Проблемно-ориентированные методы алгебраического подхода (конспект лекций)

§1 Задачи обучения по прецедентам

Опр 1 Задача *обучения по прецедентам* $Z = \langle \mathfrak{I}_i, \mathfrak{I}_f, \mathfrak{M}^u, \{x_k, y_k\}_{k=1}^q \rangle$, где

$\mathfrak{I}_i$ — множество начальных информаций,

$\mathfrak{I}_f$ — множество финальных информаций,

$\{x_k, y_k\}_{k=1}^q \subset \mathfrak{I}_i \times \mathfrak{I}_f$ — обучающая выборка,

$\mathfrak{M}^u$ — множество отображений из $\mathfrak{I}_i$ в $\mathfrak{I}_f$.

Требуется построить *корректный* алгоритм $A : \mathfrak{I}_i \to \mathfrak{I}_f$ такой, что:

1^0. $A(x_k) = y_k$, для всех $k = 1, \dots, q$ — *локальные* ограничения.

2^0. $A \in \mathfrak{M}^u$ — *универсальные* ограничения.

Опр 2 Задача *восстановления регрессии*: $\mathfrak{I}_f = \mathbb{R}$.

Опр 3 Задача *классификации на l непересекающихся классов*: $\mathfrak{I}_f = \{0, \dots, l-1\}$.

§2 Оптимизационный подход

Шаг 1 Выбирается $\mathfrak{M} \subseteq \mathfrak{M}^u$ — *эвристическая информационная модель* алгоритмов.
Шаг 2 Выбирается $Q : \mathfrak{M}^u \times (\mathfrak{I}_i \times \mathfrak{I}_f)^q \to \mathbb{R}_+$ — *функционал качества.*
Шаг 3 Решается задача минимизации Q при заданной выборке $\{x_k, y_k\}_{k=1}^q$:

$$A^* = \arg\min_{A \in \mathfrak{M}} Q(A, \{x_k, y_k\}_{k=1}^q).$$

Пример 1 (функционалы качества)
Введём w_k — веса объектов обучения, $k = 1, \dots, q$.
Функционал среднеквадратичной ошибки — для задач восстановления регрессии:

$$Q = \sum_{k=1}^q w_k \left(A(x_k) - y_k\right)^2.$$

Функционал количества ошибок — для задач классификации:

$$Q = \sum_{k=1}^q w_k [A(x_k) \neq y_k].$$

§3 Алгебраический подход

Шаг 1 Выбирается $\mathfrak{I}_e$ — *пространство оценок.*
Шаг 2 Выбирается $\mathfrak{M}^0 \subseteq \{B : \mathfrak{I}_i \to \mathfrak{I}_e\}$ — *модель алгоритмических операторов.*
Шаг 3 Выбирается $\mathfrak{M}^1 \subseteq \bigcup_{p=0}^{\infty}\{C : \mathfrak{I}_e^p \to \mathfrak{I}_f\}$ — *семейство решающих правил.*
Шаг 4 Выбирается $\mathfrak{F} \subseteq \bigcup_{p=0}^{\infty}\{F : \mathfrak{I}_e^p \to \mathfrak{I}_e\}$ — *семейство корректирующих операций.*
Шаг 5 Строится корректный алгоритм вида $A = C \circ F(B_1, \dots, B_p)$:

$$\begin{tikzcd}
\mathfrak{I}_i \arrow[r, "A=C\circ F(B_1,\dots,B_p)"] \arrow[d, "B_1,\dots,B_p"] & \mathfrak{I}_f \\
\mathfrak{I}_e^p \arrow[r, "F"] & \mathfrak{I}_e \arrow[u, "C"']
\end{tikzcd}$$

Все суперпозиции вида $C \circ B$ образуют модель алгоритмов $\mathfrak{M} = \mathfrak{M}^1 \circ \mathfrak{M}^0$.
Все суперпозиции вида $C \circ F(B_1, \dots, B_p)$ образуют её *$\mathfrak{F}$-расширение* $\mathfrak{F}(\mathfrak{M}) = \mathfrak{M}^1 \circ \mathfrak{F}(\mathfrak{M}^0)$.
Предполагается, что семейства $\mathfrak{M}^0$, $\mathfrak{F}$, $\mathfrak{M}^1$ выбраны так, что $\mathfrak{F}(\mathfrak{M}) \subseteq \mathfrak{M}^u$.

Опр 4 *Глобальный базис* для задачи Z — набор алгоритмических операторов $B_1, \dots, B_p$ из $\mathfrak{M}^0$ такой, что для любой выборки $\{v_k\}_{k=1}^q \in (\mathfrak{I}_f)^q$

$$\exists F \in \mathfrak{F},\ \exists C \in \mathfrak{M}^1 : \quad A(x_k) = v_k,\ k = 1, \dots, q.$$

Опр 5 *Локальный базис* для задачи Z — набор алгоритмических операторов $B_1, \dots, B_p$ из $\mathfrak{M}^0$ такой, что

$$\exists F \in \mathfrak{F},\ \exists C \in \mathfrak{M}^1 : \quad A(x_k) = y_k,\ k = 1, \dots, q.$$

Глобальный базис для задачи Z существует тогда и только тогда, когда Z регулярна.
Локальный базис для задачи Z существует тогда и только тогда, когда Z разрешима.
Глобальный базис является также и локальным. Обратное в общем случае неверно.

