

Методы машинного обучения. Выявление аномалий и робастное обучение

Воронцов Константин Вячеславович
www.MachineLearning.ru/wiki?title=User:Vokov
вопросы к лектору: k.v.vorontsov@yandex.ru

материалы курса:
github.com/MSU-ML-COURSE/ML-COURSE-23-24
орг.вопросы по курсу: ml.cmc@mail.ru

- 1 Выявление аномальных объектов**
 - Эвристики для оценивания аномальности объектов
 - Отсев выбросов в непараметрической регрессии
 - Систематизация подходов
- 2 Теория робастного (помехоустойчивого) обучения**
 - Робастные функции потерь
 - Робастные агрегирующие функции
 - Методы итерационного взвешивания
- 3 Задачи с аномальными или новыми классами**
 - Одноклассовая классификация
 - Обучение по выборке одного класса
 - Задачи с новыми или неизвестными классами

Задачи выявления аномалий (Anomaly Detection)

Выявление выбросов (Outlier Detection)

- ошибки в данных обучающего или тестового объекта
- неадекватность модели на некоторых объектах

Выявление «новизны» (Novelty Detection)

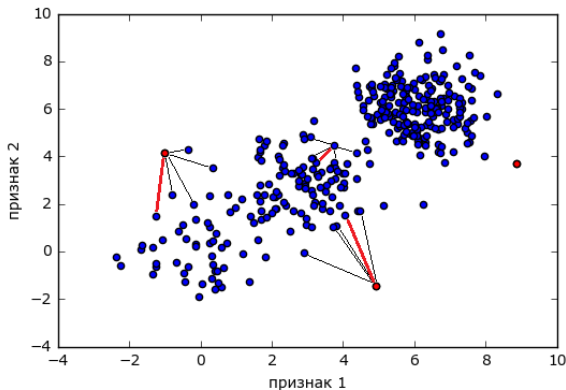
- ничего подобного не было в обучающей выборке

Примеры приложений

- обнаружение мошенничества (Fraud Detection)
- обнаружение вторжений (Intrusion Detection)
- обнаружение инсайдерской торговли на бирже
- обнаружение неполадок по показаниям датчиков
- медицинская диагностика (Medical Diagnosis)

Метрические методы

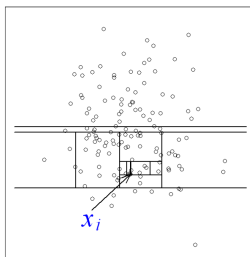
Аномальность объекта — расстояние до его k -го ближайшего соседа: чем больше, тем меньше локальная плотность выборки



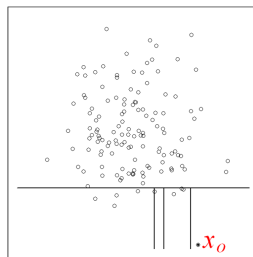
M.M.Breunig, H.-P.Kriegel, R.T.Ng, J.Sander. Local outlier factor: identifying density-based local outliers. 2000

Случайный изолирующий лес (Isolation Forest)

- Строится случайный лес деревьев
- Каждое ветвление: случайный признак и порог
- В каждом листе остаётся только один объект
- *Аномальность объекта* — средняя глубина листьев, в которые он попадает: чем меньше, тем более объект изолирован



глубина 12

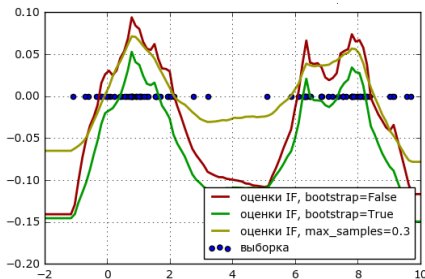


глубина 4

Fei Tony Liu, Kai Ming Ting, Zhi-Hua Zhou. Isolation Forest. 2008

Случайный изолирующий лес (IsolationForest)

- Строится случайный лес деревьев
- Каждое ветвление: случайный признак и порог
- В каждом листе остаётся только один объект
- *Аномальность объекта* — средняя глубина листьев, в которые он попадает: чем меньше, тем более объект изолирован



<https://dyakonov.org/2017/04/19/поиск-аномалий-anomaly-detection>

Разделение смеси распределений с фоновой компонентой

Порождающая модель смеси распределений:

$$p(x) = w_0 \varphi_0(x) + \sum_{j=1}^k w_j \varphi(x, \theta_j), \quad \sum_{j=0}^k w_j = 1, \quad w_j \geq 0,$$

Варианты задания **фонового распределения** $\varphi_0(x)$:

