

© 2022 г. А.И. БАЗАРОВА (bazarova.ai@phystech.edu)
(Московский физико-технический институт),
А.В. ГРАБОВОЙ (grabovoy.av@phystech.edu)
(Московский физико-технический институт),
В.В. СТРИЖОВ, д-р физ.-мат. наук (strijov@phystech.edu)
(Вычислительный центр им. А.А. Дородницына ФИЦ ИУ РАН)

Анализ свойств вероятностных моделей в задачах обучения с экспертом¹

Данная работа посвящена построению интерпретируемых моделей машинного обучения. Решается задача аппроксимации набора фигур на контурном изображении. Вводятся предположения, что фигуры являются кривыми второго порядка. При аппроксимации фигур используются информация о типе, расположении и форме кривых, а также о множестве их возможных преобразований. Такая информация называется *экспертной*, а метод машинного обучения, основанный на экспертной информации, называется *обучение с экспертом*. Предполагается, что набор фигур аппроксимируется набором *локальных моделей*. Каждая локальная модель, основанная на экспертной информации, аппроксимирует одну фигуру на контурном изображении. Для построения моделей предлагается отображать кривые второго порядка в пространство признаков, в которой каждая локальная модель является линейной. Таким образом, кривые второго порядка аппроксимируются набором линейных моделей. В вычислительном эксперименте рассматривается задача аппроксимации радужной оболочки глаза на контурном изображении.

Ключевые слова: смесь экспертов; экспертное обучение; линейные модели; интерпретируемые модели.

1. Введение

Современные решения задачи классификации изображений на основе сетей глубокого обучения ResNet, VGG, Intercept [2] представляют собой плохо интерпретируемые модели [1]. В [3, 4] показано, что сети глубокого обучения являются не устойчивыми даже к небольшому шуму в данных [5].

В данной работе предлагается метод *обучение с экспертом*. Метод предполагает использование экспертной информации для повышения качества аппроксимации, а также для получения интерпретируемых моделей машинного обучения. Предметные знания экспертов о данных называются *экспертная информация*. Предполагается, что использование экспертной информации позволяет аппроксимировать выборку

¹Исследование выполнено при финансовой поддержке Российского Фонда Фундаментальных Исследований в рамках научных проектов 20-37-90050, 19-07-01155 и проекта Национальной Технологической Инициативы 13/1251/2018.

простыми интерпретируемыми моделями, такими как линейные модели. Метод машинного обучения, учитывающий экспертные знания при построении моделей, называется *обучение с экспертом*. В данной работе решается задача аппроксимации кривых второго порядка на контурном изображении. В работе анализируются кривые второго порядка, так как они описываются линейными моделями. Параметры кривых второго порядка необходимо восстановить в задаче распознавания радужной оболочки глаза [6–8], в задаче описания трека частицы в адронном коллайдере [9]. Экспертная информация о кривой второго порядка отображает точки на плоскости в новое признаковое описание объекта. Каждая кривая аппроксимируется одной линейной моделью. Модель, аппроксимирующая кривую, называется *локальной моделью*. Для аппроксимации всего контурного изображения требуется аппроксимация нескольких кривых второго порядка с использованием локальных моделей. Вводятся следующие ограничения на изображения: а) изображение состоит только из кривых второго порядка; б) изображение аппроксимируется небольшим числом кривых второго порядка; в) известно число и тип кривых на изображении.

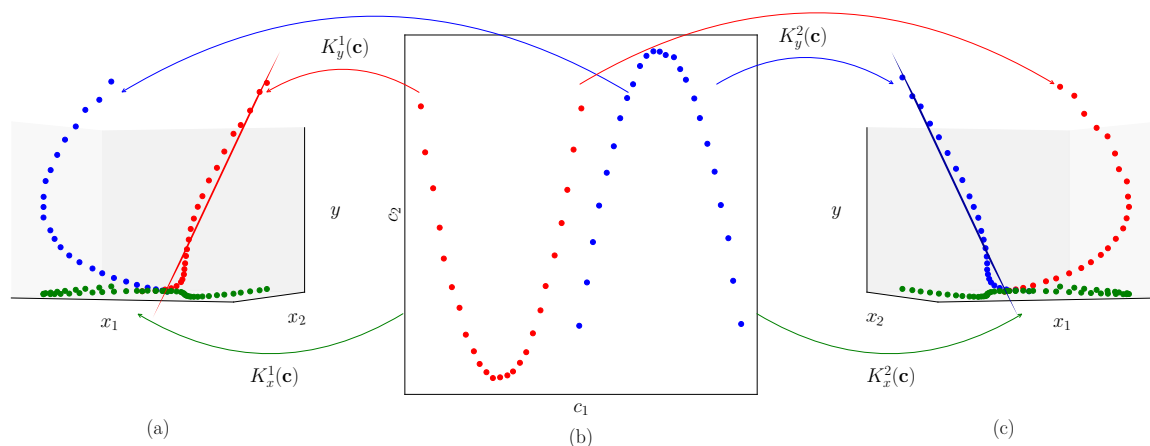


Рис. 1. Визуализация экспертной информации в случае с двумя экспертами: а) экспертная информация первого эксперта; б) исходные данные; в) экспертная информация второго эксперта

На рис. 1 показан пример кривых второго порядка, а также экспертная информация о кривых. Отображение $K_x^1(c)$, $K_x^2(c)$, $K_y^1(c)$, $K_y^2(c)$ основаны на экспертной информации представлены. На рис. 1 слева показана экспертная информация о первом эксперте. Используя эту информацию, первая кривая аппроксимируется линейной моделью, а вторая кривая представляет собой шум. На рис. 1 справа показана экспертная информация второго эксперта. Используя эту информацию, вторая кривая аппроксимируется линейной моделью, а первая кривая представляет собой шум.