Классические и проблемно-ориентированные методы алгебраического подхода:

	Классические	**Проблемно-ориентированные**
Тип базиса	глобальный	локальный
Требование к задаче	регулярность	разрешимость
Цель построения	док-во теорем существования	решение прикладных задач
Способ построения	разложение по базису	численная оптимизация
Сложность алгоритма	высокая	низкая
Качество экстраполяции	низкое	высокое

§4 Функционалы качества алгоритмических операторов

Функционал качества алгоритмических операторов:

$$Q : \mathfrak{M}^0_* \times (\mathfrak{I}_i \times \mathfrak{I}_f)^q \to \mathbb{R}_+.$$

Канонический способ определения $Q(B)$ по $Q(A)$:

$$Q(B) = \min_{C \in \mathfrak{F}^1} Q(C \circ B).$$

Пример 2 (задача восстановления регрессии)
Положим $\mathfrak{I}_f = \mathfrak{I}_e = \mathbb{R}$.
Решающие правила не используются: $\mathfrak{M}^1 = \{C(B) \equiv B\}$.
Функционал качества:

$$Q(B) = \sum_{k=1}^{q} w_k \left(B(x_k) - y_k\right)^2.$$

Для его минимизации можно использовать метод наименьших квадратов.

Пример 3 (задача классификации)
Положим $\mathfrak{I}_f = \{0, 1\}$, $\mathfrak{I}_e = \mathbb{R}$.
Семейство пороговых решающих правил: $\mathfrak{M}^1 = \{[B > \gamma] : \gamma \in \mathbb{R}\}$.
Функционал качества:

$$Q(B) = \min_{\gamma \in \mathbb{R}} \sum_{k=1}^{q} w_k \big|[B(x_k) > \gamma] - y_k\big|.$$

Его минимизация сводится к поиску и решению совместной подсистемы максимального веса в системе q неравенств с весами w_k, $k = 1, \ldots, q$, относительно оператора B и порога γ:

$$\begin{cases} B(x_k) \leqslant \gamma, & \text{если } y_k = 0; \\ B(x_k) > \gamma, & \text{если } y_k = 1; \end{cases}$$

при этом считается, что вес подсистемы равен сумме весов всех входящих в неё неравенств. Это классическая NP-полная комбинаторная задача. Для её решения можно использовать метод ветвей и границ.

§5 Оптимизационные методы построения локальных базисов

Оптимизационная задача построения оператора $F(B_1, \dots, B_p)$:

$$\bigl(F^*, B_1^*, \dots, B_p^*\bigr) = \mathop{\arg\min}_{\substack{B_1,\dots,B_p \in \mathfrak{M}^0 \\ F \in \mathfrak{F}}} Q(F(B_1, \dots, B_p)).$$

Совместная минимизация по всем $B_1, \dots, B_p$ и F является трудной проблемой.
Рассмотрим итерационный *процесс последовательного наращивания базиса*:
Шаг 1 Первый оператор строится также, как в оптимизационном подходе:

$$B_1^* = \mathop{\arg\min}_{B_1 \in \mathfrak{M}^0} Q(B_1).$$

Шаг 2 Если качество оператора B_1^* не удовлетворяет, строится второй оператор, который объединяется с первым с помощью корректирующей операции:

$$(B_2^*, F^*) = \mathop{\arg\min}_{\substack{B_2 \in \mathfrak{M}^0 \\ F \in \mathfrak{F}}} Q(F(B_1^*, B_2)).$$

Шаг 3 Процесс повторяется до получения оператора $F(B_1, \dots, B_p)$ удовлетворительного качества, либо до достижения заданного p:

$$(B_p^*, F^*) = \mathop{\arg\min}_{\substack{B_p \in \mathfrak{M}^0 \\ F \in \mathfrak{F}}} Q(F(B_1^*, \dots, B_{p-1}^*, B_p)). \tag{5.1}$$

Шаг 4 По окончании наращивания возможна повторная оптимизация операторов:

$$(B_r^*, F^*) = \mathop{\arg\min}_{\substack{B_r \in \mathfrak{M}^0 \\ F \in \mathfrak{F}}} Q(F(B_1^*, \dots, B_{r-1}^*, B_r, B_{r+1}^*, \dots, B_p^*)).$$

В дальнейшем будет исследоваться возможность сведения совместной оптимизации (B_p, F) на шаге 3 к последовательному построению B_p и F с применением готовых оптимизационных методов, разработанных ранее для решения стандартной задачи

$$B_p^* = \mathop{\arg\min}_{B_p \in \mathfrak{M}^0} Q(B_p). \tag{5.2}$$

Будут получены формулы пересчёта весов объектов w_k и ответов y_k, с помощью которых любой стандартный метод решения задачи (5.2) преобразуется в метод решения задачи (5.1). Эти формулы существенно зависят от вида семейства $\mathfrak{F}$. Будут рассмотрены семейства линейных, полиномиальных и монотонных корректирующих операций.
В некоторых случаях задача совместной оптимизации (B_p, F) не допускает непосредственного сведения к (5.2). Тогда её решение разбивается на два шага, из которых может быть образован отдельный итерационный процесс:

$$B_p^* = \mathop{\arg\min}_{B_p \in \mathfrak{M}^0} Q(F^*(B_1^*, \dots, B_{p-1}^*, B_p)); \tag{5.3}$$

$$F^* = \mathop{\arg\min}_{F \in \mathfrak{F}} Q(F(B_1^*, \dots, B_p^*)).$$

§6 Иллюстрации к методу наращивания локального базиса

Задача на плоскости: $\mathfrak{I}_i = \mathbb{R}^2$, $\mathfrak{I}_f = \{\circ, \times\}$, $\mathfrak{M}^0$ — прямые на плоскости.
Оператор $F(B_1, \ldots, B_p)$ строит кусочно-линейную разделяющую поверхность.
Каждый из операторов $B_1, \ldots, B_{p-1}$ размечает только одну полуплоскость.
Оператор B_p размечает обе полуплоскости.

Пример 4 Разделимость одной плоскостью.
Вывод: итерации могут закончиться после первого шага.

Пример 5 Разделимость одной плоскостью не достигается, двумя — достигается.
Вывод: методом последовательного наращивания можно получить локальный базис.