- равномерное на всём X
- гауссовское с огромной фиксированной дисперсией

Задача максимизации логарифма правдоподобия

$$L(w, \theta) = \ln \prod_{i=1}^{\ell} p(x_i) = \sum_{i=1}^{\ell} \ln \left(w_0 \varphi_0(x_i) + \sum_{j=1}^k w_j \varphi(x_i, \theta_j) \right) \rightarrow \max_{w, \theta}$$

Аномальность объекта x_i — вероятность $p(j=0|x_i)$ того, что он является фоновым, оценивается на E-шаге EM-алгоритма

Робастные автокодировщики (Robust AutoEncoder)

Автокодировщик реконструирует $\hat{x} = g(f(x))$ по исходным x :

$$\sum_{i=1}^{\ell} \|g(f(x_i, \alpha), \beta) - x_i\|^2 \rightarrow \min_{\alpha, \beta}$$

Аномальность объекта — неизвестный разреженный шум $(\varepsilon_i)_{i=1}^{\ell}$

Реконструируются незашумлённые объекты $\tilde{x}_i = x_i - \varepsilon_i$:

$$\sum_{i=1}^{\ell} \|g(f(x_i - \varepsilon_i, \alpha), \beta) - (x_i - \varepsilon_i)\|^2 + \lambda \sum_{i=1}^{\ell} \|\varepsilon_i\|_1 \rightarrow \min_{\alpha, \beta, \varepsilon}$$

Пример. Робастный метод главных компонент (Robust PCA):

$$\|F - GU^T - E\|^2 + \lambda \|E\|_1 \rightarrow \min_{G, U, E}$$

где GU^T — матрица низкого ранга, E — разреженная матрица

C.Zhou, R.C.Paffenroth. Anomaly detection with robust deep autoencoders. 2017.

E.J.Candès, X.Li, Y.Ma, J.Wright. Robust Principal Component Analysis. 2009.

Напоминание. Непараметрическая регрессия

Модель регрессии — константа $f(x, \alpha) = \alpha$ в окрестности x :

$$Q(\alpha; X^\ell) = \sum_{i=1}^{\ell} w_i(x) (\alpha - y_i)^2 \rightarrow \min_{\alpha \in \mathbb{R}}$$

где $w_i(x) = K\left(\frac{\rho(x, x_i)}{h}\right)$ — веса объектов x_i относительно x ;

$K(r)$ — ядро, невозрастающее, ограниченное, гладкое;

h — ширина окна сглаживания.

Формула ядерного сглаживания Надарая–Ватсона:

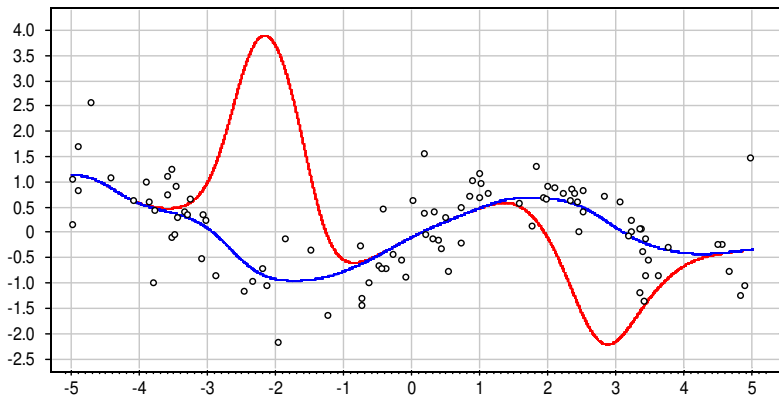
$$a_h(x; X^\ell) = \frac{\sum_{i=1}^{\ell} y_i w_i(x)}{\sum_{i=1}^{\ell} w_i(x)} = \frac{\sum_{i=1}^{\ell} y_i K\left(\frac{\rho(x, x_i)}{h}\right)}{\sum_{i=1}^{\ell} K\left(\frac{\rho(x, x_i)}{h}\right)}.$$

Проблема выбросов (эксперимент на синтетических данных)

$\ell = 100$, $h = 1.0$, гауссовское ядро $K(r) = \exp(-2r^2)$

Две из 100 точек — выбросы с ординатами $y_i = 40$ и -40

Синяя кривая — выбросов нет



Проблема выбросов и идея перевзвешивания объектов

Проблема выбросов: аномальные точки с большими случайными ошибками y_i сильно искажают функцию $a_h(x)$

Основная идея:

Аномальность объекта — LOO-ошибка $\varepsilon_i = |a_h(x_i; X^\ell \setminus x_i) - y_i|$.

Чем больше ε_i , тем меньше должен быть вес $w_i(x)$.

Повторять в итерациях: обучение, перерасчёт ошибок и весов.

Эвристика:

домножить веса $w_i(x)$ на коэффициенты $\gamma_i = \tilde{K}(\varepsilon_i)$,

где $\tilde{K}(r)$ — ядро, вообще говоря, отличное от $K(r)$.

