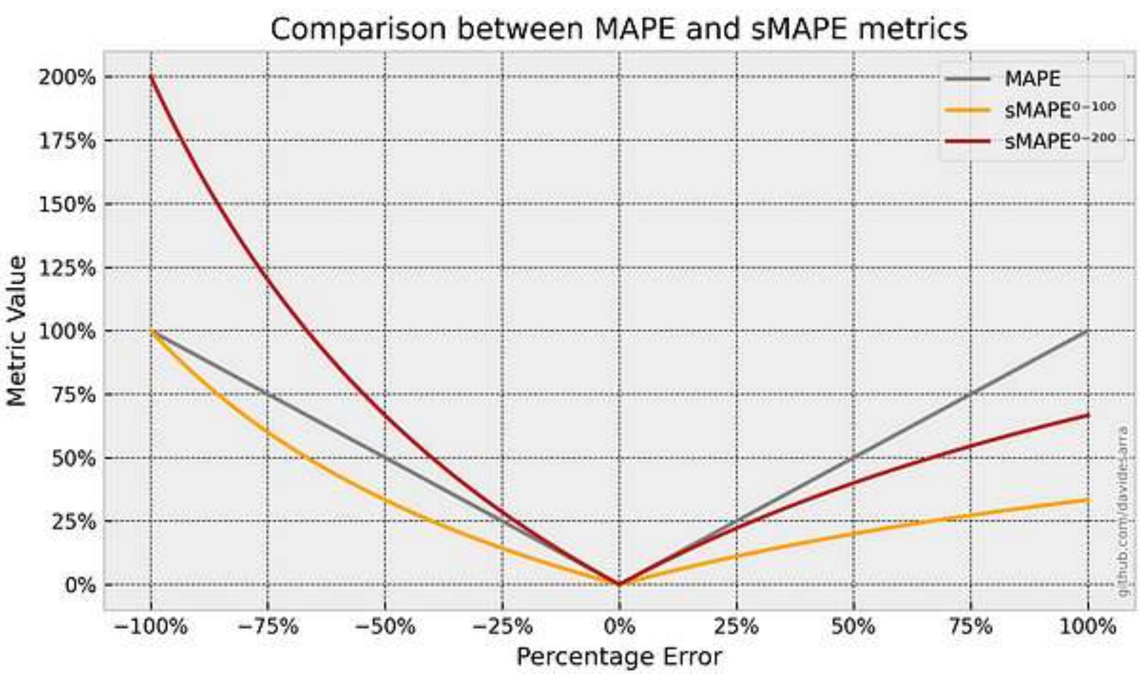


Измерение ошибки в задачах регрессии

Symmetric mean absolute percentage error:

$$\boldsymbol{SMAPE^{0-100}(y,\hat{y}) = \frac{100\%}{l}\sum_{i=1}^{l}\frac{|\hat{y}_i - y_i|}{(|y_i| + |\hat{y}_i|)}}$$