При аппроксимации нескольких кривых на одном контурном изображении строится мультимодель. Примером мультимodelей является случайный лес [10], бустинг деревьев [11], смесь экспертов [12]. В данной работе в качестве мультимodelи рассматривается смесь экспертов. Смесь экспертов — это мультимодель, которая линейно взвешивает локальные модели, которые аппроксимируют часть выборки. Значения весовых коэффициентов зависят от объекта, для которого делается прогноз. Для решения задачи используется вариационный ЕМ-алгоритм [13–15, 24]. Смесь экспертов имеет множество применений в ряде приложений. В [16] решается задача

классификации текстов. В [17–23] смесь экспертов используется для прогнозирования временных рядов, для распознавания речи, распознавания повседневной деятельности человека и прогнозирования стоимости ценных бумаг. В [14] рассматривалась смесь экспертов для решения задачи распознавания рукописных цифр на изображениях.

В качестве примера рассматривается задача аппроксимации изображения радужной оболочки глаза. На рис. 2 показан пример изображения, которое необходимо аппроксимировать. В данной работе рассматривается обработанное изображение в виде контурного изображения.

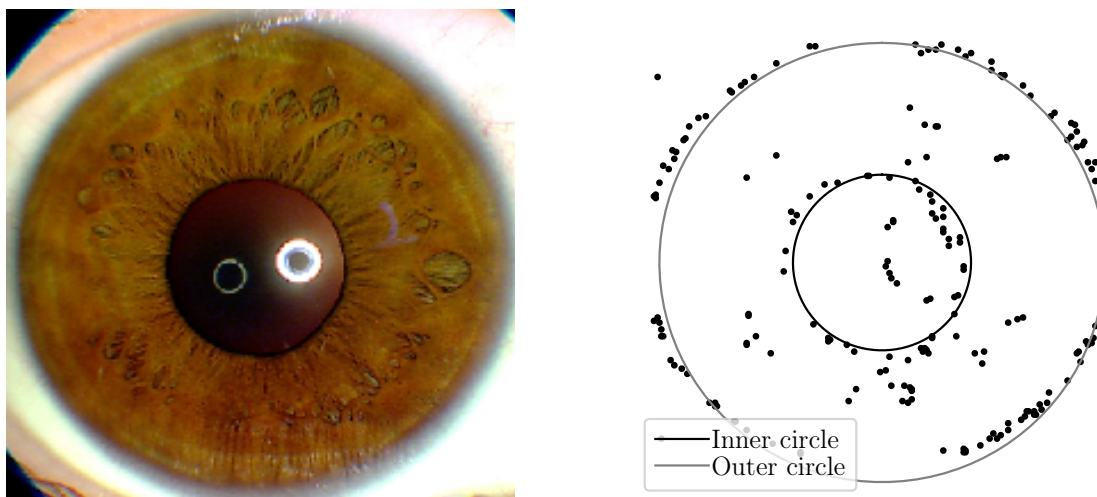


Рис. 2. Пример изображения радужной оболочки глаза и его контурное представление. Слева: изображение радужной оболочки глаза. Справа: контурное изображение радужной оболочки и аппроксимирующие заданное изображение окружности

Для задачи аппроксимации радужной оболочки используется экспертная информация: радужная оболочка глаза аппроксимируется двумя концентрическими окружностями. Экспертная информация используется для построения описания признаков точек плоскости, а также для построения регуляризатора в функции оптимизации. Часть функции ошибок для оптимизации, использующая экспертную информацию, называется регуляризатором. Таким образом, информация о том, что изображение являются окружностями задается видом признакового описания, а информация о том, что окружности концентрические задается с помощью специального регуляризатора.

Вычислительный эксперимент анализирует качество аппроксимации контурного изображения в зависимости от экспертной информации и уровня шума в синтетически сформированных данных. Проведен анализ качества аппроксимации радужной оболочки в зависимости от объема экспертной информации, которая использовалась для построения модели. Каждое аппроксимированное изображение представляет собой отдельный набор точек, которые необходимо аппроксимировать.

2. Постановка задачи восстановления параметров кривых второго порядка на изображении

Задано бинарное изображение:

$$\mathbf{M} \in \{0, 1\}^{m_1 \times m_2},$$

где 1 соответствует черной точке изображения, а 0 соответствует белой точке фона. Используя изображение \mathbf{M} строится выборка \mathbf{C} , элементами которой являются координаты (x_i, y_i) черных точек:

$$\mathbf{C} \in \mathbb{R}^{N \times 2}.$$

Каждый эксперт предполагает, что изображение состоит из кривой второго порядка Ω . Пусть для множества точек $\mathbf{C} \in \mathbb{R}^{N \times 2}$, образующих кривую Ω , задана экспертная информация о фигуре $E(\Omega)$. Множество $E(\Omega)$ состоит из ожидаемого экспертом образа Ω и множества его допустимых преобразований. На основе экспертного описания введем отображения в задачу аппроксимации:

$$(1) \quad K_x(E(\Omega)) : \mathbb{R}^2 \rightarrow \mathbb{R}^n, \quad K_y(E(\Omega)) : \mathbb{R}^2 \rightarrow \mathbb{R},$$

где K_x — отображение объектов в признаковое описание объектов, n — число признаков, а K_y — отображение в целевую переменную для аппроксимации. Применяя отображения K_x, K_y для выборки \mathbf{C} поэлементно получаем:

$$(2) \quad K_x(E(\Omega), \mathbf{c}) = \mathbf{x}, \quad K_y(E(\Omega), \mathbf{c}) = y,$$

где $\mathbf{c} = (x_i, y_i)$ точка из множества точек \mathbf{C} .