Рекомендация:

брать квартическое ядро $\tilde{K}(\varepsilon) = K_Q\left(\frac{\varepsilon}{6 \operatorname{med}\{\varepsilon_i\}}\right)$,

где $\operatorname{med}\{\varepsilon_i\}$ — медиана вариационного ряда ошибок.

Gary W. Moran. Locally-Weighted-Regression Scatter-Plot Smoothing (LOWESS): a graphical exploratory data analysis technique. 1984

Алгоритм LOWESS (LOcally WEighted Scatter plot Smoothing)

Вход: X^ℓ — обучающая выборка;

Выход: коэффициенты γ_i , $i = 1, \dots, \ell$;

инициализация: $\gamma_i := 1$, $i = 1, \dots, \ell$;

повторять

оценки скользящего контроля в каждом объекте:

$$a_i := a_h(x_i; X^\ell \setminus \{x_i\}) = \frac{\sum_{j=1, j \neq i}^{\ell} y_j \gamma_j K\left(\frac{\rho(x_i, x_j)}{h(x_i)}\right)}{\sum_{j=1, j \neq i}^{\ell} \gamma_j K\left(\frac{\rho(x_i, x_j)}{h(x_i)}\right)}, \quad i = 1, \dots, \ell;$$

$$\gamma_i := \tilde{K}(|a_i - y_i|), \quad i = 1, \dots, \ell;$$

пока коэффициенты γ_i не стабилизируются;

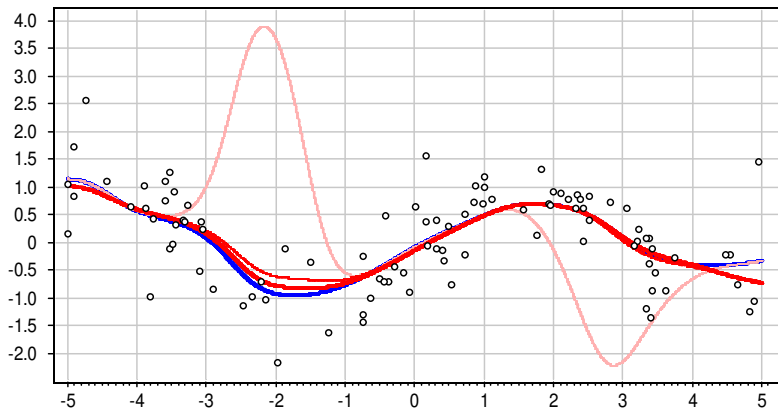
Gary W. Moran. Locally-Weighted-Regression Scatter-Plot Smoothing (LOWESS): a graphical exploratory data analysis technique. 1984

Пример работы LOWESS на синтетических данных

$\ell = 100$, $h = 1.0$, гауссовское ядро $K(r) = \exp(-2r^2)$

Две из 100 точек — выбросы с ординатами $y_i = 40$ и -40

В данном случае LOWESS сходится за несколько итераций:



Выявление новизны (novelty detection) и другие задачи

Аномальность объекта (anomaly/novelty/surprise score) — это значение функции потерь $\mathcal{L}(a(x_i), y_i)$ на данном объекте

Варианты оценивания аномальности:

- аномальность оценивается для объекта обучающей выборки (outlier) или для нового объекта (novelty)
- потеря зависит от y_i (supervised) или нет (unsupervised)
- при оценивании аномальности обучающего объекта он исключается из выборки ($a(x_i; X^\ell \setminus x_i)$) или нет ($a(x_i; X^\ell)$)
- функция потерь та же, что в критерии обучения или нет

Варианты использования оценок аномальности:

- жёсткое удаление аномальных объектов из выборки
- мягкое перевзвешивание весов объектов

M. Salehi et al. A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: solutions and future challenges. 2021.

Оптимизационные задачи машинного обучения

Постановки задач регрессии, классификации, кластеризации, восстановления плотности, снижения размерности и других отличаются функциями потерь $\mathcal{L}_i(\alpha)$ и регуляризацией $\tau R(\alpha)$:

$$Q(\alpha) = \sum_{i=1}^{\ell} w_i \mathcal{L}_i(\alpha) + \tau R(\alpha) \rightarrow \min_{\alpha}$$

Проблема: выбросы могут исказить $\mathcal{L}_i(\alpha)$ и критерий $Q(\alpha)$

Идея: уменьшать веса w_i выбросов с большими $\mathcal{L}_i(\alpha)$

$$Q(\alpha) = \sum_{i=1}^{\ell} \mu(\mathcal{L}_i(\alpha)) + \tau R(\alpha) \rightarrow \min_{\alpha}$$

$$\nabla Q(\alpha) = \sum_{i=1}^{\ell} \underbrace{\mu'(\mathcal{L}_i(\alpha))}_{w_i} \nabla \mathcal{L}_i(\alpha) + \tau \nabla R(\alpha) = 0$$

