

Анализ свойств вероятностных моделей в задачах обучения с экспертом

А. И. Базарова¹, А. В. Грабовой², В. В. Стрижов³

Аннотация: Решается задача аппроксимации заданного набора фигур на изображении. Вводятся предположения о том, что фигуры являются кривыми второго порядка. При оптимизации параметров вероятностных моделей используется информация о виде этих кривых и множестве их возможных преобразований. Такая информация называется *экспертной*, а метод машинного обучения, основанный на *экспертной* информации, называется *обучением с экспертом*. В работе для построения моделей предлагается отобразить кривые второго порядка в новое признаковое пространство, в котором каждая локальная модель является линейной моделью. Таким образом, кривые высших порядков распознаются при помощи композиции линейных моделей. В работе поставлена и решена задача оптимизации для поиска оптимальных параметров мультимодели. В вычислительном эксперименте проводится анализ качества аппроксимации в зависимости от уровня шума в данных, также в зависимости от экспертной информации.

Ключевые слова: смесь моделей, обучение с экспертом, линейные модели

DOI: 00.00000/0000000000000000

1 Введение

В данной работе предлагается метод *обучения с экспертом*. Данный метод предполагает использование предметных знаний экспертов для повышения качества аппроксимации моделей машинного обучения. Предметные знания экспертов о выборке назовем *экспертной информацией*. В данной работе в качестве экспертной информации рассматривается информация о виде кривой второго порядка. Методы машинного обучения, которые учитывают экспертные знания при построении моделей, назовем *обучением с экспертом*.

¹Московский физико-технический институт, bazarova.ai@phystech.edu

²Московский физико-технический институт, grabovoy.av@phystech.edu

³Московский физико-технический институт, strijov@ccas.ru

В работе решается задача распознавания кривых второго порядка на изображении. Кривые второго порядка выбраны для анализа, так как их легко аппроксимировать линейными моделями. При этом эти фигуры требуется восстановить в таких прикладных задачах, как задача распознавания радужки глаза [13, 14, 15], задача описания трека частицы в адронном коллайдере. Коэффициенты уравнений кривых второго порядка, описывающих данные кривые, аналитически выражаются через параметры построенных моделей.

Для каждого класса перечисленных кривых предлагается, согласно экспертной информации, отобразить точки на изображении в новое признаковое пространство и затем оптимально аппроксимировать кривую. Отображение строится таким образом, чтобы в новом признаковом описании каждая кривая описывалась линейной моделью. В данной работе рассматривается задача поиска фигур на изображении в предположении, что число и тип фигур на изображении известны.

При распознавании нескольких кривых на одном изображении строится мульти-модель. Примерами мультимodelей являются случайный лес [2], бустинг деревьев [1], смесь экспертов [3]. В данной работе в качестве мультимodelи рассматривается смесь экспертов. Смесь экспертов — это мультимodelь, которая линейно взвешивает локальные модели, аппроксимирующие выборку. Значения весовых коэффициентов зависят от того объекта, для которого производится предсказание. Для решения задачи смеси экспертов используется вариационный ЕМ-алгоритм [5, 18, 16]. Смесь экспертов имеет множество применений в ряде прикладных задач. В работе [10] решается задача классификации текстов. В работах [7, 8, 4, 11, 12, 17, 6] смесь экспертов используется для предсказания временных рядов в задачах распознавания речи, дневной активности человека и предсказания стоимости ценных бумаг. В работе [9] смесь экспертов рассматривалась для решения задачи распознавания рукописных цифр на изображениях.