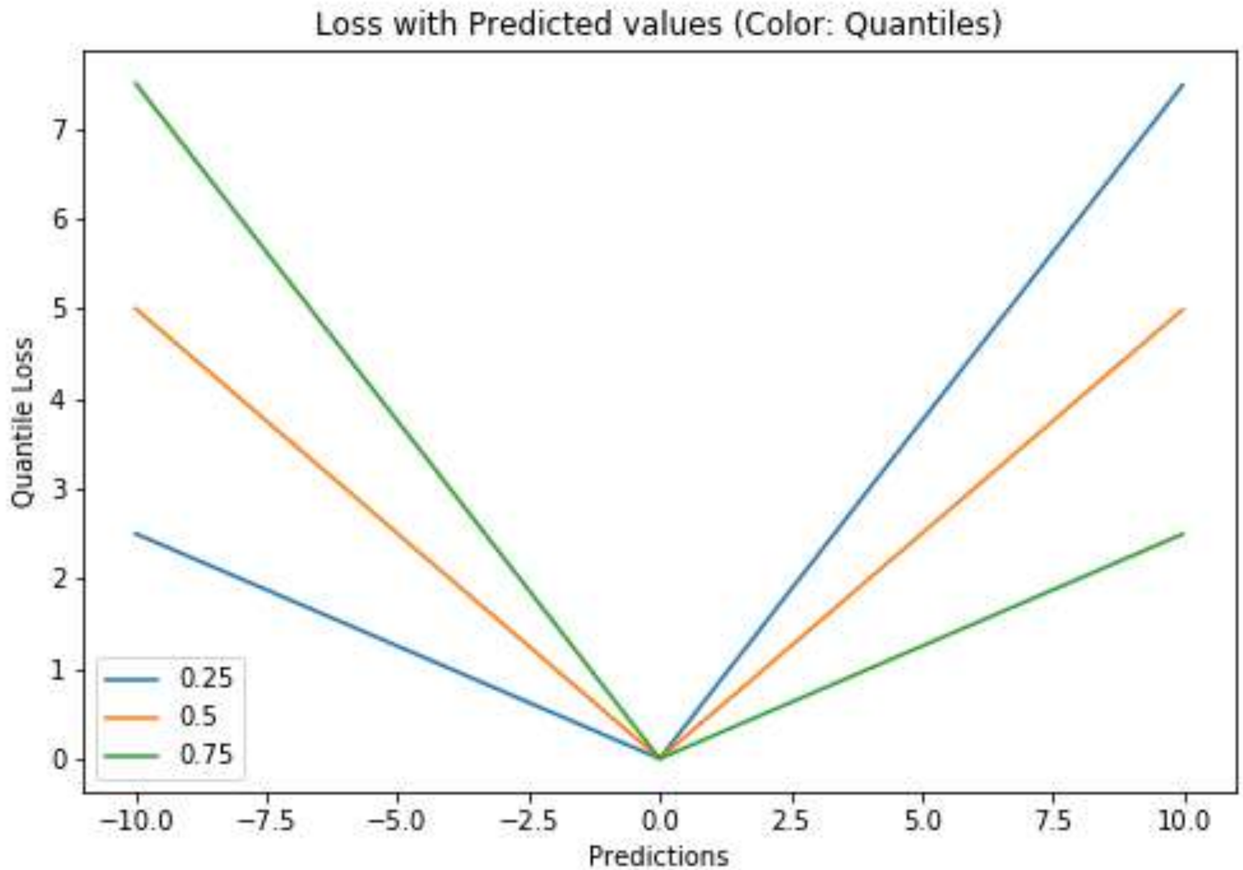
$$\boldsymbol{SMAPE^{0-200}(y,\hat{y}) = \frac{100\%}{l}\sum_{i=1}^{l}\frac{|\hat{y}_i - y_i|}{(|y_i| + |\hat{y}_i|)/2}}$$



Измерение ошибки в задачах регрессии

Quantile Loss

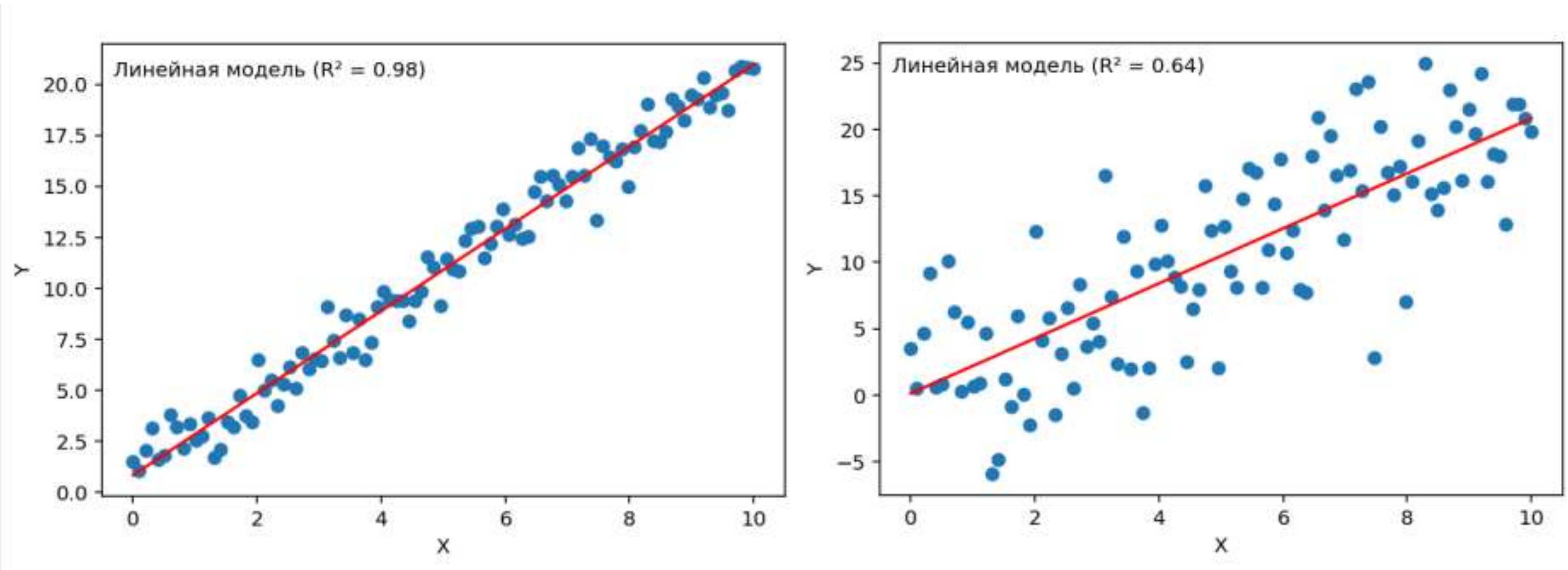
$$\boldsymbol{Q(y,\widehat{y}) = \frac{1}{l}\sum_{i=1}^{l}\rho_\tau(y_i - \widehat{y}_i)}$$