Итерационное взвешивание (Iterative Reweighting Scheme, IRS)

Пусть $\mu(r)$ — функция медленного роста:

$\mu(r) \geq 0$, $\mu'(r) \geq 0$, $\mu'(r)$ убывает, $\mu'(r) \rightarrow 0$ при $r \rightarrow +\infty$

Вход: $\mathcal{L}_i(\alpha)$ — функции потерь на обучающей выборке;

Выход: параметры модели α , веса объектов w_i ;

инициализация: $w_i := \frac{1}{\ell}$, $i = 1, \dots, \ell$;

повторять

$$\alpha := \arg \min_{\alpha} \sum_{i=1}^{\ell} w_i \mathcal{L}_i(\alpha) + \tau R(\alpha);$$

$$w_i := \text{norm}_i(\mu'(\mathcal{L}_i(\alpha))), \quad i = 1, \dots, \ell;$$

пока веса w_i не стабилизируются;

где $\text{norm}_i(v_i) = \frac{v_i}{\sum_j v_j}$ — операция нормирования вектора.

Недостаток: всё плохо, когда выбросы большие или их много

Итерационное взвешивание наименьших квадратов

(Iteratively Reweighted Least Squares, IRLS)

Робастная регрессия: $\mathcal{L}_i(\alpha) = |f(x_i, \alpha) - y_i|$ **Вход:** $(x_i, y_i)_{i=1}^{\ell}$ — обучающая выборка;**Выход:** параметры модели α , веса объектов w_i ;инициализация: $w_i := \frac{1}{\ell}$, $i = 1, \dots, \ell$;**повторять**

$$\alpha := \arg \min_{\alpha} \sum_{i=1}^{\ell} w_i \underbrace{(f(x_i, \alpha) - y_i)^2}_{\mathcal{L}_i^2(\alpha)} + \tau R(\alpha);$$

$$w_i := \operatorname{norm}_i \frac{\mu'(\mathcal{L}_i(\alpha))}{\mathcal{L}_i(\alpha)}, \quad i = 1, \dots, \ell;$$

пока веса w_i не стабилизируются;**Недостаток:** всё плохо, когда выбросы большие или их много

Функции потерь для робастной регрессии

Потеря $\mathcal{L}_i(\alpha) = \mu(r)$ — функция μ от ошибки $r = f(x_i, \alpha) - y_i$
Квадратичная функция потерь $\mu(r) = r^2$ — не робастная.

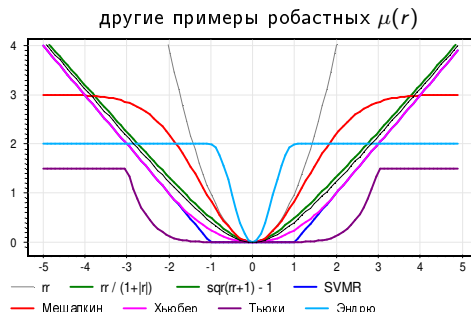
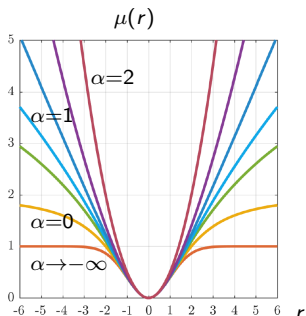
Робастные функции потерь $\mu(r)$, с параметром c :

- $\max(0, |r| - c)$ — кусочно-линейная (SVM-regression)
- $c(1 - \exp(-\frac{r^2}{2c}))$ — Мешалкина
- $\begin{cases} \frac{1}{2c}r^2, & |r| < c \\ |r| - \frac{c}{2}, & |r| \geq c \end{cases}$ — Хьюбера
- $\begin{cases} \frac{c^2}{6}(1 - (1 - \frac{r}{c})^2)^3, & |r| < c \\ \frac{c^2}{6}, & |r| \geq c \end{cases}$ — Тьюки
- $\begin{cases} (1 - \cos \frac{\pi r}{c}), & |r| < c \\ 2c, & |r| \geq c \end{cases}$ — Эндрю