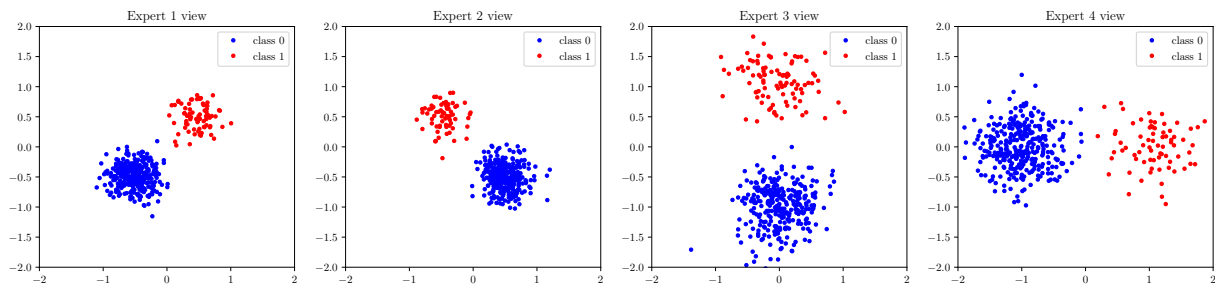


Рис. 1: Представление объектов разными экспертами

В отличие от постановки задачи построения смеси экспертов [3], задача обучения с экспертом предполагает, что каждая модель имеет собственное признаковое описание объектов, которое построено на основе экспертной информации. Например, рассмотрим следующую задачу: пусть объекты это люди, а в качестве экспертов рассмотрим врачей, требуется определить больной человек или нет. На рис. 1 показаны представления о людях разными врачами — экспертами. Класс 1 у каждого эксперта отвечает

тому, что человек имеет болезнь по профессии соответствующего врача. Видно, что в пространстве каждого эксперта люди с соответствующей болезнью линейно отделимы от остальных людей с точностью 100%. Заметим, что если объединить все признаки в одно пространство и построить линейную модель, то получим качество 95% точности.

Для оптимизации параметров ансамбля моделей вводится функция ошибки. Она состоит из двух слагаемых: функции регуляризации, вид которой основан на экспертной информации, и ошибки локальных линейных моделей. Качество мультимодели оценивается с помощью интерпретируемого критерия качества.

В вычислительном эксперименте проводится анализ качества аппроксимации в зависимости от заданных априорных предположений и от уровня шума в данных.

2 Постановка задачи нахождения параметров кривых второго порядка на изображении

Задано бинарное изображение. Эксперт предполагает, что на нем изображен конечный набор кривых второго порядка:

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

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

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

Построение признакового описания на основе экспертной информации. Пусть для набора точек $\mathbf{C} \in \mathbb{R}^{N \times 2}$, образующих кривую Ω , задано экспертное описание $E(\Omega)$. Множество $E(\Omega)$ состоит из ожидаемого экспертом вида фигуры Ω и множества ее допустимых преобразований. На основе экспертного описания введем отображения

$$K_x(E(\Omega)) : \mathbb{R}^2 \rightarrow \mathbb{R}^n, \quad K_y(E(\Omega)) : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (2.1)$$

иными словами, поэлементно:

$$K_x(E(\Omega), \mathbf{c}) = \mathbf{x}, \quad K_y(E(\Omega), \mathbf{c}) = y. \quad (2.2)$$

Здесь n — число признаков, $\mathbf{c} = (x_i, y_i)$ — точка из выборки \mathbf{C} . Рассмотрим линейную модель

$$g : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, \quad (2.3)$$

$$g(\mathbf{x}, \mathbf{w}) = \mathbf{x}^\top \mathbf{w}. \quad (2.4)$$

Вектор $\hat{\mathbf{w}} = [w_0, w_1, \dots, w_n]$ является решением следующей оптимизационной задачи:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \|g(\mathbf{x}, \mathbf{w}) - y\|_2^2. \quad (2.5)$$

Применяя отображения (2.2) к исходному набору точек \mathbf{C} , получим выборку

$$\mathfrak{D} = \{(\mathbf{x}, y) \mid \forall \mathbf{c} \in \mathbf{C} \mathbf{x} = K_x(\mathbf{c}), y = K_y(\mathbf{c})\}. \quad (2.6)$$

Таким образом, задача определения коэффициентов уравнения исходной фигуры сводится к решению задачи линейной регрессии, т. е. нахождения компонент вектора $\hat{\mathbf{w}}$, связывающего полученные \mathbf{x} и y .

Мультимодель. В случае, когда на изображении K кривых второго порядка $\Omega_1, \dots, \Omega_K$, для каждой из которых имеется экспертная информация $E_k = E(\Omega_k)$, $k \in \{1, \dots, K\}$, ставится задача построения мультимодели, называемой смесью K экспертов.

Определение 2.1. Назовем мультимодель f смесью K экспертов

$$f = \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{V}) g_k(\mathbf{w}_k), \quad \pi_k(\mathbf{x}, \mathbf{V}) : \mathbb{R}^{2 \times |\mathbf{V}|} \rightarrow [0, 1], \quad \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{V}) = 1, \quad (2.7)$$

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

Для каждой кривой второго порядка заданы отображения (2.1). Тогда, используя локальные линейные модели, построим универсальную мультимодель, описывающую все множество кривых $\Omega_1, \dots, \Omega_K$ на изображении \mathbf{M} :

$$f = \sum_{(\mathbf{x}, y) \in \mathfrak{D}} \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{V}) g_k(\mathbf{x}, \mathbf{w}_k), \quad (2.8)$$

где π_k — шлюзовая функция:

$$\pi_k(\mathbf{x}, \mathbf{V}) : \mathbb{R}^{2 \times |\mathbf{V}|} \rightarrow [0, 1], \quad \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{V}) = 1, \quad (2.9)$$

\mathbf{V} — параметры шлюзовой функции, а g_k — локальная линейная модель вида (2.4).