Применяя отображения (2) к исходному множеству точек \mathbf{C} , получаем выборку

$$(3) \quad \mathfrak{D} = \{(\mathbf{x}, y) \mid \forall \mathbf{c} \in \mathbf{C} \mathbf{x} = K_x(\mathbf{c}), y = K_y(\mathbf{c})\}.$$

Получаем, что исходная задача аппроксимации кривой Ω сводится к аппроксимации выборки \mathfrak{D} . В работе предполагается, что выборка \mathfrak{D} аппроксимируется линейной моделью:

$$g(\mathbf{x}, \mathbf{w}) = \mathbf{x}^T \mathbf{w},$$

где \mathbf{w} — вектор параметров для аппроксимации.

Для нахождения оптимального вектора параметров $\hat{\mathbf{w}}$ решается оптимизационная задача:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \sum_{(\mathbf{x}, y) \in \mathfrak{D}} \|g(\mathbf{x}, \mathbf{w}) - y\|_2^2.$$

Задача аппроксимации исходной кривой Ω сводится к решению задачи линейной регрессии, т.е. к нахождению компонент вектора $\hat{\mathbf{w}}$.

В случае, когда на изображении K кривых второго порядка $\Omega_1, \dots, \Omega_K$, где для каждой задана экспертная информация $E_k = E(\Omega_k)$, $k \in \{1, \dots, K\}$. Получаем задачу построения мультимодели, которая называется смесью K экспертов.

Определение 1. Назовем мультимодель f смесью K экспертов

$$f = \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{V}) g_k(\mathbf{w}_k), \quad \pi_k(\mathbf{x}, \mathbf{V}) : \mathbb{R}^{n \times |\mathbf{V}|} \rightarrow [0, 1], \quad \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{V}) = 1,$$

где g_k — локальная модель, называемая экспертом. Вектор \mathbf{x} — признаковое описание объекта, π_k — шлюзовая функция, вектор \mathbf{w}_k — параметры локальной модели, \mathbf{V} являются параметрами шлюзовой функции.

Для каждой кривой второго порядка Ω_k заданы отображения (1). Введем обозначения: $K_x^k(\mathbf{c}) = K_x(\Omega_k, \mathbf{c})$ и $K_y^k(\mathbf{c}) = K_y(\Omega_k, \mathbf{c})$. Используя локальные линейные модели g_k строится мультимодель f , описывающая кривые $\Omega_1, \dots, \Omega_K$ на изображении \mathbf{M} :

$$(4) \quad f = \sum_{\mathbf{c} \in \mathbf{C}} \sum_{k=1}^K \pi_k(\mathbf{c}, \mathbf{V}) g_k(K_x^k(\mathbf{c}), \mathbf{w}_k),$$

где π_k — шлюзовая функция. В работе рассматривается случай, когда $\mathbf{x} = K_x^1(\mathbf{c}) = \dots = K_x^K(\mathbf{c})$. В этом случае выражение (4) переписывается в виде:

$$f = \sum_{\mathbf{c} \in \mathbf{C}} \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{V}) g_k(\mathbf{x}, \mathbf{w}_k),$$

где шлюзовая функция π_k имеет вид:

$$\pi_k(\mathbf{x}, \mathbf{V}) : \mathbb{R}^{n \times |\mathbf{V}|} \rightarrow [0, 1], \quad \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{V}) = 1,$$

где \mathbf{V} — параметры функции шлюза, а g_k — локальная модель.

В работе рассматривается следующий вид шлюзовой функции:

$$\pi(\mathbf{x}, \mathbf{V}) = \text{softmax}(\mathbf{V}_1^T \boldsymbol{\sigma}(\mathbf{V}_2^T \mathbf{x})),$$

где $\mathbf{V} = \{\mathbf{V}_1, \mathbf{V}_2\}$ — параметры шлюзовой функции, $\mathbf{V}_1 \in \mathbb{R}^{p \times k}$, $\mathbf{V}_2 \in \mathbb{R}^{n \times p}$.

Для нахождения оптимальных параметров мультимодели решается оптимизационная задача:

$$(5) \quad \mathcal{L} = \sum_{(\mathbf{x}, y) \in \mathcal{D}} \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{V}) (y - \mathbf{w}_k^T \mathbf{x})^2 + R(\mathbf{V}, \mathbf{W}, E(\Omega)) \rightarrow \min_{\mathbf{V}, \mathbf{W}},$$

где $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_k]$ — параметры локальных моделей, $R(\mathbf{V}, \mathbf{W}, E(\Omega))$ — параметры регуляризации, основанные на экспертной информации.

3. Построение признакового описания фигур

Единое пространство для кривых второго порядка. Произвольная кривая второго порядка, главная ось которой не параллельна оси ординат, задается уравнением:

$$x^2 = B'xy + C'y^2 + D'x + E'y + F',$$

где на коэффициенты B', C' накладываются ограничения, зависящие от типа кривой. Выражение (2) для кривой второго порядка принимает вид:

$$K_x(\mathbf{c}_i) = [x_i y_i, y_i^2, x_i, y_i, 1], \quad K_y(\mathbf{c}_i) = x_i^2,$$

откуда получаем задачу линейной регрессии для восстановления параметров B', C', D', E', F' по заданной выборке.