Функции потерь для робастной регрессии

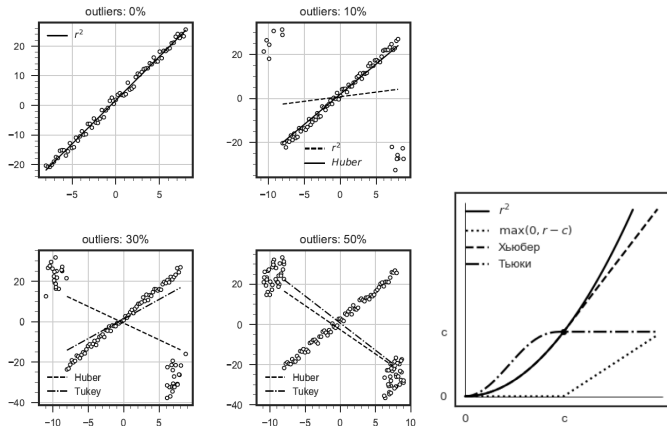
Семейство функций потерь Баррона с параметром α :

$$\mu(r) = \frac{|\alpha - 2|}{\alpha} \left(\left(\frac{r^2}{|\alpha - 2|} + 1 \right)^{\alpha/2} - 1 \right)$$



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

Пример. Робастная регрессия



Недостаток: всё плохо, когда выбросы большие или их много

З.М.Шибзухов. Методы машинного обучения на основе минимизации
сглаженных оценок средних, нечувствительных к выбросам. ММРО-2019.

Робастные (устойчивые к выбросам) способы усреднения

Среднее арифметическое (неустойчивое к большим выбросам):

$$\frac{1}{\ell} \sum_{i=1}^{\ell} z_i = \arg \min_u \sum_{i=1}^{\ell} (z_i - u)^2$$

Робастные (устойчивые) способы усреднения, определяемые через *вариационный ряд* $z^{(1)} \leq \dots \leq z^{(\ell)}$ значений z_1, \dots, z_{ℓ} :

- медиана $\frac{1}{2} (z^{(\lfloor \frac{\ell+1}{2} \rfloor)} + z^{(\lceil \frac{\ell+1}{2} \rceil)}) = \arg \min_u \sum_{i=1}^{\ell} |z_i - u|$
- γ -квантиль $z^{(\lfloor \gamma \ell \rfloor)} = \arg \min_u \sum_{i=1}^{\ell} |z_i - u| \cdot \begin{cases} \gamma, & z_i \geq u \\ 1-\gamma, & z_i < u \end{cases}$
- цензурированное среднее $\frac{1}{\ell} \sum_{i=1}^{\ell} \min(z_i, z^{(m)})$

Недостаток: эти функции усреднения недифференцируемы

Общий вид и свойства агрегирующих функций

Идея 1: среднее заменить одномерной минимизацией по u

Идея 2: затем модуль заменить его гладкой аппроксимацией

$$Q(\alpha) = M(\underbrace{\mathcal{L}_1(\alpha)}_{z_1}, \dots, \underbrace{\mathcal{L}_\ell(\alpha)}_{z_\ell}) = \arg \min_u \sum_{i=1}^{\ell} d(z_i - u)$$

Свойства функции несходства (dissimilarity function) $d(r)$:

- строго выпуклая, $d(r) \geq 0$, $d(0) = 0$

Свойства агрегирующей функции $M(z_1, \dots, z_\ell)$:

- $M(z_1) = z_1$
- монотонность: $z_i \leq z'_i \rightarrow M(z_1, \dots, z_\ell) \leq M(z'_1, \dots, z'_\ell)$
- симметричность: $M(z_1, \dots, z_\ell) = M(z_{\pi(1)}, \dots, z_{\pi(\ell)})$ для $\forall \pi$
- $\min(z_1, \dots, z_\ell) \leq M(z_1, \dots, z_\ell) \leq \max(z_1, \dots, z_\ell)$

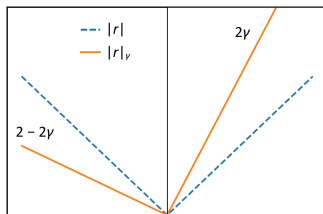
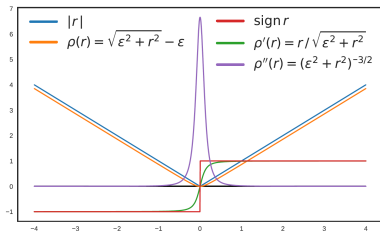
Примеры сглаженных функций несходства

Сглаженный модуль (для аппроксимации медианы):

$$d_\varepsilon(r) = \sqrt{\varepsilon^2 + r^2} - \varepsilon \xrightarrow{\varepsilon \rightarrow 0} |r|$$

Сглаженный несимметричный модуль (для γ -квантили):

$$d_{\gamma\varepsilon}(r) = \begin{cases} 2\gamma d_\varepsilon(r), & r \geq 0 \\ 2(1-\gamma)d_\varepsilon(r), & r < 0 \end{cases} \xrightarrow{\varepsilon \rightarrow 0} |r|_\gamma = \begin{cases} 2\gamma|r|, & r \geq 0 \\ 2(1-\gamma)|r|, & r < 0 \end{cases}$$