В данной работе

$$\pi(\mathbf{x}, \mathbf{V}) = \text{softmax}(\mathbf{V}_1^T \boldsymbol{\sigma}(\mathbf{V}_2^T \mathbf{x})), \quad (2.10)$$

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 необходимо решить следующую задачу оптимизации:

$$\mathcal{L} = \sum_{(\mathbf{x}, y) \in \mathfrak{D}} \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{V}) (y - \mathbf{w}_k^T \mathbf{x})^2 + R(\mathbf{V}, \mathbf{W}, E(\Omega)) \rightarrow \min_{\mathbf{V}, \mathbf{W}}, \quad (2.11)$$

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

3 Построение признакового описания фигур

Окружность. Пусть (x_0, y_0) — центр окружности, которую необходимо найти на бинарном изображении \mathbf{M} , а r — ее радиус. Элементы выборки $(x_i, y_i) \in \mathbf{C}$ являются геометрическим местом точек, которое аппроксимируется уравнением окружности:

$$(x_i - x_0)^2 + (y_i - y_0)^2 = r^2. \quad (3.1)$$

Раскрыв скобки, получим:

$$(2x_0) \cdot x_i + (2y_0) \cdot y_i + (r^2 - x_0^2 - y_0^2) \cdot 1 = x_i^2 + y_i^2. \quad (3.2)$$

Тогда отображения (2.2) примут вид:

$$K_x(\mathbf{c}_i) = [x_i, y_i, 1] = \mathbf{x}, K_y(\mathbf{c}_i) = x_i^2 + y_i^2 = y. \quad (3.3)$$

Поставим задачу линейной регрессии (2.6). Компоненты вектора $\mathbf{w} = [w_0, w_1, w_2]^T$, связывающего \mathbf{x} и y , восстанавливают параметры окружности:

$$x_0 = \frac{w_0}{2}, y_0 = \frac{w_1}{2}, r = \sqrt{w_2 + x_0^2 + y_0^2}. \quad (3.4)$$

Эллипс. Элементы выборки $(x_i, y_i) \in \mathbf{C}$ являются ГМТ, которое аппроксимируется уравнением общим уравнением эллипса:

$$Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + Ey_i + F = 0, \quad B^2 - 4AC < 0. \quad (3.5)$$

Из условия на коэффициенты A, B, C следует, что $A \neq 0, C \neq 0$. Перепишем уравнение:

$$B'x_iy_i + C'y_i^2 + D'x_i + E'y_i + F' = x_i^2. \quad (3.6)$$

В этом случае отображения (2.2) имеют вид:

$$K_x(\mathbf{c}_i) = [x_iy_i, y_i^2, x_i, y_i, 1] = \mathbf{x}, K_y(\mathbf{c}_i) = x_i^2 = y. \quad (3.7)$$

Поставим задачу линейной регрессии (2.6). Оптимальные параметры линейной регрессии и коэффициенты уравнения связаны следующим образом:

$$w_0 = B' = -\frac{B}{A}, w_1 = C' = -\frac{C}{A}, w_2 = D' = -\frac{D}{A}, w_3 = E' = -\frac{E}{A}, w_4 = F' = -\frac{F}{A}, \quad (3.8)$$

с учетом

$$B^2 - 4AC < 0, \quad (3.9)$$

то есть

$$w_0^2 + 4w_1 < 0. \quad (3.10)$$

Парабола. Элементы выборки \mathbf{C} аппроксимируются общим уравнением параболы:

$$Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + Ey_i + F = 0, \quad B^2 = 4AC. \quad (3.11)$$

где Рассмотрим различные варианты расположения параболы относительно координатных осей.

