Анализ априорных распределений в задаче смеси экспертов

Грабовой Андрей Валериевич

Московский физико-технический институт

МФТИ, г. Долгопрудный

Цель: предложить метод построения ансамбля локально аппроксимирующих локальных моделей для поиска окружностей на изображении.

Задачи

1. Предложить метод поиска окружности при помощи линейной модели, для поиска нескольких окружностей предложить метод построения смеси локальных аппроксимирующих моделей.
2. Предложить априорные распределения на параметры локальных моделей.

Исследуемая проблема

1. Снижение размерности пространства описаний изображения.

Метод решения

В качестве мультимодели предлагается использовать смесь экспертов. Для повышения качества мультимодели предлагается ввести априорные распределения параметров локальных моделей. Вводится *регуляризация* априорных распределений для учета взаимосвязи между априорными распределениями разных локальных моделей смеси.

1. *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.
2. *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.
3. *Bishop C.* Pattern Recognition and Machine Learning. — Berlin: Springer, 2006. 758 p.
4. *I. Matveev* Detection of iris in image by interrelated maxima of brightness gradient projections // Appl.Comput. Math. 9 (2), 252–257, 2010.
5. *K. Bowyer, K. Hollingsworth, P. Flynn* A Survey of Iris Biometrics Research: 2008–2010.

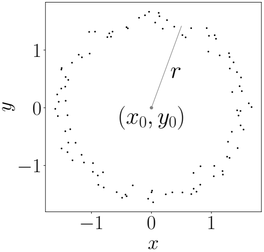
Задано бинарное изображение:

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

где 1 — черная точка, 0 — белая точка фона.
По изображению $\mathbf{M}$ строится выборка $\mathbf{C}$:

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

где N — число черных точек на изображении $\mathbf{M}$.



Пусть x_0, y_0 — центр окружности, которую требуется найти, а r ее радиус.
Точки $(x_i, y_i) \in \mathbf{C}$ должны удовлетворять уравнению окружности:

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

Задачу линейной регрессии для нахождения окружности:

$$\mathbf{Xw} \approx \mathbf{y}, \quad \mathbf{X} = [\mathbf{C}, \mathbf{1}], \quad \mathbf{y} = [x_1^2 + y_1^2, x_2^2 + y_2^2, \cdots, x_N^2 + y_N^2]^\mathsf{T},$$

где найденые оптимальные параметры линейной регрессии $\mathbf{w} = [w_1, w_2, w_3]^\mathsf{T}$ восстанавливают параметры окружности:

$$x_0 = \frac{w_1}{2}, \quad y_0 = \frac{w_2}{2}, \quad r = \sqrt{w_3 + x_0^2 + y_0^2}.$$

Задана выборка:

$$\mathbf{X} \in \mathbb{R}^{N \times n},$$

где N — число объектов в выборке, а n — размерность признакового пространства.

Definition

Смесь экспертов — мультимодель, определяющая правдоподобие веса π_k каждой локальной модели $\mathbf{f}_k$ на признаковом описании объекта $\mathbf{x}$.

$$\hat{\mathbf{f}} = \sum_{k=1}^{K} \pi_k \mathbf{f}_k, \qquad \pi_k\left(\mathbf{x}, \mathbf{V}\right) : \mathbb{R}^{n \times |\mathbf{V}|} \to [0, 1], \qquad \sum_{k=1}^{K} \pi_k\left(\mathbf{x}, \mathbf{V}\right) = 1,$$

где $\hat{\mathbf{f}}$ — мультимодель, а $\mathbf{f}_k$ является локальной моделью, π_k — шлюзовая функция, $\mathbf{w}_k$ — параметры k-й локальной модели, $\mathbf{V}$ — параметры шлюзовой функции.