Окружность. В качестве частного случая кривой второго порядка рассматривается окружность. Пусть (x_0, y_0) — центр окружности, которую требуется восстановить на бинарном изображении \mathbf{M} , а r — ее радиус. Элементы выборки $(x_i, y_i) \in \mathbf{C}$ представляют собой геометрическое место точек, которое аппроксимируется уравнением окружности:

$$(2x_0) \cdot x_i + (2y_0) \cdot y_i + (r^2 - x_0^2 - y_0^2) \cdot 1 = x_i^2 + y_i^2.$$

Тогда выражение (2) принимают следующий вид:

$$K_x(\mathbf{c}_i) = [x_i, y_i, 1] = \mathbf{x}, \quad K_y(\mathbf{c}_i) = x_i^2 + y_i^2 = y.$$

Получаем задачу линейной регрессии (3). Компоненты вектора $\mathbf{w} = [w_0, w_1, w_2]^T$, восстанавливают параметры окружности:

$$x_0 = \frac{w_0}{2}, \quad y_0 = \frac{w_1}{2}, \quad r = \sqrt{w_2 + x_0^2 + y_0^2}.$$

4. Композиция фигур

Для построения композиции фигур используется уравнение (5), которое принимает вид:

$$\mathcal{L} = \sum_{\mathbf{c} \in \mathbf{C}} \sum_{k=1}^K \pi_k(\mathbf{c}, \mathbf{V}) (K_y^k(\mathbf{c}) - \mathbf{w}_k^T K_x^k(\mathbf{c}))^2 + R(\mathbf{V}, \mathbf{W}, E(\Omega)) \rightarrow \min_{\mathbf{V}, \mathbf{W}},$$

где K_x^k, K_y^k экспертное представление k -го эксперта. Предполагая, что все кривые на изображении описываются одним описанием признаков $\mathbf{x} = K_x^1(\mathbf{c}) = \dots = K_x^K(\mathbf{c})$, $y = K_y^1(\mathbf{c}) = \dots = K_y^K(\mathbf{c})$, получаем оптимизационную задачу:

$$(6) \quad \mathcal{L} = \sum_{(\mathbf{x}, y) \in \mathcal{D}} \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{V}) (y - \mathbf{w}_k^T \mathbf{x})^2 + R(\mathbf{V}, \mathbf{W}, E(\Omega)) \rightarrow \min_{\mathbf{V}, \mathbf{W}},$$

где регуляризатор R учитывает дополнительные ограничения параметров локальных моделей. Для решения задачи оптимизации (6) используется ЕМ-алгоритм описанный в работе [24].

5. Вычислительный эксперимент

Проведен вычислительный эксперимент по анализу качества аппроксимации кривых второго порядка на изображении. Эксперимент разделен на несколько частей. Первая часть описывает эксперимент с несколькими окружностями на изображении. Во второй части анализируется сходимость метода в зависимости от уровня шума в данных и от заданной экспертной информации. В третьей части проводится эксперимент по аппроксимации радужной оболочки глаза.

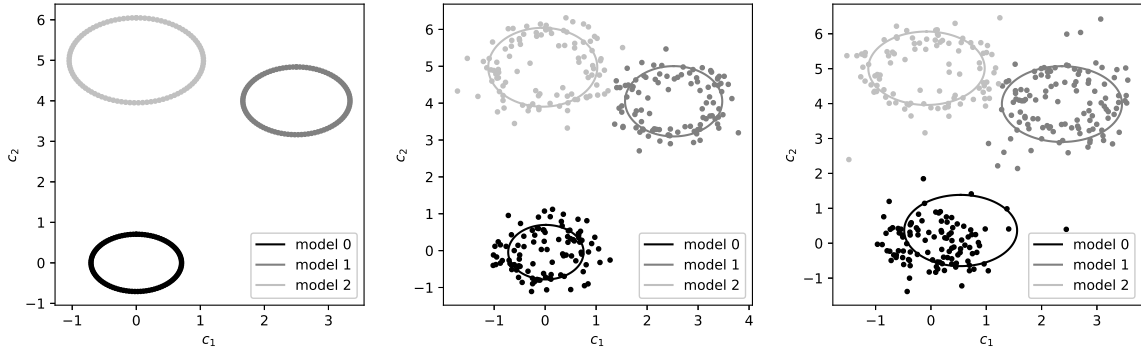


Рис. 2. Мультимодель в зависимости от уровня шума в выборке. Слева направо: окружности без шума; шум в радиусе круга; шум в радиусе круга, шум по всему изображению