Пример 6 После первого шага 1 ошибка, после второго 3 ошибки.
Вывод 1: приходится применять различные ухищрения, чтобы избежать повтора $B_2^* = B_1^*$.
Вывод 2: на промежуточных итерациях качество может ухудшаться.

Пример 7 После третьего шага 0 ошибок. Альтернатива: после построения B_2 вернуться к оптимизации B_1. Тоже 0 ошибок, при этом B_3 уже не нужен.
Вывод: при повторной оптимизации некоторые операторы могут оказаться лишними.

Пример 8 Оптимизация B_1 ничего не даёт. Возьмём B_1 наобум и оптимизируем B_2.
Вывод: низкое качество оператора может быть скомпенсировано на следующих шагах.

§7 Линейные корректирующие операции

Семейство линейных корректирующих операций при $\mathfrak{I}_e = \mathbb{R}$:

$$\mathfrak{F}_L = \left\{ F(B_1, \dots, B_p) = \sum_{i=1}^{p} \alpha_i B_i \;\middle|\; \alpha_i \in \mathbb{R},\ i = 1, \dots, p \right\}.$$

Для задачи восстановления регрессии положим

$$\mathfrak{I}_f = \mathfrak{I}_e = \mathbb{R}, \quad \mathfrak{M}^1 = \{C(B) \equiv B\},$$
$$Q(B) = \sum_{k=1}^{q} (B(x_k) - y_k)^2.$$

Введём матричные обозначения:

$$Z = [B_i^*(x_k)]_{k=1,q}^{i=1,p-1}, \qquad a = [\alpha_i]_{i=1,p-1},$$
$$z(B_p) = [B_p(x_k)]_{k=1,q}, \quad y = [y_k]_{k=1,q}.$$

Теорема 1 (о сводимости)

В случае восстановления регрессии построение оператора B_p в задачах (5.2) и (5.3) при $\alpha_p \neq 0$ сводится к минимизации одного и того же квадратичного функционала

$$Q(B_p) = \|z(B_p) - v\|^2,$$

причём

$$v = \begin{cases} y, & \text{для задачи (5.2);} \\ \frac{1}{\alpha_p}(y - Za), & \text{для задачи (5.3).} \end{cases}$$

Доказательство.

Запишем функционал для задачи (5.2) в матричных обозначениях:

$$Q(B_p) = \|z(B_p) - y\|^2.$$

Теперь запишем функционал для задачи (5.3):

$$Q(F^*(B_1^*, \dots, B_{p-1}^*, B_p)) = \|Za + z(B_p)\alpha_p - y\|^2 = \alpha_p^2 \|z(B_p) - \tfrac{1}{\alpha_p}(y - Za)\|^2.$$

Отбрасывая положительный множитель α_p^2, заключаем, что минимизация этого функционала эквивалентна минимизации $\|z(B_p) - v\|^2$. ■

Упражнение 1 В отличие от (5.3), задача (5.1) не сводится к той же постановке, что (5.2). Показать, что решение (5.1) эквивалентно минимизации функционала

$$Q(B_p) = \min_{F \in \mathfrak{F}} Q(F(B_1^*, \dots, B_{p-1}^*, B_p)) = \|P_Z(y - \alpha_p z(B_p))\|^2, \quad \text{где}$$
$$\alpha_p = \frac{z^T(B_p) P_Z y}{z^T(B_p) P_Z z(B_p)},$$
$$P_Z = Z(Z^T Z)^{-1} Z^T.$$

Для задачи классификации в 2 непересекающихся класса положим:

$$\mathfrak{I}_f = \{0, 1\}, \quad \mathfrak{I}_e = \mathbb{R}, \quad \mathfrak{M}^1 = \{[B > \gamma] \mid \gamma \in \mathbb{R}\},$$
$$Q(B) = \min_{\gamma \in \mathbb{R}} \sum_{k=1}^{q} \bigl|[B(x_k) > \gamma] - y_k\bigr|.$$

Теорема 2 (о сводимости)
В случае классификации построение оператора B_p в задачах (5.2) и (5.3) при $\alpha_p \neq 0$ сводится к поиску и решению максимальной совместной подсистемы в системе q неравенств

$$z(B_p) \lesseqgtr v$$

относительно оператора $B_p \in \mathfrak{M}^0$ и порога $\gamma \in \mathbb{R}$, причём

$$v = \begin{cases} \gamma, & \text{для задачи (5.2);} \\ \frac{1}{\alpha_p}(\gamma - Za), & \text{для задачи (5.3).} \end{cases}$$
$$(\lesseqgtr)_k = \begin{cases} \leqslant, & \text{если } y_k = 0\text{, для задачи (5.2);} \\ >, & \text{если } y_k = 1\text{, для задачи (5.2);} \\ \leqslant, & \text{если } y_k = 0 \text{ и } \alpha_p > 0\text{, для задачи (5.3);} \\ >, & \text{если } y_k = 1 \text{ и } \alpha_p > 0\text{, для задачи (5.3);} \\ \geqslant, & \text{если } y_k = 0 \text{ и } \alpha_p < 0\text{, для задачи (5.3);} \\ <, & \text{если } y_k = 1 \text{ и } \alpha_p < 0\text{, для задачи (5.3);} \end{cases}$$

Доказательство.
Запишем функционал для задачи (5.2) в виде системы неравенств:

$$\begin{cases} B_p(x_k) \leqslant \gamma, & \text{если } y_k = 0; \\ B_p(x_k) > \gamma, & \text{если } y_k = 1. \end{cases}$$

В матричных обозначениях это то же самое, что $z(B_p) \lesseqgtr \gamma$, где $(\lesseqgtr)_k = \begin{cases} \leqslant, & \text{если } y_k = 0; \\ >, & \text{если } y_k = 1. \end{cases}$