В случае, если ось симметрии параболы параллельна Ox , из экспертных данных следует, что коэффициенты общего уравнения параболы

$$A = B = 0. \quad (3.12)$$

Тогда общее уравнение параболы, аппроксимирующее выборку \mathbf{C} , приобретает вид

$$Cy_i^2 + Dx_i + Ey_i + F = 0. \quad (3.13)$$

Перепишем:

$$y_i^2 = D'x_i + E'y_i + F'. \quad (3.14)$$

Тогда вид преобразований (2.2):

$$K_x(\mathbf{c}_i) = [x_i, y_i, 1] = \mathbf{x}, \quad K_y(\mathbf{c}_i) = y_i^2 = y. \quad (3.15)$$

Поставим задачу линейной регрессии (2.6). Параметры параболы выражаются через оптимальные параметры линейной регрессии:

$$w_0 = D' = -\frac{D}{C}, \quad w_1 = E' = -\frac{E}{C}, \quad w_2 = F' = -\frac{F}{C}. \quad (3.16)$$

В случае, если ось симметрии параболы параллельна Oy , то аналогично предыдущему случаю, общее уравнение приобретет вид:

$$x_i^2 = D'x_i + E'y_i + F'. \quad (3.17)$$

Преобразования (2.2) будут иметь вид:

$$K_x(\mathbf{c}_i) = [x_i, y_i, 1] = \mathbf{x}, \quad K_y(\mathbf{c}_i) = x_i^2 = y. \quad (3.18)$$

В случае, если ось симметрии параболы не параллельна ни одной из координатных осей, то в таком случае $A \neq 0$, $B \neq 0$, $C \neq 0$.

Тогда общее уравнение параболы, аппроксимирующее выборку, имеет вид:

$$x_i^2 = B'x_iy_i + C'y_i^2 + D'x_i + E'y_i + F'. \quad (3.19)$$

Преобразования (2.2):

$$K_x(\mathbf{c}_i) = [x_iy_i, y_i^2, x_i, y_i, 1] = \mathbf{x}, \quad K_y(\mathbf{c}_i) = x_i^2 = y. \quad (3.20)$$

Поставим задачу линейной регрессии (2.6). Исходные параметры параболы восстанавливаются по вектору оптимальных параметров линейной регрессии \mathbf{w} следующим образом:

$$w_0 = B' = -\frac{B}{A}, \quad w_1 = C' = -\frac{C}{A}, \quad w_2 = D' = -\frac{D}{A}, \quad w_3 = E' = -\frac{E}{A}, \quad w_4 = F' = -\frac{F}{A}. \quad (3.21)$$

Гипербола. Элементы выборки \mathbf{C} аппроксимируются общим уравнением гиперболы:

$$Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + Ey_i + F = 0, \quad (3.22)$$

где

$$B^2 - 4AC > 0. \quad (3.23)$$

В случае, если полуоси гиперболы параллельны осям координат, то уравнение гиперболы имеет вид:

$$Ax_i^2 + Cy_i^2 + Dx_i + Ey_i + F = 0, \quad (3.24)$$

где

$$AC < 0. \quad (3.25)$$

Перепишем уравнение:

$$C'y_i^2 + D'x_i + E'y_i + F' = x_i^2. \quad (3.26)$$

Преобразования (2.2):

$$K_x(\mathbf{c}_i) = [y_i^2, x_i, y_i, 1] = \mathbf{x}, \quad K_y(\mathbf{c}_i) = x_i^2 = y. \quad (3.27)$$

В случае, если полуоси гиперболы непараллельны осям координат, то $B \neq 0$. Перепишем уравнение гиперболы:

$$A'x_i^2 + C'y_i^2 + D'x_i + E'y_i + F' = x_iy_i. \quad (3.28)$$

Преобразования (2.2):

$$K_x(\mathbf{c}_i) = [x_i^2, y_i^2, x_i, y_i, 1] = \mathbf{x}, \quad K_y(\mathbf{c}_i) = x_iy_i = y. \quad (3.29)$$

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

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

где на коэффициенты B', C' накладываются ограничения, которые зависят от вида кривой. Выражение (2.2) принимает следующий вид:

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

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

4 Композиция фигур

Для построения композиции фигур воспользуемся выражением (2.11) принимает следующий вид:

$$\mathcal{L} = \sum_{\mathbf{c} \in \mathbf{C}} \sum_{k=1}^K \pi_k(\mathbf{c}, \mathbf{V}) (K_y^k(\mathbf{c}) - \mathbf{w}_k^\top K_x^k(\mathbf{c}))^2 + R(\mathbf{V}, \mathbf{W}, E(\Omega)) \rightarrow \min_{\mathbf{V}, \mathbf{W}}, \quad (4.1)$$

где K_x^k, K_y^k экспертное представление k -го эксперта. В качестве регуляризатора R рассматриваются дополнительные ограничения на вектора параметров моделей: априорное распределение параметров $\mathbf{w}_1, \dots, \mathbf{w}_k$, а также связь между параметрами вида (3.10). Для решения оптимизационной задачи (4.1) предлагается использовать ЕМ-алгоритм.