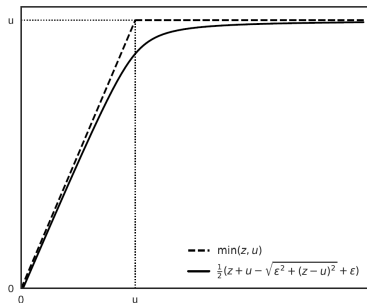


Ещё пример: сглаженное цензурированное среднее

Воспользуемся тождеством $\min(z_i, u) = \frac{1}{2}(z_i + u) - \frac{1}{2}|z_i - u|$

$$M(z_1, \dots, z_\ell) = \frac{1}{2\ell} \sum_{i=1}^{\ell} z_i + z_{\gamma\epsilon} - d_\epsilon(z_i - z_{\gamma\epsilon}) \xrightarrow{\epsilon \rightarrow 0} \frac{1}{\ell} \sum_{i=1}^{\ell} \min(z_i, z^{(m)})$$

$$z_{\gamma\epsilon} = \arg \min_u \sum_{i=1}^{\ell} d_{\gamma\epsilon}(z_i - u) \xrightarrow{\epsilon \rightarrow 0} z^{(m)}, \quad m = \gamma\ell$$



Итерационное взвешивание для агрегирующей функции

Обобщённая минимизация эмпирического риска (ERM):

$$Q(\alpha) = M(\mathcal{L}_1(\alpha), \dots, \mathcal{L}_\ell(\alpha)) + \tau R(\alpha) \rightarrow \min_{\alpha}$$

$$\nabla Q(\alpha) = \sum_{i=1}^{\ell} \underbrace{\frac{\partial M}{\partial \mathcal{L}_i}(\mathcal{L}_1(\alpha), \dots, \mathcal{L}_\ell(\alpha))}_{w_i} \nabla \mathcal{L}_i(\alpha) + \tau \nabla R(\alpha) = 0$$

Алгоритм итерационного взвешивания (IR-ERM):

повторять

$$\alpha := \arg \min_{\alpha} \sum_{i=1}^{\ell} w_i \mathcal{L}_i(\alpha) + \tau R(\alpha);$$

$$w_i := \frac{\partial M}{\partial \mathcal{L}_i}(\mathcal{L}_1(\alpha), \dots, \mathcal{L}_\ell(\alpha)), \quad i = 1, \dots, \ell;$$

пока веса w_i не стабилизируются;

Теперь разберёмся, как вычислять производные $\frac{\partial M}{\partial z_i}(z_1, \dots, z_\ell)$

Вычисление частных производных $\frac{\partial M}{\partial z_k}$

Запишем необходимые условия экстремума по $u \equiv M$

$$M(z_1, \dots, z_\ell) = \arg \min_u \sum_{i=1}^{\ell} d(z_i - u) \quad (*)$$

в виде уравнения $\sum_{i=1}^{\ell} d'(z_i - M) = 0$ относительно M , продифференцируем его по z_k и выразим отсюда $\frac{\partial M}{\partial z_k}$:

$$\sum_{i=1}^{\ell} d''(z_i - M) \frac{\partial}{\partial z_k} (z_i - M) = 0$$

$$d''(z_k - M) = \frac{\partial M}{\partial z_k} \sum_{i=1}^{\ell} d''(z_i - M)$$

$$\frac{\partial M}{\partial z_k} = \frac{d''(z_k - M)}{\sum_{i=1}^{\ell} d''(z_i - M)} = \text{norm}_k d''(z_k - M)$$

Осталось разобраться, как вычислять $u \equiv M$ в задаче (*)

Одномерная задача оптимизации по M

Чтобы решать уравнение $\sum_{i=1}^{\ell} d'(z_i - M) = 0$ относительно M методом простой итерации, представим его в виде $M = f(M)$:

$$\sum_{i=1}^{\ell} \frac{d'(z_i - M)}{z_i - M} (z_i - M) = 0$$

$$\sum_{i=1}^{\ell} z_i \frac{d'(z_i - M)}{z_i - M} = M \sum_{i=1}^{\ell} \frac{d'(z_i - M)}{z_i - M}$$

$$M = \frac{\sum_{i=1}^{\ell} z_i \varphi(z_i - M)}{\sum_{i=1}^{\ell} \varphi(z_i - M)} = \sum_{i=1}^{\ell} z_i \text{norm}_i \varphi(z_i - M), \quad \text{где } \varphi(r) = \frac{d'(r)}{r}$$

Интересно, что M — средневзвешенное значений $\{z_i\}$.