Измерение ошибки в задачах регрессии

Coefficient of determination

$$R^2(y,\widehat{y}) = 1 - \frac{\sum_{i=1}^{l}(\widehat{y}_i - y_i)^2}{\sum_{i=1}^{l}(y_i - \overline{y})^2}$$



Градиентные методы оптимизации

Градиент функции $f\ :\ \mathbb{R}^n \to \mathbb{R}$ - вектор частных производных

$$\nabla f(x_1, \dots, x_n) = \left(\frac{\partial f}{\partial x_j}\right)_{j=1}^{n}$$

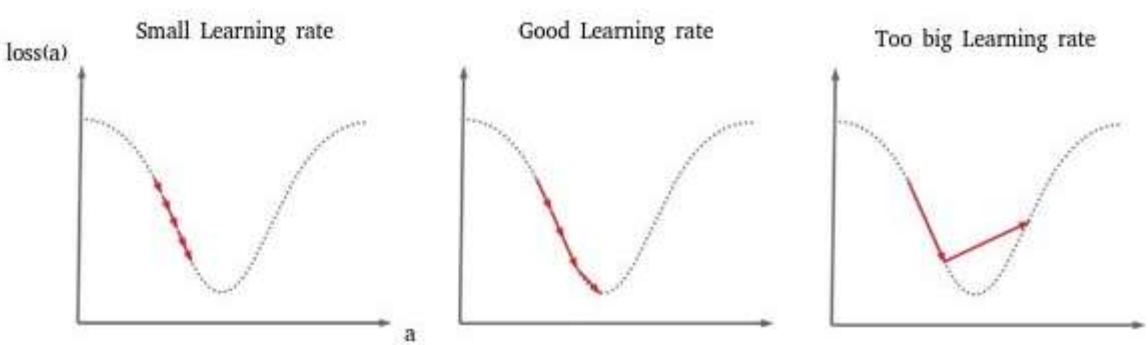
Градиентный шаг:

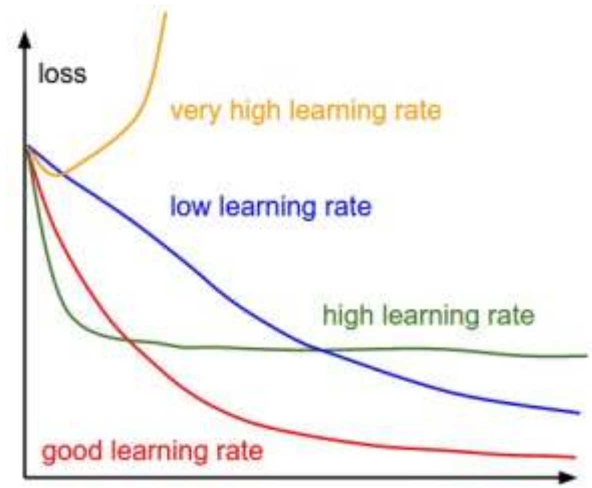
$$w^{(k)} = w^{(k-1)} - \alpha\, \nabla Q\left(w^{(k-1)}\right)$$

$w^{(0)}$- начальный набор параметров

$Q(w)$ - значение функционала ошибки для набора параметров w

α - learning rate (размер шага, темп обучения)





Оценивание градиента

- Полный градиент:

$$\nabla_w Q(w) = \frac{1}{l}\sum_{i=1}^{l} \nabla_w q_i(w)$$

- Стохастический градиентный спуск (SGD):

$$w^{(k)} = w^{(k-1)} - \alpha\, \nabla q_{i_k}\left(w^{(k-1)}\right)$$

- Средний стохастический градиент (SAG):

$$w^{(k)} = w^{(k-1)} - \alpha \frac{1}{l}\sum_{i=1}^{l} z_i^{(k)}$$

Источник: https://education.yandex.ru/handbook/ml/article/optimizaciya-v-ml

Модификации градиентного спуска

- Метод инерции (momentum):

$$\boldsymbol{w^{(k)} = w^{(k-1)} - h_k}$$

$$\boldsymbol{h_k = \beta\, h_{k-1} + \alpha \nabla_w Q\left(w^{(k-1)}\right)}$$

вектор инерции

параметр метода, определяющий скорость затухания градиентов с предыдущих шагов

- Nesterov momentum:

$$\boldsymbol{h_k = \beta\, h_{k-1} + \alpha \nabla_w Q\left(w^{(k-1)} + \beta\, h_{k-1}\right)}$$