5 Вычислительный эксперимент

Проведен вычислительный эксперимент для анализа качества моделей кривых второго порядка на изображении. Эксперимент разделен на несколько частей. В первой части проводится анализ сходимости метода в случае, когда задана единственная кривая. Во второй части проводится эксперимент с несколькими окружностями на изображении. В третьей части проводится анализ сходимости метода в зависимости от уровня шума в данных и от априорных предположений.

5.1 Эксперимент с окружностями

В данной части эксперимента показан пример обучения мультимодели для аппроксимации нескольких фигур второго порядка одновременно. В качестве данных используется синтетическая выборка Synthetic 2, которая получена при помощи генерации трех произвольных непересекающихся окружностей, а также добавления к данным окружностям шума. Шум добавлялся к радиусу окружности для каждой точки, также в выборку были добавлены случайные точки, которые не относятся к окружностям. В эксперименте сравниваются две модели: в первой модели регуляризатор $R(\mathbf{V}, \mathbf{W}, E(\Omega)) = 0$, то есть модель без задания регуляризатора, во второй модели регуляризатор:

$$R(\mathbf{V}, \mathbf{W}, E(\Omega)) = - \sum_{k=1}^K \gamma (\mathbf{w}_k - \mathbf{w}_k^0)^\top (\mathbf{w}_k - \mathbf{w}_k^0),$$

где \mathbf{w}_k^0 априорные предположения о векторе параметров.

На рис. 2 показан результат построения ансамбля локально аппроксимирующих моделей, которые аппроксимируют выборку. Каждая локальная модель аппроксимирует одну окружность, причем при добавлении разного шума, качество аппроксимации падает. На рис. 3 показан график зависимости радиуса окружностей r и их центров

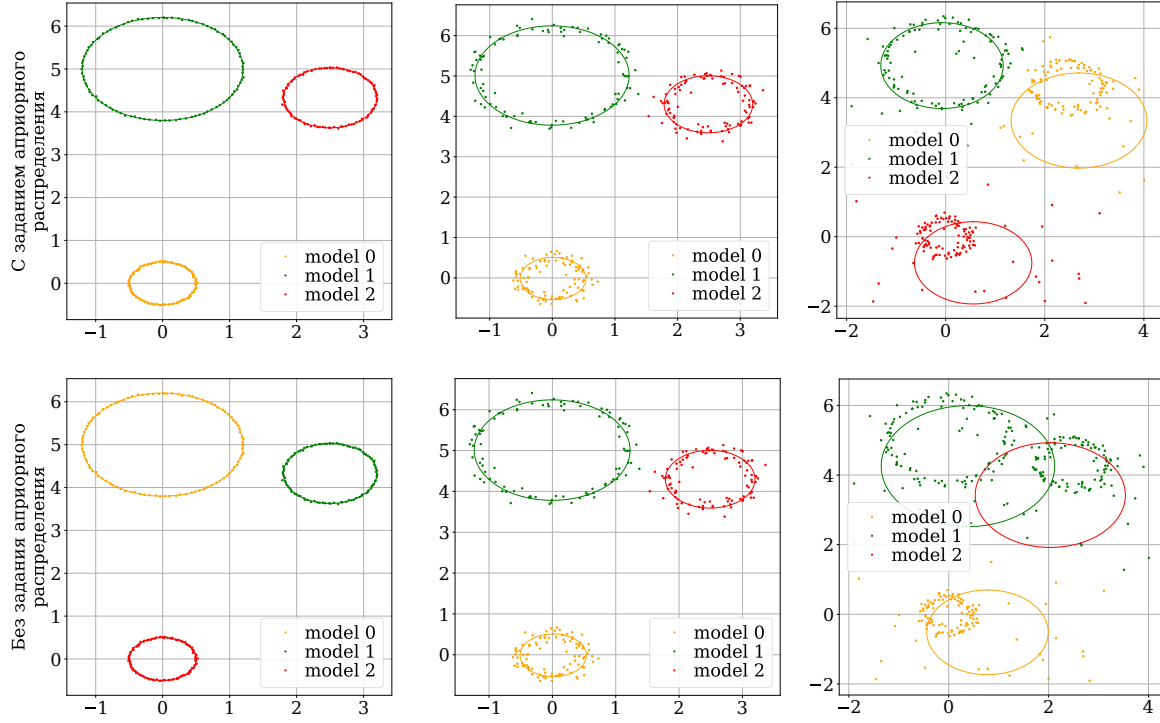


Рис. 2: Мультимодель в зависимости от разных априорных предположений и уровня шума. Сверху вниз: построение с заданием априорного распределения; без задания априорного распределения. Слева на право: окружности без шума; шум в радиусе окружности; шум в радиусе окружности а также произвольные точки по всему изображению.