В качестве локальных моделей $\mathbf{f}_k$ и шлюзовой функции $\boldsymbol{\pi}$ рассматриваются следующие функции:

$$\mathbf{f}_k\left(\mathbf{x}\right) = \mathbf{w}_k^{\mathsf{T}}\mathbf{x}, \quad \boldsymbol{\pi}\left(\mathbf{x}, \mathbf{V}\right) = \text{softmax}\Bigl(\mathbf{V}_1^{\mathsf{T}}\boldsymbol{\sigma}\bigl(\mathbf{V}_2^{\mathsf{T}}\mathbf{x}\bigr)\Bigr),$$

где $\mathbf{V} = \bigl\{\mathbf{V}_1, \mathbf{V}_2\bigr\}$ — параметры шлюзовой функции.

Параметры локальных моделей оптимизируются согласно принципу максимального правдоподобия модели:

$$p\bigl(\mathbf{y}, \mathbf{W}|\mathbf{X}, \mathbf{V}\bigr) = \prod_{k=1}^{K} p^k\left(\mathbf{w}_k\right) \prod_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k p_k\left(y_i|\mathbf{w}_k, \mathbf{x}_i\right)\right),$$

где $\mathbf{W} = \bigl[\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_K\bigr]^{\mathsf{T}}$.
Задача оптимизации параметров локальных моделей и параметров смеси:

$$\hat{\mathbf{W}}, \hat{\mathbf{V}} = \arg\max_{\mathbf{W}, \mathbf{V}} p\bigl(\mathbf{y}, \mathbf{W}|\mathbf{X}, \mathbf{V}\bigr).$$

Рассматривается вероятностная постановка задачи:

1) правдоподобие выборки $p_k\left(y_i|\mathbf{w}_k, \mathbf{x}_i\right) = \mathcal{N}\left(y_i|\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i, \beta^{-1}\right)$, где β уровень шума,
2) априорное распределение параметров $p^k\left(\mathbf{w}_k\right) = \mathcal{N}\left(\mathbf{w}_k|\mathbf{w}_k^0, \mathbf{A}_k\right)$, где $\mathbf{w}_k^0$ — вектор размера $n \times 1$, $\mathbf{A}_k$ — ковариационная матрица параметров,
3) регуляризация априорного распределения $p\left(\boldsymbol{\varepsilon}_{k,k'}|\boldsymbol{\alpha}\right) = \mathcal{N}\left(\boldsymbol{\varepsilon}_{k,k'}|\mathbf{0}, \boldsymbol{\Xi}\right)$, где $\boldsymbol{\Xi}$ — ковариационная матрица общего вида, $\boldsymbol{\varepsilon}_{k,k'} = \mathbf{w}_k^0 - \mathbf{w}_{k'}^0$.

Правдоподобие модели включает правдоподобие выборки, априорное распределение параметров, а также их регуляризацию

$$p\bigl(\mathbf{y}, \mathbf{W}|\mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \boldsymbol{\Xi}, \beta\bigr) = \prod_{k,k'=1}^{K} \mathcal{N}\left(\boldsymbol{\varepsilon}_{k,k'}|\mathbf{0}, \boldsymbol{\Xi}\right) \cdot$$
$$\cdot \prod_{k=1}^{K} \mathcal{N}\left(\mathbf{w}_k|\mathbf{w}_k^0, \mathbf{A}_k\right) \prod_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k \mathcal{N}\left(y_i|\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i, \beta^{-1}\right)\right),$$

где $\mathbf{A} = \bigl\{\mathbf{A}_1, \cdots, \mathbf{A}_K\bigr\}$. Введем скрытые переменные $\mathbf{Z} = [z_{ik}]$, где $z_{ik} = 1$ тогда и только тогда, когда $k_i = k$:

$$\log p\bigl(\mathbf{y}, \mathbf{Z}, \mathbf{W}|\mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \boldsymbol{\Xi}, \beta\bigr) =$$
$$= \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \left[\log \pi_k\left(\mathbf{x}_i, \mathbf{V}\right) - \frac{\beta}{2}\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right)^2 + \frac{1}{2}\log\frac{\beta}{2\pi}\right] +$$
$$+ \sum_{k=1}^{K}\left[-\frac{1}{2}\left(\mathbf{w}_k - \mathbf{w}_k^0\right)^{\mathsf{T}} \mathbf{A}_k^{-1}\left(\mathbf{w}_k - \mathbf{w}_k^0\right) + \frac{1}{2}\log\det \mathbf{A}_k^{-1} - \frac{n}{2}\log 2\pi\right] +$$
$$+ \sum_{k=1}^{K}\sum_{k'=1}^{K}\left[-\frac{1}{2}\left(\mathbf{w}_k^0 - \mathbf{w}_{k'}^0\right)^{\mathsf{T}} \hat{\boldsymbol{\alpha}}^{-1}\left(\mathbf{w}_k^0 - \mathbf{w}_{k'}^0\right) + \frac{1}{2}\log\det \boldsymbol{\Xi} - \frac{n}{2}\log 2\pi\right].$$