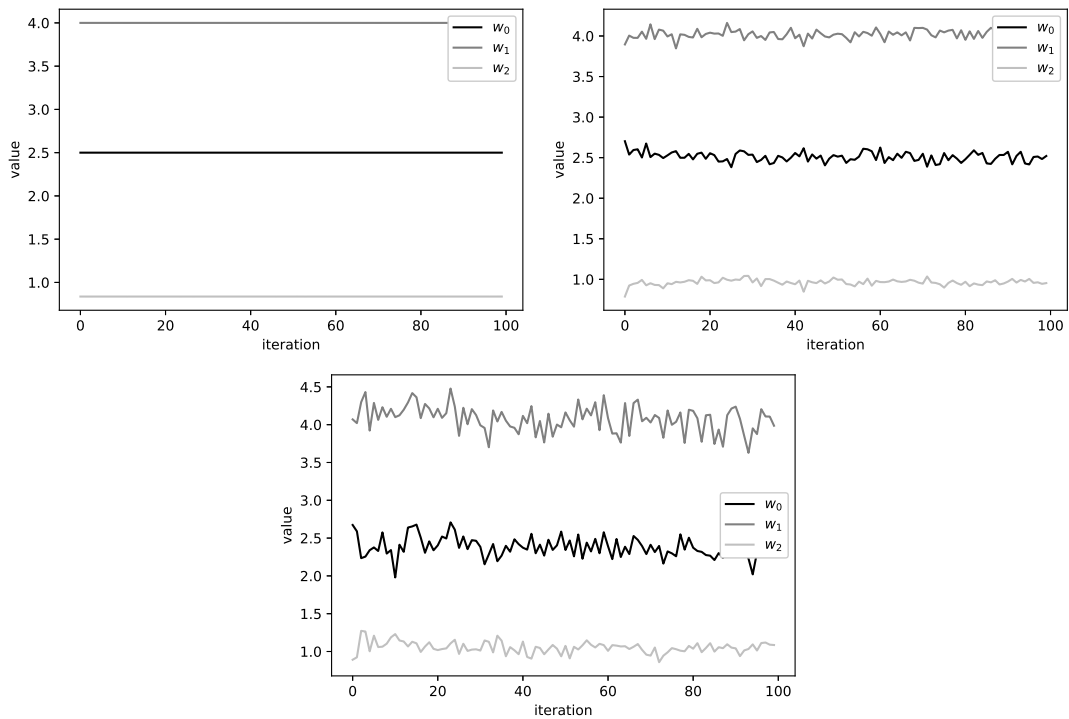


Рис. 3. Зависимость параметров r , x_0 и y_0 от номера итерации в зависимости от уровня шума в выборке. Слева направо: окружности без шума; шум в радиусе круга; шум в радиусе круга, шум по всему изображению

5.1. Эксперимент по восстановлению параметров окружности

В этой части эксперимента анализируется аппроксимация нескольких кривых второго порядка предложенной мультимоделью. Аппроксимируется сгенерированная синтетическая выборка. Выборка состоит из трех произвольных непересекающихся окружностей. К окружностям добавлен шум. Шум добавлялся к каждой точке окружности по отдельности, а также в выборку добавлялись случайные точки, не относящиеся к окружностям.

На рис. 3 показан результат построения ансамбля локальных аппроксимирующих моделей. Каждая локальная модель аппроксимирует одну окружность. Видно, что при добавлении шума качество аппроксимации падает. На рис. 4 показан график зависимости радиуса окружностей r и их центров (x_0, y_0) от номера итерации.

5.2. Анализ качества аппроксимации для выборки с разным уровнем шума

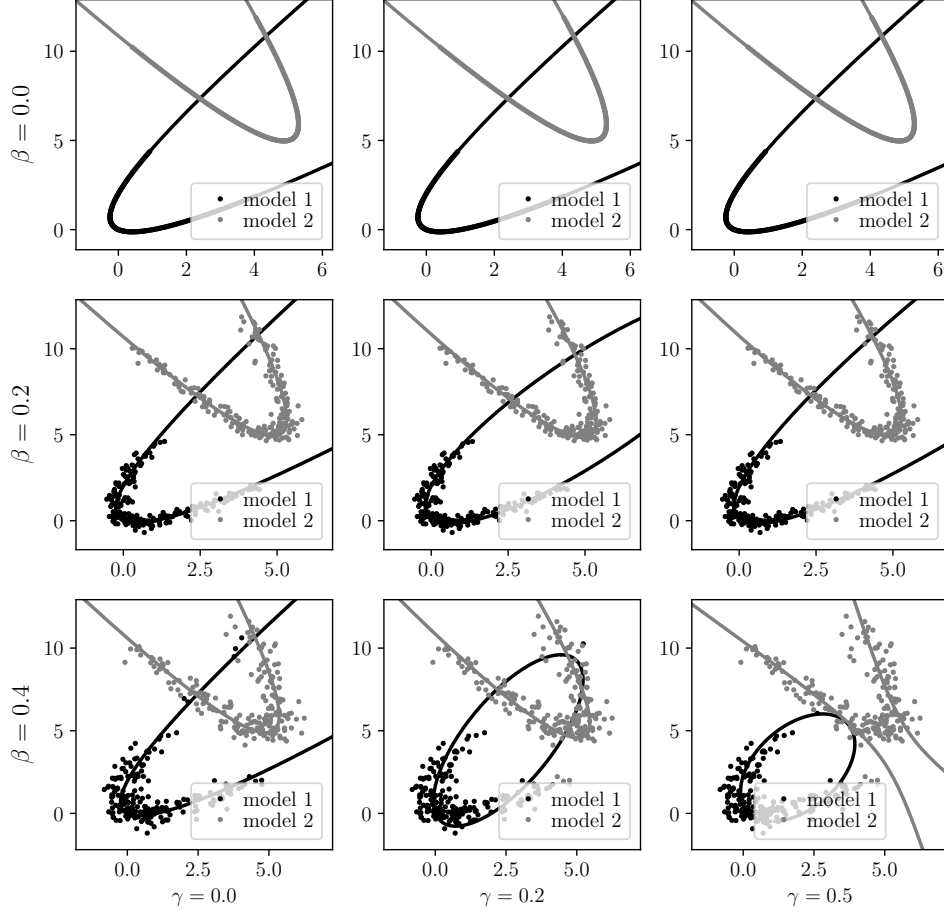


Рис. 4. Результат аппроксимации данных с разным уровнем шума β и по дисперсии априорного распределения γ