Теперь запишем функционал для задачи (5.3):

$$Za + z(B_p)\alpha_p \lesseqgtr \gamma, \quad (\lesseqgtr)_k = \begin{cases} \leqslant, & \text{если } y_k = 0; \\ >, & \text{если } y_k = 1. \end{cases}$$

Выражая отсюда $z(B_p)$, получаем требуемое. ∎

§8 Полиномиальные корректирующие операции

Семейство полиномиальных корректирующих операций, $\mathfrak{I}_e = \mathbb{R}$:

$$\mathfrak{F}_P = \left\{ F(B_1, \ldots, B_p) = \sum_{i=1}^{s} \alpha_i B_{p_{i-1}+1} \ldots B_{p_i} \;\middle|\; \begin{array}{l} \alpha_i \in \mathbb{R},\ i = 1, \ldots, s \\ 0 = p_0 < p_1 < \cdots < p_s = p \end{array} \right\}.$$

Для задачи восстановления регрессии положим

$$\mathfrak{I}_f = \mathfrak{I}_e = \mathbb{R}, \quad \mathfrak{M}^1 = \{C(B) \equiv B\},$$
$$Q(B) = \sum_{k=1}^{q} (B(x_k) - y_k)^2.$$

Обозначим через $\otimes$ и $\oslash$ операции покомпонентного умножения и деления векторов.

Теорема 3 (о сводимости)
В случае восстановления регрессии построение оператора B_p в задачах (5.2) и (5.3) сводится к минимизации одного и того же квадратичного функционала

$$Q(B_p) = \sum_{k=1}^{q} w_k^2 (B_p(x_k) - v_k)^2,$$

причём

$$v = \begin{cases} y, & \text{для задачи (5.2);} \\ (y - Za) \oslash w, & \text{для задачи (5.3);} \end{cases}$$
$$w_k = \begin{cases} 1, & \text{для задачи (5.2);} \\ [B^*_{p_{s-1}+1}(x_k) \ldots B^*_{p-1}(x_k)]_{k=1,q}; & \text{для задачи (5.3);} \end{cases}$$
$$Z = [B^*_{p_{i-1}+1}(x_k) \ldots B^*_{p_i}(x_k)]_{k=1,q}^{i=1,s-1}.$$

Доказательство.
Запишем функционал для задачи (5.2) в матричных обозначениях:

$$Q(B_p) = \|z(B_p) - y\|^2 = \|w \otimes (z(B_p) - v)\|^2,$$

поскольку w — в данном случае вектор, состоящий из единиц.
Теперь запишем функционал для задачи (5.3):

$$\begin{aligned} Q(F^*(B_1^*, \ldots, B_{p-1}^*, B_p)) &= \|Za + w \otimes z(B_p) - y\|^2 \\ &= \|w \otimes (z(B_p) - (y - Za) \oslash w)\|^2 \\ &= \|w \otimes (z(B_p) - v)\|^2. \end{aligned}$$

Таким образом, обе задачи сводятся к минимизации одного и того же функционала. ∎

§9 Дробно-квадратичные корректирующие операции

Семейство квадратичных корректирующих операций, $\mathfrak{I}_e = \mathbb{R}$:

$$\mathfrak{F}_Q = \left\{ F(B_1, W_1, \ldots, B_p, W_p) = \sum_{i=1}^{p} \alpha_i B_i W_i \;\middle|\; \alpha_i \in \mathbb{R},\ i = 1, \ldots, p \right\}.$$

Семейство дробно-квадратичных корректирующих операций, $\mathfrak{I}_e = \mathbb{R}$:

$$\mathfrak{F}_{DQ} = \left\{ F(B_1, W_1, \ldots, B_p, W_p) = \frac{\sum_{i=1}^{p} \alpha_i B_i W_i}{\sum_{i=1}^{p} W_i} \;\middle|\; \alpha_i \in \mathbb{R},\ i = 1, \ldots, p \right\}, \quad \text{где}$$

$B_i \in \mathfrak{M}_i^0$ — модели алгоритмических операторов,
$W_i \in \tilde{\mathfrak{M}}_i^0$ — модели *областей компетентности*,
$W_i(x) \in [0, 1]$ — функция, характеризующая компетентность оператора B_i — чем больше $W_i(x)$, тем выше степень доверия к результату $B_i(x)$.
Итерационный процесс наращивания локального базиса позволяет настроить не только сами операторы B_i, но и соответствующие им области компетентности W_i.

Пример 9 (некоторые простые разновидности областей компетентности)
Пусть $f(x)$ — некоторый признак, $\rho(x, x')$ — заданная метрика на множестве объектов $\mathfrak{I}_i$.
1. Функция с параметрами K и M, определяющая компетентность оператора B_i по правилу «чем ближе значение признака f к числу M, тем выше компетентность B_i»:

$$W_i(x) = \frac{1}{1 + K(f(x) - M)^2}.$$

2. Функция с параметрами K и M, определяющая компетентность оператора B_i по правилу «чем больше значение признака f, тем выше компетентность B_i»:

$$W_i(x) = \frac{1}{2} + \frac{1}{2}\,\mathrm{th}(K(f(x) - M)).$$

3. Функция с параметрами K и x_0, определяющая компетентность оператора B_i по правилу «чем ближе к точке x_0, тем выше компетентность B_i»:

$$W_i(x) = \frac{1}{1 + K\rho(x, x_0)}.$$

§10 Монотонные корректирующие операции

Семейство монотонных корректирующих операций:

$$\mathfrak{F}_M = \bigcup_{p=0}^{\infty} \left\{ F : \mathfrak{I}_e^p \to \mathfrak{I}_f \;\middle|\; (\forall u, v \in \mathfrak{I}_e^p)\;\; u \leqslant v \to F(u) \leqslant F(v) \right\},$$

где $\mathfrak{I}_f$ и $\mathfrak{I}_e$ — произвольные частично упорядоченные множества, решающие правила не используются.