Достаточное условие сходимости метода простой итерации

Процесс $M_{t+1} = f(M_t)$ сходится, если $|f'(M)| < 1$
в окрестности неподвижной точки $M = f(M)$.

$$\left| \frac{\partial}{\partial M} \frac{\sum_i z_i \varphi(z_i - M)}{\sum_i \varphi(z_i - M)} \right| < 1$$

После взятия производной по M :

$$\frac{|\sum_i (z_i - M) \varphi'(z_i - M)|}{|\sum_i \varphi(z_i - M)|} < 1$$

Данное условие нетрудно проверяется для каждой конкретной функции $d(r)$, и для большинства полезных d оно выполнено.

Beliakov G., Sola H., Calvo T. A practical guide to averaging functions. 2016.

З. М. Шибзухов. Минимизации робастных оценок сумм параметризованных функций. 2019.

Собираем всё воедино: алгоритм IR-ERM

Вход: $\mathcal{L}_i(\alpha)$ — функции потерь на обучающей выборке;

Выход: параметры модели α , веса объектов w_i ;

инициализация $w_i := \frac{1}{\ell}$, $i = 1, \dots, \ell$;

повторять

$$\alpha := \arg \min_{\alpha} \sum_{i=1}^{\ell} w_i \mathcal{L}_i(\alpha) + \tau R(\alpha);$$

$$z_i := \mathcal{L}_i(\alpha); \text{ инициализация } M := \sum_{i=1}^{\ell} w_i z_i;$$

повторять

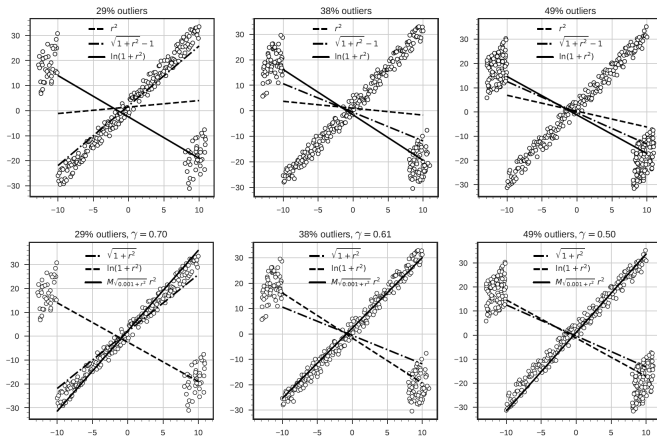
$$M = \sum_{i=1}^{\ell} z_i \operatorname{norm}_i \varphi(z_i - M), \text{ где } \varphi(r) = \frac{d'(r)}{r};$$

пока значение M не сойдётся;

$$w_i := \operatorname{norm}_i d''(z_i - M), \quad i = 1, \dots, \ell;$$

пока веса w_i не стабилизируются;

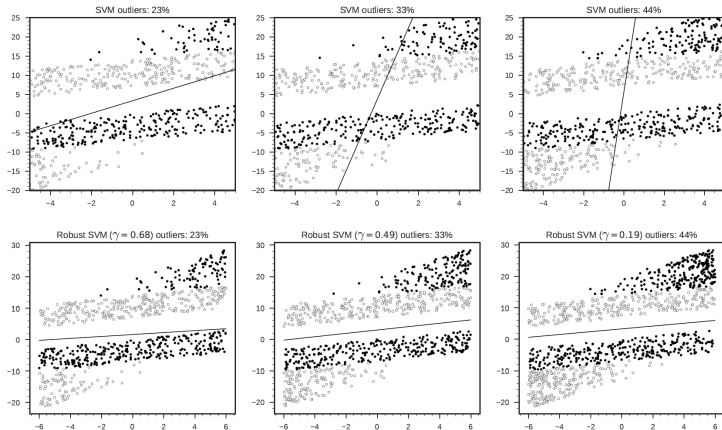
Пример 1. Робастная регрессия (линейная)



Агрегирующая функция справляется даже с 49% выбросов

З.М.Шибзухов. Методы машинного обучения на основе минимизации
сглаженных оценок средних, нечувствительных к выбросам. ММРО-2019.

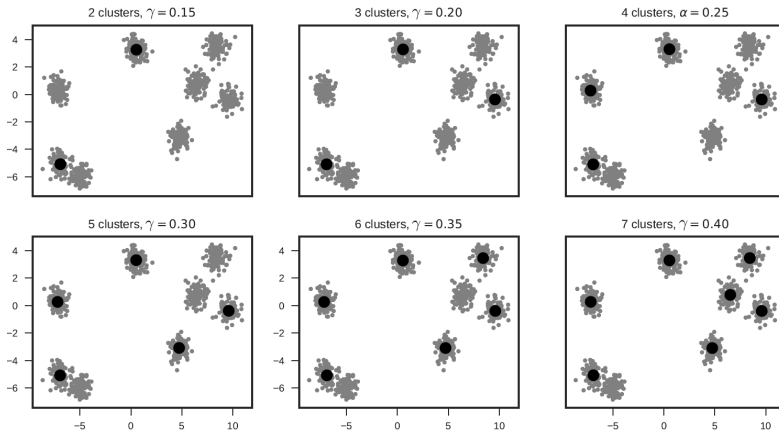
Пример 2. Робастная классификация (SVM)



Агрегирующая функция справляется даже с 44% выбросов

З.М.Шибзухов. Методы машинного обучения на основе минимизации
сглаженных оценок средних, нечувствительных к выбросам. ММРО-2019.