(x_0, y_0) от номера итерации. Видно, что модель с заданием априорного распределения сходится быстрее чем модель без задания априорного распределения.

5.2 Эксперимент с разным уровнем шума и дисперсии априорного распределения

В данной части эксперимента проводится анализ качества аппроксимации S от уровня шума β в данных и от параметра априорных распределений γ . Выборка получена следующим образом: сначала случайным образом выбирается два вектора параметров $\mathbf{w}_1^{\text{true}}$ и $\mathbf{w}_2^{\text{true}}$ — коэффициенты двух парабол. На основе векторов $\mathbf{w}_1^{\text{true}}$ и $\mathbf{w}_2^{\text{true}}$ выполняется генерация точек x_i и y_i с добавлением нормального шума $\varepsilon \sim \mathcal{N}(0, \beta)$. При обучении мультимодели в качестве априорного распределения параметров рассматривается $\mathbf{w}_1 \sim \mathcal{N}(\mathbf{w}_1^{\text{true}}, \gamma \mathbf{I})$, $\mathbf{w}_2 \sim \mathcal{N}(\mathbf{w}_2^{\text{true}}, \gamma \mathbf{I})$.

Рассматривается следующий критерий качества:

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

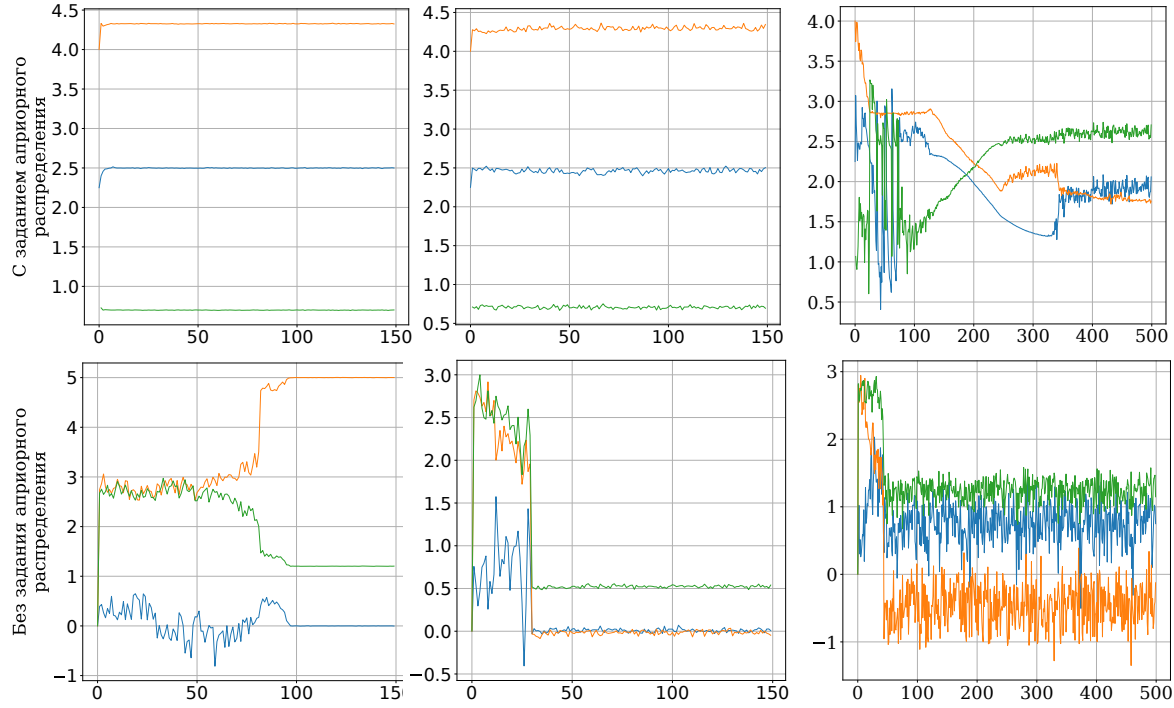


Рис. 3: Зависимость параметров r , x_0 и y_0 от номера итерации при разных априорных распределениях. Сверху вниз: построение с заданием априорного распределения; без задания априорного распределения. Слева на право: окружности без шума; шум в радиусе окружности; шум в радиусе окружности а также произвольные точки по всему изображению.