Элементы u и v несравнимы: $u \parallel v$, если $\neg(u \leqslant v) \wedge \neg(v \leqslant u)$.

Последовательность пар $\{u_k, v_k\}_{k=1}^n$ монотонна, если $u_j \leqslant u_k \to v_j \leqslant v_k$ для всех j, k.

Лемма 4 (о проведении монотонной функции через заданные точки)
Монотонная функция $f : U \to V$ такая, что $f(u_k) = v_k$ для всех $k = 1, \ldots, n$ существует тогда и только тогда, когда последовательность $\{u_k, v_k\}_{k=1}^n$ монотонна.

Доказательство.

Необходимость. Если f монотонна, то последовательность $\{u_k, f(u_k)\}_{k=1}^n$ монотонна.

Достаточность. Пусть последовательность $\{u_k, v_k\}_{k=1}^n$ монотонна. Построим функцию f в классе кусочно постоянных функций. Положим для произвольного $u \in U$:

$$I(u) = \{k \mid u_k \leqslant u\};$$

$$f(u) = \begin{cases} \min\limits_{k=1,\ldots,n} v_k, & \text{если } I(u) = \varnothing; \\ \max\limits_{k \in I(u)} v_k, & \text{если } I(u) \neq \varnothing. \end{cases}$$

Функция f монотонна, поскольку для произвольных u и u' из U

$$u \leqslant u' \quad \to \quad I(u) \subseteq I(u') \quad \to \quad f(u) \leqslant f(u').$$

Функция f принимает заданные значения в точках u_k, так как для всех $k = 1, \ldots, n$

$$k \in I(u_k) \quad \to \quad I(u_k) \neq \varnothing \quad \to \quad f(u_k) = \max_{j \in I(u_k)} v_j = v_k,$$

причём максимум достигается при $j = k$ в силу монотонности $\{u_k, v_k\}_{k=1}^n$:

$$j \in I(u_k) \quad \to \quad u_j \leqslant u_k \quad \to \quad v_j \leqslant v_k. \quad \blacksquare$$

§11 Дефектные пары и теорема сходимости

Обозначения:

$$S = \{1, \dots, q\};$$
$$a_k = [B_i(x_k)]_{i=1}^p, \quad k \in S;$$
$$f_k = F(B_1, \dots, B_p)(x_k) = F(a_k), \quad k \in S.$$

Условие корректности алгоритма $F(B_1, \dots, B_p)$ в этих обозначениях:

$$F(a_k) = y_k,\ k \in S.$$

Опр 6 Пара индексов $(j,k) \in S^2$ *дефектная* для B, если $y_j < y_k$ и $B(x_j) \geqslant B(x_k)$.

Опр 7 *Дефект оператора* B — множество всех его дефектных пар:

$$\mathbb{D}(B) = \left\{(j,k) \in S^2 \;\middle|\; y_j < y_k \wedge B(x_j) \geqslant B(x_k)\right\}.$$

Опр 8 *Совокупный дефект операторов* $B_1, \dots, B_p$:

$$\mathbb{D}_p(B_1, \dots, B_p) = \mathbb{D}(B_1) \cap \dots \cap \mathbb{D}(B_p) = \left\{(j,k) \in S^2 \;\middle|\; y_j < y_k \wedge a_j \geqslant a_k\right\}.$$

Сокращённое обозначение: $\mathbb{D}_p \equiv \mathbb{D}_p(B_1, \dots, B_p)$.

Теорема 5 Следующие три условия эквивалентны:

$$\mathbb{D}_p(B_1, \dots, B_p) = \varnothing;$$
$$(\exists F \in \mathfrak{F}_M) \quad \mathbb{D}(F(B_1, \dots, B_p)) = \varnothing;$$
$$(\exists F \in \mathfrak{F}_M) \quad F(a_k) = y_k,\ k \in S;$$

Доказательство.

$$
\begin{aligned}
&\mathbb{D}_p(B_1, \dots, B_p) = \varnothing \ \leftrightarrow \quad \cdots \text{по определению совокупного дефекта}\\
&\leftrightarrow \forall (j,k) \in S^2 \quad \neg(a_k \leqslant a_j \wedge y_j < y_k) \ \leftrightarrow \quad \cdots \text{из логики: } \neg(A \wedge B) = (A \to \neg B)\\
&\leftrightarrow \forall (j,k) \in S^2 \quad (a_k \leqslant a_j) \to (y_k \leqslant y_j) \ \leftrightarrow\\
&\leftrightarrow \{a_k, y_k\}_{k \in S} \text{ — монотонная последовательность } \leftrightarrow \quad \cdots \text{по Лемме 4}\\
&\leftrightarrow (\exists F \in \mathfrak{F}_M) \quad (\forall k \in S) \quad F(a_k) = y_k \ \leftrightarrow\\
&\leftrightarrow (\exists F \in \mathfrak{F}_M) \quad \forall (j,k) \in S^2 \quad (a_k \leqslant a_j) \to (F(a_k) \leqslant F(a_j)) \ \leftrightarrow\\
&\leftrightarrow (\exists F \in \mathfrak{F}_M) \quad \forall (j,k) \in S^2 \quad \neg(a_k \leqslant a_j \wedge F(a_k) > F(a_j)) \ \leftrightarrow \quad \cdots \text{по опр. дефекта}\\
&\leftrightarrow (\exists F \in \mathfrak{F}_M) \quad \mathbb{D}(F(B_1, \dots, B_p)) = \varnothing. \quad \blacksquare
\end{aligned}
$$

Следствие 1 Чтобы установить существование корректного алгоритма $F(B_1, \dots, B_p)$, достаточно проверить, что совокупный дефект пуст. Для этого не требуется строить F.
Следствие 2 Можно ввести функционал качества

$$Q(B) = |\mathbb{D}(B)|.$$

Следствие 3 При добавлении оператора B_p совокупный дефект не увеличивается:

$$\mathbb{D}(B_1, \dots, B_p) \subseteq \mathbb{D}(B_1, \dots, B_{p-1}).$$

Устранение дефектной пары (j, k) происходит при условии

$$B_p(x_j) < B_p(x_k), \quad (j, k) \in \mathbb{D}(B_1, \dots, B_{p-1}),$$

при этом пара векторов $a_k \leqslant a_j$ переходит в разряд несравнимых.