Пример 3. Робастная кластеризация



Если в данных смешано несколько зависимостей, то вместо компромиссного «натягивания» одной модели на все данные *робастные методы* моделируют основную, игнорируя остальные

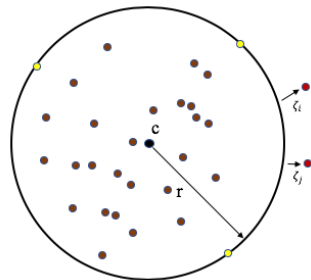
Одноклассовый SVM (one-class SVM, OSVM)

Дано: обучающая выборка $\{x_i \in \mathbb{R}^n : i = 1, \dots, \ell\}$

Найти: центр $c \in \mathbb{R}^n$ и радиус r шара, охватывающего всю выборку кроме аномальных объектов-выбросов

Критерий: минимизация радиуса шара и суммы штрафов за выход из шара:

$$\nu r^2 + \sum_{i=1}^{\ell} \mathcal{L}(\underbrace{r^2 - \|x_i - c\|^2}_{\zeta_i = \text{margin}(c, r)}) \rightarrow \min_{c, r}$$



При $\mathcal{L}(\zeta) = (-\zeta)_+$ свойства решения аналогичны SVM:

- Выпуклая задача квадратичного программирования
- Решение *разрежено* — зависит только от *опорных объектов*
- Обобщение на нелинейные модели: $\langle x_i, x_j \rangle \rightarrow K(x_i, x_j)$

Частный случай SSL: PU-learning (Positive and Unlabeled)

Примеры задач, когда известны объекты только одного класса:

- обнаружение мошеннических транзакций
- рекомендательные системы, персонализация рекламы
- автоматическое пополнение базы знаний фактами

Модель двухклассовой классификации $a(x_i, w)$.

Неразмеченные трактуются как негативные с весом $C_- \ll C_+$:

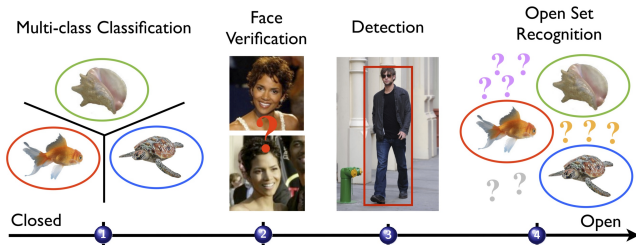
$$\sum_{i=1}^k \frac{C_+}{k} \mathcal{L}(a(x_i, w), +1) + \sum_{i=k+1}^{\ell} \frac{C_-}{\ell-k} \mathcal{L}(a(x_i, w), -1) + \tau R(w) \rightarrow \min_w$$

Один из успешных методов — Biased SVM.

Gang Li. A Survey on Positive and Unlabelled Learning. 2013.

J.Bekker, J.Davis. Learning From Positive and Unlabeled Data: A Survey. 2020.

Задачи классификации с нефиксированным набором классов



- 1 Обычная многоклассовая классификация
- 2 Дообучение модели на каждом новом классе
- 3 Детектирование объектов одного класса против остальных (One-Class Classification)
- 4 Распознавание с открытым набором классов (Open-Set Recognition → Open-World Recognition)

- Природа аномальности объектов:
 - помехи (ошибки, шум, грязь) в исходных данных,
 - модель плохо описывает примеси посторонних явлений,
 - регулярно появляется что-то принципиально новое.
- Простой способ отсева наиболее грубых выбросов — исключать объекты с наибольшими значениями потерь.
- Редкий для ML случай: минимизируется не сумма потерь $\mathcal{L}_1 + \dots + \mathcal{L}_l$, а обобщённое среднее $M(\mathcal{L}_1, \dots, \mathcal{L}_l)$.
- Природа аномальности классов:
 - невозможность собрать обучающие объекты класса,
 - динамическое увеличение числа классов
- Не существует идеального способа определения аномалий. Явно или неявно предполагается «модель аномалии».

З.М.Шибзухов. Минимизации робастных оценок сумм параметризованных функций. 2019.

M.Salehi et al. A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: solutions and future challenges. 2021.