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. 5 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 падает.

На рис. 5 показан пример работы алгоритма при разных параметрах β и γ . Видно, что в случае отсутствия шума β обе локальные модели аппроксимируют выборку. При увеличении уровня шума качество аппроксимации падает: при $\beta = 0,2$ при увеличении γ первая локальная модель из параболы переходит в эллипс; при $\beta = 0,4$ при увеличении γ первая локальная модель из параболы переходит в эллипс, а вторая модель из параболы переходит в гиперболу.

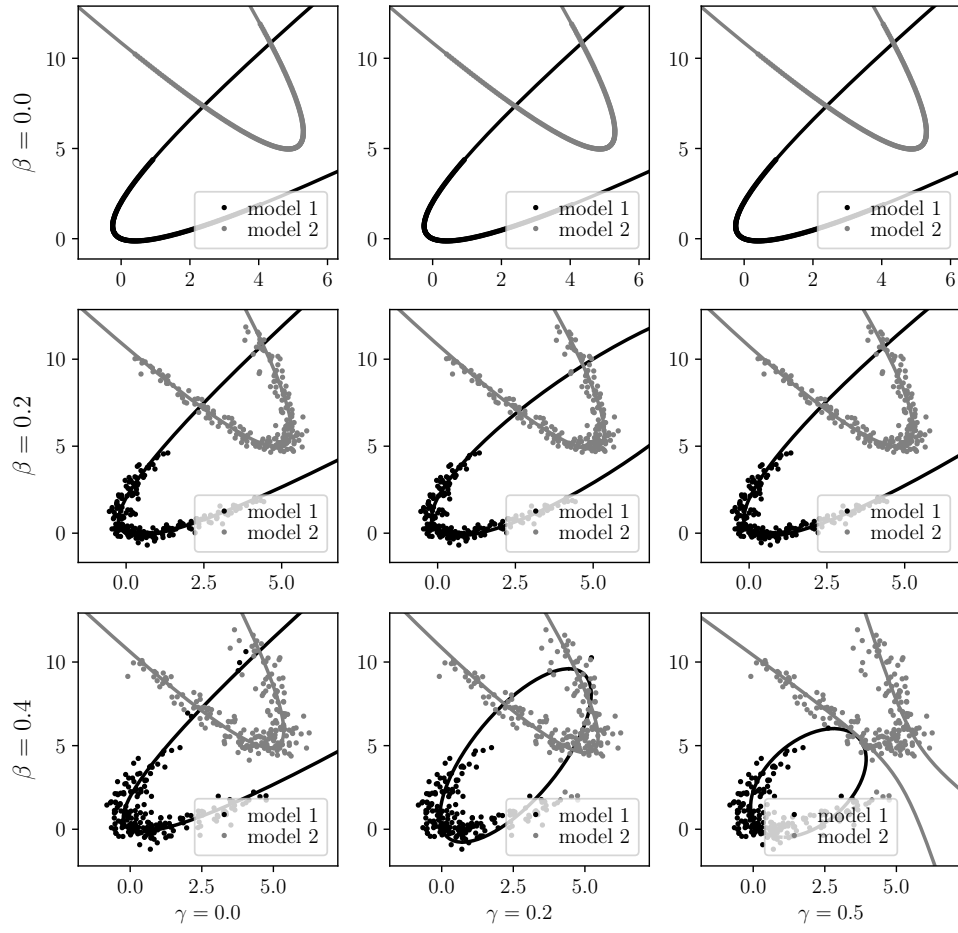


Рис. 4: Результат аппроксимации для данных с разным уровнем шума β и от дисперсии априорного распределения γ

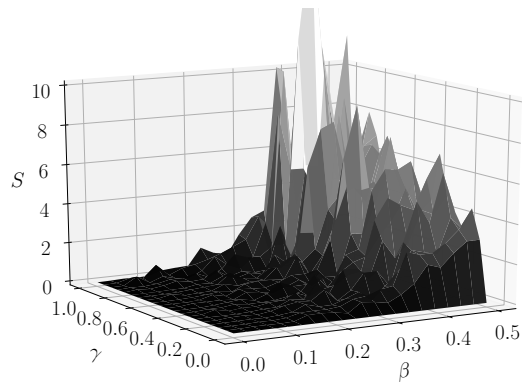


Рис. 5: Зависимость моделей от уровня шума β в данных, а также от дисперсии априорного распределения γ

6 Заключение

В данной работе проведен анализ смеси экспертов при различных априорных распределениях параметров локальных моделей. В качестве данных использовались кривые второго порядка: парабола, гипербола, эллипс. В качестве шлюзовой функции использовалась двухслойная нейросеть.