Теорема 6 (о сходимости) Пусть задан оператор B_1, и модель $\mathfrak{M}^0$ такова, что для любого m-элементного подмножества $\Delta \subseteq \{(j, k) \in S^2 \mid y_j < y_k\}$, $m \geqslant 1$, можно построить оператор B, для которого $B(x_j) < B(x_k)$ при всех $(j, k) \in \Delta$. Тогда итерационный процесс (5.1) строит локальный базис $B_1, B_2, \dots, B_p$, в котором $p \leqslant \left\lceil \frac{Q(B_1)}{m} \right\rceil + 1$.

Доказательство.
Будем последовательно строить операторы B_r из $\mathfrak{M}^0$, $r = 2, 3, \dots$, для которых

$$B_r(x_j) < B_r(x_k), \quad (j, k) \in \Delta_r.$$

Согласно условию теоремы можно выбрать оператор B_r так, что $\Delta_r \subseteq \mathbb{D}_{r-1}(B_1, \dots, B_{r-1})$ и $|\Delta_r| \geqslant m$, $r = 2, 3, \dots$ Тогда

$$|\mathbb{D}_r(B_1, \dots, B_r)| \leqslant |\mathbb{D}_{r-1}(B_1, \dots, B_{r-1})| - m.$$

Следовательно, дефект будет исчерпан при $r = \left\lceil \frac{Q(B_1)}{m} \right\rceil + 1.$ ■

§12 Дефектные тройки и двусторонняя оценка $Q(F(B_1, \dots, B_p))$

Опр 9 Тройка индексов $(j, s, k) \in S^3$ *дефектная* для $B_1, \dots, B_p$, если

$$a_k \leqslant a_j, \quad a_s \parallel a_k, \quad a_s \parallel a_j;$$
$$y_j < y_k, \quad y_j \leqslant y_s \leqslant y_k.$$

Дефектная тройка имеет основание (j, k), вершину s, ребра (s, k) и (s, j). Если $y_j < y_s < y_k$, то тройка (j, s, k) называется строго дефектной.

Основание дефектной тройки принадлежит совокупному дефекту: $(j, k) \in \mathbb{D}_p$.

Лемма 7 Совокупный дефект неустраним монотонными корректирующими операциями:

$$(\forall F \in \mathfrak{F}_M) \quad \mathbb{D}_p \subseteq \mathbb{D}(F(B_1, \dots, B_p)).$$

Лемма 8 Если в S^3 есть строго дефектная тройка, то

$$(\forall F \in \mathfrak{F}_M) \quad \mathbb{D}_p \subset \mathbb{D}(F(B_1, \dots, B_p)).$$

Лемма 9 Если в S^3 нет дефектных троек, то

$$(\exists F \in \mathfrak{F}_M) \quad \mathbb{D}_p = \mathbb{D}(F(B_1, \dots, B_p)).$$

Следствие Для минимизации $Q(F(B_1, \dots, B_p))$ следует строить B_p так, чтобы устранить как можно больше дефектных пар из $\mathbb{D}(B_1, \dots, B_{p-1})$. В первую очередь должны устраняться пары, лежащие в основание наибольшего количества дефектных троек.

Гипотеза Возможно построить монотонную корректирующую операцию F так, чтобы разность $\mathbb{D}(F(B_1, \dots, B_p)) \backslash \mathbb{D}_p$ состояла только из рёбер дефектных троек.

Следствие Пусть T_{jk} — суммарное число рёбер дефектных троек с основанием (j, k). Тогда существует корректирующая операция $F \in \mathfrak{F}_M$ такая, что

$$|\mathbb{D}_p| \leqslant Q(F(B_1, \dots, B_p)) \leqslant |\mathbb{D}_p| + \sum_{(j,k) \in \mathbb{D}_p} T_{jk}. \tag{12.1}$$

§13 Совместная оптимизация B_p и F

Опр 10 Будем говорить, что минимизация функционала $g(z)$ *сводится* к минимизации функционала $g'(z)$ по множеству Z, если выполняются два условия:

$$\min_{z\in Z} g(z) = \min_{z\in Z} g'(z);$$
$$g(z) < g'(z) \quad \text{для всех } z \in Z.$$

Теорема 10 Минимизация функционала $Q(F(B_1^*,\dots,B_{p-1}^*,B_p))$ совместно по оператору B_p и корректирующей операции F сводится к минимизации правой части (12.1), которая не зависит от F. Данная задача, в свою очередь, эквивалентна задаче поиска и решения совместной подсистемы максимального веса в системе неравенств

$$B_p(x_j) < B_p(x_k), \text{ с весом } T_{jk}, \quad (j,k) \in \mathbb{D}_{p-1}(B_1^* \dots, B_{p-1}^*). \tag{13.1}$$

Таким образом, задача (5.1) совместной оптимизации оператора B_p и корректирующий операции F сводится к последовательному выполнению трёх шагов:

Шаг 1.

Построить оператор B_p путём решения совместной подсистемы максимального веса в системе неравенств (13.1).

Проблема: постановка задачи (13.1) нестандартная, так как обычно ограничения относятся к отдельным объектам, а не парам объектов.

Шаг 2.