В этой части эксперимента анализируется качество аппроксимации S в зависимости от уровня шума β в данных и от параметра априорных распределений γ . Выборка сгенерирована следующим образом: сначала случайным образом выбираются два вектора параметров $\mathbf{w}_1^{\text{true}}$ и $\mathbf{w}_2^{\text{true}}$ — коэффициенты двух парабол. Векторы $\mathbf{w}_1^{\text{true}}$ и $\mathbf{w}_2^{\text{true}}$ используются для генерации точек x_i и y_i с добавлением нормального шума $\varepsilon \sim \mathcal{N}(0, \beta)$. При обучении мультимодели учитывается априорное распределение параметров $\mathbf{w}_1 \sim \mathcal{N}(\mathbf{w}_1^{\text{true}}, \gamma \mathbf{I})$, $\mathbf{w}_2 \sim \mathcal{N}(\mathbf{w}_2^{\text{true}}, \gamma \mathbf{I})$.

Рассматривается критерий качества:

$$S = \|\mathbf{w}_1^{\text{pred}} - \mathbf{w}_1^{\text{true}}\|_2^2 + \|\mathbf{w}_2^{\text{pred}} - \mathbf{w}_2^{\text{true}}\|_2^2,$$

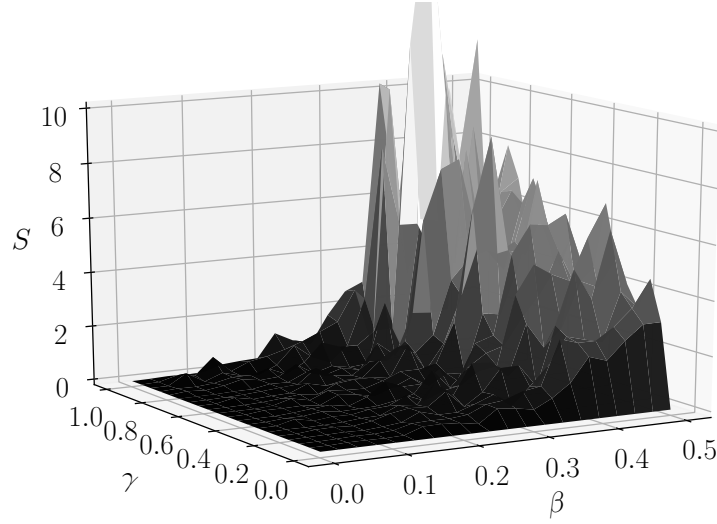


Рис. 5. Зависимость моделей от уровня шума β в данных, а также от дисперсии априорного распределения γ

где $\mathbf{w}_1^{\text{pred}}$ аппроксимация вектора параметров первой локальной модели, а $\mathbf{w}_2^{\text{pred}}$ аппроксимация вектора параметров второй локальной модели.

На рис. 6 показана зависимость критерия качества S от уровня шума β и параметра априорного распределения γ . Из графика видно, что при малом уровне шума β качество аппроксимации не зависит от параметра γ , а при увеличении шума β качество аппроксимации S уменьшается.

На рис. 5 показан пример работы алгоритма с разными параметрами β и γ . Видно, что в отсутствие шума β обе локальные модели аппроксимируют выборку корректно. С ростом уровня шума качество аппроксимации падает: при $\beta = 0,2$ с ростом γ первая локальная модель из параболы преобразовывается в эллипс; для $\beta = 0,4$ с ростом γ первая локальная модель из параболы преобразовывается в эллипс, а вторая модель из параболы в гиперболу.

5.3. Аппроксимация радужки глаза

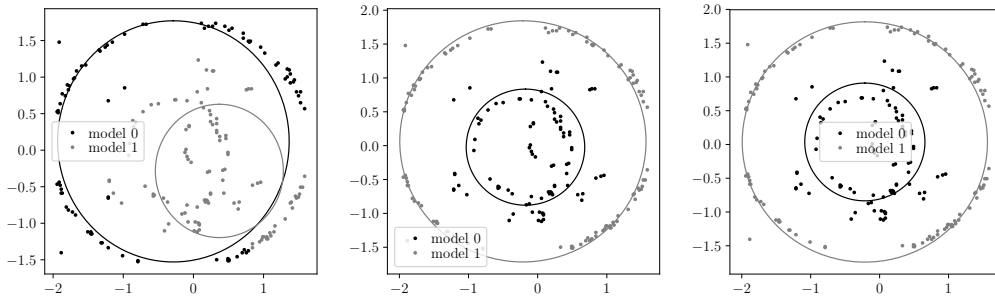


Рис. 6. Визуализация аппроксимации радужной оболочки. Слева на право: если указан регуляризатор R_0 ; если указан регуляризатор R_1 ; если указан регуляризатор R_2