В вычислительно эксперименте проведено сравнение сходимости ЕМ-алгоритма при задании априорного распределения и без него. Проведен эксперимент, в котором анализируется качество аппроксимации в зависимости от начального уровня шума в данных, а также в зависимости от параметров регуляризатора. В ходе эксперимента показано, что при увеличении уровня шума в начальных данных, точность аппроксимации падает: при большом шума вид аппроксимируемой фигуры изменяется с параболы на гиперболу.

Список литературы

- [1] *Chen Tianqi, Guestrin Carlos* XGBoost: A Scalable Tree Boosting System // KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016.
- [2] *Chen Xi, Ishwaran Hemant* Random Forests for Genomic Data Analysis // Genomics. 2012. Issues. 99, No 6. pp. 323–329.
- [3] *Yuksel Seniha Esen, Wilson Joseph N., Gader Paul D* Twenty Years of Mixture of Experts // IEEE Transactions on Neural Networks and Learning Systems. 2012. Issues. 23, No 8. pp. 1177–1193.
- [4] *L. Cao* Support vector machines experts for time series forecasting // Neurocomputing, vol. 51, pp. 321–339, Apr. 2003.
- [5] *A. P. Dempster, N. M. Laird and D. B. Rubin* Maximum Likelihood from Incomplete Data via the EM Algorithm // Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, No. 1 pp. 1-38, 1977.
- [6] *M. S. Yumlu, F. S. Gurgen, N. Okay* Financial time series prediction using mixture of experts // in Proc. 18th Int. Symp. Comput. Inf. Sci., 2003, pp. 553–560.
- [7] *Y. M. Cheung, W. M. Leung, and L. Xu* Application of mixture of experts model to financial time series forecasting // in Proc. Int. Conf. Neural Netw. Signal Process., 1995, pp. 1–4.
- [8] *A. S. Weigend, S. Shi* Predicting daily probability distributions of S&P500 returns // J. Forecast., vol. 19, no. 4, pp. 375–392, 2000.

- [9] *R. Ebrahimpour, M. R. Moradian, A. Esmkhani, F. M. Jafarlou* Recognition of Persian handwritten digits using characterization loci and mixture of experts // J. Digital Content Technol. Appl., vol. 3, no. 3, pp. 42–46, 2009.
- [10] *A. Estabrooks, N. Japkowicz* A mixture-of-experts framework for text classification //in Proc. Workshop Comput. Natural Lang. Learn., Assoc. Comput. Linguist., 2001, pp. 1–8.
- [11] *S. Mossavat, O. Amft, B. de Vries, P. Petkov, W. Kleijn* A Bayesian hierarchical mixture of experts approach to estimate speech quality // in Proc. 2nd Int. Workshop Qual. Multimedia Exper., pp. 200–205., 2010
- [12] *C. Sminchisescu, A. Kanaujia, and D. Metaxas* B M3 E: Discriminative density propagation for visual tracking // IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 11, pp. 2030–2044, 2007.
- [13] *I. Matveev* Detection of iris in image by interrelated maxima of brightness gradient projections // Appl.Comput. Math. 9 (2), 252–257, 2010.
- [14] *I. Matveev, I. Simonenko.* Detecting precise iris boundaries by circular shortest path method // Pattern Recognition and Image Analysis. 24. 304-309. 2014.
- [15] *K. Bowyer, K. Hollingsworth, P. Flynn* A Survey of Iris Biometrics Research: 2008–2010.
- [16] *F. Peng, R. A. Jacobs, M. A. Tanner* Bayesian inference in mixtures-of-experts and hierarchical mixtures-of-experts models with an application to speech recognition // J. Amer. Stat. Assoc., vol. 91, no. 435, pp. 953–960, 1996.
- [17] *A. Tuerk* The state based mixture of experts HMM with applications to the recognition of spontaneous speech // Ph.D. thesis, Dept. Eng., Univ. Cambridge, Cambridge, U.K., 2001.
- [18] *Bishop C.* Pattern Recognition and Machine Learning. — Berlin: Springer, 2006. 758 p.