Найти значения $f_k \in \mathfrak{I}_f$, $k \in S$, наименее отклоняющиеся от y_k, для которых последовательность $\{a_k, f_k\}_{k\in S}$ монотонна:

$$\begin{cases} \sum\limits_{k\in S}(f_k - y_k)^2 \to \min; \\ f_j \leqslant f_k, \text{ если } a_j \leqslant a_k. \end{cases}$$

Это частная задача квадратичного программирования. Известны эффективные методы её решения, но в данных лекциях они не рассматриваются.

Шаг 3.

Построить монотонную (в случае восстановления регрессии непрерывную и достаточно гладкую) функцию F, проходящую через q точек, заданных монотонной последовательностью $\{a_k, f_k\}_{k\in S}$:

$$F(a_k) = f_k, \quad k \in S.$$

§14 Теоремы сведения

Введём сокращённые обозначения:

$$\mathbb{D}_{p-1} = \mathbb{D}_{p-1}(B_1^*, \ldots, B_{p-1}^*);$$
$$\mathbb{D}_p = \mathbb{D}_p(B_1^*, \ldots, B_{p-1}^*, B_p).$$

Обозначим через T_k верхнюю оценку числа дефектных пар, в которые может входить k-ый объект при оптимальном выборе корректирующей операции:

$$T_k = \sum_{j \in D_k}^{q} (1 + T_{jk});$$
$$D_k = \bigl\{ j \in S \bigm| (k,j) \in \mathbb{D}_{p-1} \text{ или } (j,k) \in \mathbb{D}_{p-1} \bigr\}.$$

Теорема 11 (о сводимости в задаче классификации)
В случае классификации на 2 непересекающихся класса, когда $\mathfrak{I}_f = \{0,1\}$, $\mathfrak{I}_e = \mathbb{R}$ и

$$Q(B) = \min_{\gamma \in \mathbb{R}} \sum_{k=1}^{q} w_k \bigl| [B(x_k) > \gamma] - y_k \bigr|,$$

построение оператора B_p в задачах (5.2) и (5.1) сводится к поиску и решению совместной подсистемы максимального веса в системе q неравенств относительно оператора $B_p \in \mathfrak{M}^0$ и порога $\gamma \in \mathbb{R}$:

$$\begin{cases} B_p(x_k) \leqslant \gamma, & \text{при } y_k = 0; \\ B_p(x_k) > \gamma, & \text{при } y_k = 1. \end{cases}$$

В задаче (5.2) объекты имеют веса $w_k = 1$, $k \in S$.
В задаче (5.1) объекты имеют веса $w_k = T_k$, $k \in S$.

Теорема 12 (о сводимости в задаче восстановления регрессии)
В случае восстановления регрессии, когда $\mathfrak{I}_f = \mathfrak{I}_e = \mathbb{R}$, $\mathfrak{M}^1 = \{C(B) \equiv B\}$ и

$$Q(B) = \sum_{k=1}^{q} w_k (B(x_k) - y_k)^2,$$

построение оператора B_p в задачах (5.2) и (5.1) сводится к минимизации квадратичного функционала $Q(B_p)$.
В задаче (5.2) объекты имеют веса $w_k = 1$, $k \in S$.
В задаче (5.1) объекты имеют веса $w_k = T_k / d_k^2$, $k \in S$, где

$$d_k = \tfrac{1}{2} \min_{j \in D_k} |y_k - y_j|.$$

Доказательство теорем 11 и 12.

Для задачи (5.2) утверждение теоремы очевидно.

Первая часть доказательства одинакова для теорем 11 и 12.

1. Оценим сверху функционал, минимизируемый в (5.1), с помощью неравенства (12.1):

$$Q(F(B_1^*,\dots,B_{p-1}^*,B_p)) \leqslant |\mathbb{D}_p| + \sum_{(j,k)\in\mathbb{D}_p} T_{jk} =$$
$$= \sum_{(j,k)\in\mathbb{D}_{p-1}} (1+T_{jk})\bigl[(j,k)\in\mathbb{D}(B_p)\bigr] =$$
$$= \sum_{(j,k)\in\mathbb{D}_{p-1}} (1+T_{jk})\bigl[b_j \geqslant b_k\bigr],$$

где введено обозначение $b_k = B_p(x_k)$, для всех $k \in S$.

2. В случае классификации воспользуемся оценкой

$$[b_j \geqslant b_k] \leqslant [b_j > \gamma] + [b_k \leqslant \gamma], \quad \text{для любых } b_j,\ b_k,\ \gamma \text{ из } \mathbb{R}.$$

Учтём также, что $y_j < y_k$ эквивалентно $y_j = 0$, $y_k = 1$ в силу двухэлементности $\mathfrak{I}_f$.

$$\sum_{(j,k)\in\mathbb{D}_{p-1}} (1+T_{jk})\bigl[b_j \geqslant b_k\bigr] \leqslant$$
$$\leqslant \min_{\gamma\in\mathbb{R}} \sum_{(j,k)\in\mathbb{D}_{p-1}} (1+T_{jk})\bigl([b_j>\gamma]+[b_k\leqslant\gamma]\bigr) =$$
$$= \min_{\gamma\in\mathbb{R}} \sum_{(j,k)\in\mathbb{D}_{p-1}} (1+T_{jk})\Bigl(\bigl|[b_j>\gamma]-y_j\bigr| + \bigl|[b_k>\gamma]-y_k\bigr|\Bigr) =$$
$$= \min_{\gamma\in\mathbb{R}} \sum_{\substack{j\in S\\ y_j=0}} \bigl|[b_j>\gamma]-y_j\bigr| \sum_{k:(j,k)\in\mathbb{D}_{p-1}} (1+T_{jk}) + \sum_{\substack{k\in S\\ y_k=1}} \bigl|[b_k>\gamma]-y_k\bigr| \sum_{j:(j,k)\in\mathbb{D}_{p-1}} (1+T_{jk}) =$$
$$= \min_{\gamma\in\mathbb{R}} \sum_{k\in S} \bigl|[b_k>\gamma]-y_k\bigr| \sum_{j\in D_k} (1+T_{jk}) = \min_{\gamma\in\mathbb{R}} \sum_{k\in S} T_k \bigl|[b_k>\gamma]-y_k\bigr|.$$