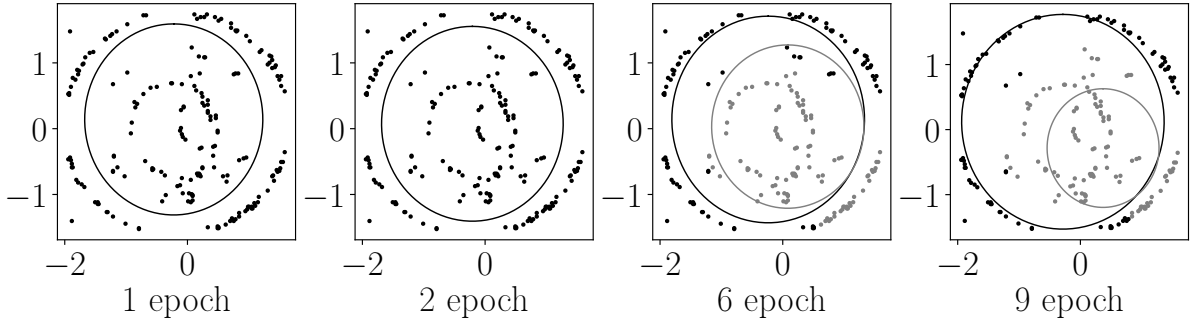


Рис. 7. Визуализация мультимодели в случае регуляризатора R_0

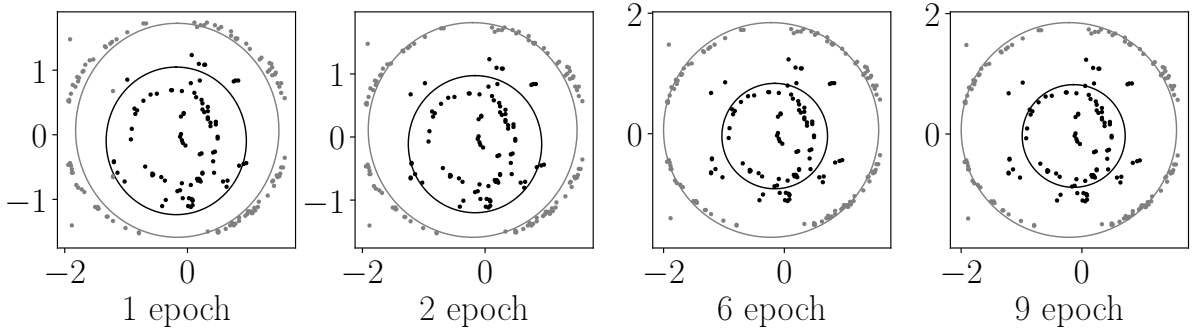


Рис. 8. Визуализация мультимодели в случае регуляризатора R_1

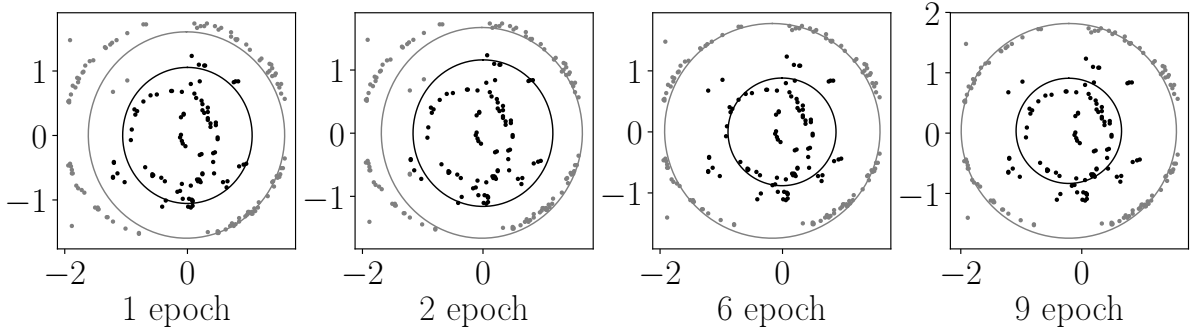


Рис. 9. Визуализация мультимодели в случае регуляризатора R_2

Анализ качества аппроксимации проводится для задачи аппроксимации радужной оболочки глаза на изображении. Радужная оболочка глаза состоит из двух концентрических окружностей, поэтому рассматривается мультимодель, состоящая из двух экспертов: каждый эксперт аппроксимирует одну из окружностей. В вычислительном эксперименте сравнивается качество аппроксимации окружностей в случае задания разных регуляризаторов R_0, R_1, R_2 . Регуляризатор $R_0(\mathbf{V}, \mathbf{W}, E(\Omega)) = 0$, что

соответствует отсутствию регуляризатора. Регуляризатор

$$R_1(\mathbf{V}, \mathbf{W}, E(\Omega)) = - \sum_{k=1}^K \mathbf{w}_k^T \mathbf{w}_k$$

способствует околонулевым параметрам локальных моделей. Регуляризатор

$$R_2(\mathbf{V}, \mathbf{W}, E(\Omega)) = - \sum_{k=1}^K \mathbf{w}_k^T \mathbf{w}_k + \sum_{k=1}^K \sum_{k'=1}^K \sum_{j=1}^2 (w_k^j - w_{k'}^j)^2$$

способствует совпадению центров окружностей и близким к нулю параметрам локальных моделей.

Рис. 7 показывает результат аппроксимации радужной оболочки глаза после 10 итераций. Видно, что при отсутствии регуляризатора одна из окружностей находится некорректно. В случае задания регуляризатора R_1 , модель аппроксимирует обе окружности, но окружности не концентричны. В случае задания регуляризатора R_2 получаем концентрические окружности на изображении.

На рис. 8–10 показан процесс сходимости мультимodelей в случае указания разных регуляризаторов R_0, R_1, R_2 . Видно, что модели с регуляризатором типа R_1 и R_2 аппроксимируют обе окружности, а мультимodelь с регуляризатором R_0 аппроксимирует только большую окружность.