Источник: https://education.yandex.ru/handbook/ml/article/optimizaciya-v-ml

Adaptive learning rate

- Adagrad

$$G_{k+1} = G_k + (\nabla f(x_k))^2$$
$$x_{k+1} = x_k - \frac{\alpha}{\sqrt{G_{k+1} + \varepsilon}} \nabla f(x_k).$$

- RMSProp

$$G_{k+1} = \gamma G_k + (1 - \gamma)(\nabla f(x_k))^2$$
$$x_{k+1} = x_k - \frac{\alpha}{\sqrt{G_{k+1} + \varepsilon}} \nabla f(x_k).$$

- Adam (ADAptive Momentum)

$$v_{k+1} = \beta_1 v_k + (1 - \beta_1) \nabla f(x_k)$$
$$G_{k+1} = \beta_2 G_k + (1 - \beta_2)(\nabla f(x_k))^2$$
$$x_{k+1} = x_k - \frac{\alpha}{\sqrt{G_{k+1} + \varepsilon}} v_{k+1}.$$

Регуляризация

- При наличии в выборке линейно зависимых (мультиколлинеарных) признаков всегда найдется такой вектор $\boldsymbol{v}$, что для любого объекта $\boldsymbol{x}$:

$$\langle \boldsymbol{v}, \boldsymbol{x} \rangle = \boldsymbol{0}$$

Множество решений:

$$\langle \boldsymbol{w} + \boldsymbol{\alpha v}, \boldsymbol{x} \rangle = \langle \boldsymbol{w}, \boldsymbol{x} \rangle + \boldsymbol{\alpha} \langle \boldsymbol{v}, \boldsymbol{x} \rangle = \langle \boldsymbol{w}, \boldsymbol{x} \rangle$$

тоже решение
оптимизационной задачи
для любого $\boldsymbol{\alpha}$

оптимальный вектор весов
(решение оптимизационной задачи)

Подмена задачи:

$$\boldsymbol{Q_\lambda(w) = Q(w) + \lambda R(w)}$$

коэффициент регуляризации

регуляризатор

- L_1-регуляризация:

$$\boldsymbol{R_1(w) = \|w\|_1 = \sum_{i=1}^{d} |w|_i}$$

- L_2-регуляризация:

$$\boldsymbol{R_2(w) = \|w\|_2 = \sum_{i=1}^{d} w^2}$$

- Elastic net regularization:

$$\boldsymbol{R_E(w) = R_1(w) + R_2(w)}$$

Преобразования признаков

Нелинейные преобразования исходного признакового пространства вида

$$\boldsymbol{x} = (\boldsymbol{x_1}, \dots, \boldsymbol{x_d}) \to \boldsymbol{\varphi}(\boldsymbol{x}) = (\boldsymbol{\varphi_1}(\boldsymbol{x}), \dots, \boldsymbol{\varphi_m}(\boldsymbol{x}))$$

- Полиномиальное (квадратичные или более высоких порядков): $\boldsymbol{\varphi}(\boldsymbol{x}) = (\boldsymbol{x_1}, \dots, \boldsymbol{x_d}, \boldsymbol{x_1^2}, \dots, \boldsymbol{x_d^2}, \boldsymbol{x_1 x_2}, \dots, \boldsymbol{x_{d-1}}, \boldsymbol{x_d})$
- Логарифмическое: $\boldsymbol{log}(\boldsymbol{x_i})$
- Экспоненциальное: $\boldsymbol{exp}(\frac{\|\boldsymbol{x}-\boldsymbol{\mu}\|^2}{\boldsymbol{\sigma}})$
- Синусоидальное: $\boldsymbol{sin}({}^{\boldsymbol{x_i}}/_{\boldsymbol{T}})$
- Любое другое, если вы можете обосновать и реализовать это преобразование

Масштабирование признаков

- Стандартизация (standard scaling):

$$\boldsymbol{z} = \frac{\boldsymbol{x} - \boldsymbol{\mu}}{\boldsymbol{\sigma}}$$

- Масштабирование к диапазону ***[a,b]*** (min-max normalization):

$$\boldsymbol{x}' = \boldsymbol{a} + \frac{(\boldsymbol{x} - \boldsymbol{min}(\boldsymbol{x}))(\boldsymbol{b} - \boldsymbol{a})}{\boldsymbol{max}(\boldsymbol{x}) - \boldsymbol{min}(\boldsymbol{x})}$$

- Robust scaling (standardization using median and IQR):

$$\boldsymbol{x}' = \frac{\boldsymbol{x} - \boldsymbol{Q}_{\boldsymbol{2}}(\boldsymbol{x})}{\boldsymbol{Q}_{\boldsymbol{3}}(\boldsymbol{x}) - \boldsymbol{Q}_{\boldsymbol{1}}(\boldsymbol{x})}$$