3. В случае восстановления регрессии воспользуемся оценкой

$$[b_j \geqslant b_k] \leqslant \bigl[|b_j-y_j| \geqslant d_j\bigr] + \bigl[|b_k-y_k| \geqslant d_k\bigr], \quad \text{для любых } (j,k)\in\mathbb{D}_{p-1},$$

затем неравенством $[x \geqslant A] \leqslant \bigl(\frac{x}{A}\bigr)^2$, справедливым при любых неотрицательных x, A.

$$\sum_{(j,k)\in\mathbb{D}_{p-1}} (1+T_{jk})\bigl[b_j \geqslant b_k\bigr] \leqslant$$
$$\leqslant \sum_{(j,k)\in\mathbb{D}_{p-1}} (1+T_{jk})\bigl[|b_j-y_j| \geqslant d_j\bigr] + \bigl[|b_k-y_k| \geqslant d_k\bigr] \leqslant$$
$$= \sum_{j\in S} \bigl[|b_j-y_j| \geqslant d_j\bigr] \sum_{k:(j,k)\in\mathbb{D}_{p-1}} (1+T_{jk}) + \sum_{k\in S} \bigl[|b_k-y_k| \geqslant d_k\bigr] \sum_{j:(j,k)\in\mathbb{D}_{p-1}} (1+T_{jk}) =$$
$$= \sum_{k\in S} T_k \bigl[|b_k-y_k| \geqslant d_k\bigr] \leqslant$$
$$\leqslant \sum_{k\in S} \frac{T_k}{d_k^2}\bigl(b_k-y_k\bigr)^2 = \sum_{k\in S} w_k\bigl(B(x_k)-y_k\bigr)^2. \quad \blacksquare$$

§15 Монотонная аппроксимация в $\mathbb{R}^p$

Пусть $\mathfrak{I}_e = \mathbb{R}$, последовательность $\{a_k, f_k\}_{k\in S}$ монотонная, точки пронумерованы по возрастанию значений $f_1 \leqslant f_2 \ldots \leqslant f_q$.
Рассмотрим задачу построения функции $F \in \mathfrak{F}_M$ такой, что $F(a_k) = f_k$, $k \in S$.

Опр 11 *Области доминирования* вектора a_k, $k \in S$:

$$M_k^{\top} = \{a \in \mathfrak{I}_e^p \mid a_k \leqslant a\} \quad \text{— верхняя область;}$$
$$M_k^{\perp} = \{a \in \mathfrak{I}_e^p \mid a \leqslant a_k\} \quad \text{— нижняя область.}$$

Опр 12 *Функции расстояния от вектора* $a = (a^1, \ldots, a^p)$
до областей доминирования:

$$r_k^{\top}(a) = \mu\left((a_k^1 - a^1)_+, \ldots, (a_k^p - a^p)_+\right);$$
$$r_k^{\perp}(a) = \mu\left((a^1 - a_k^1)_+, \ldots, (a^p - a_k^p)_+\right);$$

где $\mu : \mathbb{R}_+^p \to \mathbb{R}_+$ — заданная неубывающая функция, принимающая нулевое значение только в точке $(0, \ldots, 0)$.

Опр 13 *Функции расстояния от вектора* $a = (a^1, \ldots, a^p)$ *до ближайшей области доминирования на уровне* γ, где $f_1 \leqslant \gamma < f_q$:

$$h_k^{\top}(a, \gamma) = \min_{k: f_k > \gamma} r_k^{\top}(a);$$
$$h_k^{\perp}(a, \gamma) = \min_{k: f_k \leqslant \gamma} r_k^{\perp}(a);$$

Теорема 13 (о монотонном корректоре для задачи классификации)
Пусть в задаче классификации $\mathfrak{I}_f = \{0, 1\}$ и $f_1 < f_q$. Тогда функция

$$F(a) = \left[h_k^{\top}(a, 0) \leqslant h_k^{\perp}(a, 0)\right]$$

определена и монотонно не убывает на $\mathbb{R}^p$, причём $F(a_k) = f_k$ для всех $k \in S$.

Теорема 14 (о монотонном корректоре для задачи восстановления регрессии)
Пусть в задаче восстановления регрессии $\mathfrak{I}_f = \mathbb{R}$ и $f_1 < f_q$. Тогда функция

$$F(a) = f_1 + \sum_{k=1}^{q-1} (f_{k+1} - f_k)\Phi(a, f_k), \text{ где } \Phi(a, \gamma) = \frac{h_k^{\perp}(a, \gamma)}{h_k^{\perp}(a, \gamma) + h_k^{\top}(a, \gamma)},$$

определена и монотонно не убывает на $\mathbb{R}^p$, причём $F(a_k) = f_k$ для всех $k \in S$.
Функция $F(a)$ непрерывна, если функция $\mu(a)$ непрерывна.

§16 Примеры монотонной аппроксимации в $\mathbb{R}^2$

Дискретная ступенька $F(a)$, $a \in \mathbb{R}^2$, в задаче классификации.

Непрерывная ступенька $\Phi(a, f_k)$, $a \in \mathbb{R}^2$, на уровне f_k в задаче восстановления регрессии.

Монотонная корректирующая операция $F(a)$, $a \in \mathbb{R}^2$, в задаче восстановления регрессии.

Классический гладкий сплайн, построенный по монотонной выборке, может оказаться немонотонной функцией.