6. Заключение

В статье предлагается метод построения интерпретируемых моделей машинного обучения на основе экспертной информации. В качестве задачи рассматривается задача аппроксимации кривых второго порядка: параболы, гиперболы, эллипса. Аппроксимация кривых второго порядка применяется в задаче об аппроксимации радужной оболочки.

Проведен эксперимент, в ходе которого анализируется качество аппроксимации кривых второго порядка в зависимости от начального уровня шума в данных, а также в зависимости от регуляризатора функции ошибки. В ходе эксперимента показано, что с увеличением уровня шума в исходных данных точность аппроксимации снижается: при большом шуме форма аппроксимируемой фигуры меняется с параболы на гиперболу. Проведен вычислительный эксперимент по аппроксимации радужной оболочки глаза двумя концентрическими окружностями. Эксперимент показывает, что регуляризация на основе экспертной информации улучшает качество аппроксимации.

СПИСОК ЛИТЕРАТУРЫ

1. *Ribeiro M., Singh S., Guestrin C.* Why Should I Trust You?: Explaining the Predictions of Any Classifier // Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. P. 1135–1144.

2. *He K., Ren S., Sun J., Zhang X.* Deep Residual Learning for Image Recognition // Proc. IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas. 2016. P. 770–778.
3. *Han X., Yao M., Debayan D., Hui L., Ji-Liang T., Anil J.* Adversarial Attacks and Defenses in Images, Graphs and Text: A Review // International Journal of Automation and Computing. 2020. V. 17. P. 151–178.
4. *Akhtar N., Mian A.* Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey // IEEE Access. 2018. Vol. 6. P. 14410–14430.
5. *Grabovoy A., Strijov V.* Probabilistic Interpretation of the Distillation Problem // Automat. Remote Control. 2022. V. 83. P. 123–137.
6. *Matveev I.* Detection of iris in image by interrelated maxima of brightness gradient projections // Appl. Comput. Math. 2010. Vol. 9. P. 252–257.
7. *Matveev I., Simonenko I.* Detecting precise iris boundaries by circular shortest path method // Pattern Recognition and Image Analysis. 2014. Vol. 24. P. 304–309.
8. *Bowyer K., Hollingsworth K., Flynn P.* A Survey of Iris Biometrics Research: 2008–2010 // Handbook of iris recognition. 2010. P. 15–54.
9. *Salamani D., Gadatsch S., Golling T., Stewart G., Ghosh A., Rousseau D., Hasib A., Schaarschmidt J.* Deep Generative Models for Fast Shower Simulation in ATLAS // IEEE 14th International Conference on e-Science. 2018. P. 348–348.
10. *Chen Xi., Ishwaran H.* Random Forests for Genomic Data Analysis // Genomics. 2012. Vol. 6. P. 323–329.
11. *Chen T., Guestrin C.* XGBoost: A Scalable Tree Boosting System // Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. P. 785–794.
12. *Yuksel S., Wilson J., Gader P.* Twenty Years of Mixture of Experts // IEEE Transactions on Neural Networks and Learning Systems. 2012. Vol. 8. P. 1177–1193.
13. *Dempster A., Laird N., Rubin D.* Maximum Likelihood from Incomplete Data via the EM Algorithm // Journal of the Royal Statistical Society. Series B (Methodological). 1977. Vol. 39. P. 1–38.
14. *Ebrahimpour R., Moradian R., Esmkhani A., Jafarlou F.* Recognition of Persian handwritten digits using characterization loci and mixture of experts // J. Digital Content Technol. Appl. 2009. P. 42–46.
15. *Peng F., Jacobs R., Tanner M.* Bayesian inference in mixtures-of-experts and hierarchical mixtures-of-experts models with an application to speech recognition // J. Amer. Stat. Assoc. 1996. Vol. 91. P. 953–960.
16. *Estabrooks A., Japkowicz N.* A mixture-of-experts framework for text classification // Proc. Workshop Comput. Natural Lang. Learn., Assoc. Comput. Linguist. 2001. P. 1–8.

17. *Cheung Y., Leung W., Xu L.* Application of mixture of experts model to financial time series forecasting // Proc. Int. Conf. Neural Netw. Signal Process. 1995. P. 1–4.
18. *Weigend A., Shi S.* Predicting daily probability distributions of S&P500 returns // J. Forecast. 2000. Vol. 19. P. 375–392.
19. *Cao L.* Support vector machines experts for time series forecasting // Neurocomputing. 2003. Vol. 51. P. 321–339.
20. *Mossavat S., Amft O., Vries B., Petkov P., Kleijn W.* A Bayesian hierarchical mixture of experts approach to estimate speech quality // Proc. 2nd Int. Workshop Qual. Multimedia Exper. 2010. P. 200–205.
21. *Sminchisescu C., Kanaujia A., Metaxas D.* Discriminative density propagation for visual tracking // IEEE Trans. Pattern Anal. Mach. Intell. 2007. Vol. 29. P. 2030–2044.
22. *Tuerk A.* The state based mixture of experts HMM with applications to the recognition of spontaneous speech // Ph.D. thesis, University of Cambridge. 2001.
23. *Yumlu M., Gorgen F., Okay N.* Financial time series prediction using mixture of experts // Proc. 18th Int. Symp. Comput. Inf. Sci. 2003. P. 553–560.
24. *Grabovoy A., Strijov V.* Prior Distribution Selection for a Mixture of Experts // Comput. Math. and Math. Phys. 2021. P. 1140–1152.