Задача оптимизации параметров локальных моделей и параметров смеси принимает следующий вид:

$$\mathbf{W}, \mathbf{Z}, \mathbf{V}, \mathbf{W}^0, \mathbf{A}, \beta = \arg\max_{\mathbf{W}, \mathbf{Z}, \mathbf{V}, \mathbf{W}^0, \mathbf{A}, \beta} \log p\big(\mathbf{y}, \mathbf{Z}, \mathbf{W} | \mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \mathbf{\Xi}, \beta\big).$$

Для оптимизации используется вариационный EM–алгоритм с аппроксимацией среднего поля $q(\mathbf{Z}, \mathbf{W}) = q(\mathbf{Z})\, q(\mathbf{W})$.
E и M шаги алгоритма имеют следующий вид:

1. E–шаг:

$$\log q(\mathbf{Z}^s) \propto \mathsf{E}_{q/\mathbf{Z}} \log p\big(\mathbf{y}, \mathbf{Z}, \mathbf{W} | \mathbf{X}, \mathbf{V}^{s-1}, \mathbf{A}^{s-1}, \mathbf{W}^{0,s-1}, \mathbf{\Xi}, \beta^{s-1}\big),$$
$$\log q(\mathbf{W}^s) \propto \mathsf{E}_{q/\mathbf{W}} \log p\big(\mathbf{y}, \mathbf{Z}, \mathbf{W} | \mathbf{X}, \mathbf{V}^{s-1}, \mathbf{A}^{s-1}, \mathbf{W}^{0,s-1}, \mathbf{\Xi}, \beta^{s-1}\big),$$

2. M–шаг:

$$\mathbf{W}^{0,s}, \mathbf{A}^s, \mathbf{V}^s, \beta^s = \arg\max_{\mathbf{W}^0, \mathbf{A}, \mathbf{V}, \beta} \mathsf{E}_{q^s} \log p\big(\mathbf{y}, \mathbf{Z}, \mathbf{W} | \mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \mathbf{\Xi}, \beta\big),$$

где s — номер итерации.

Итерационные формулы EM–алгоритма:

1. E–шаг:

$$p\left(z_{ik}=1\right)=\frac{\exp\left(\log\pi_k\left(\mathbf{x}_i,\mathbf{V}\right)-\frac{\beta}{2}\left(\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i-\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\right)\right)}{\sum_{k'=1}^{K}\exp\left(\log\pi_{k'}\left(\mathbf{x}_i,\mathbf{V}\right)-\frac{\beta}{2}\left(\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_{k'}\mathbf{w}_{k'}^{\mathsf{T}}\mathbf{x}_i-\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_{k'}\right)\right)},$$

$$q(\mathbf{w}_k)=\mathcal{N}(\mathbf{w}_k|\mathbf{m}_k,\mathbf{B}_k),$$

$$\mathbf{m}_k=\mathbf{B}_k\left(\mathbf{A}_k^{-1}\mathbf{w}_k^0+\beta\sum_{i=1}^{N}\mathbf{x}_iy_i\mathsf{E}z_{ik}\right),\quad \mathbf{B}_k=\left(\mathbf{A}_k^{-1}+\beta\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}z_{ik}\right)^{-1}.$$

2. M–шаг:

$$\mathbf{A}_k=\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}-\mathbf{w}_k^0\mathsf{E}\mathbf{w}_k^{\mathsf{T}}-\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{0\mathsf{T}}+\mathbf{w}_k^0\mathbf{w}_k^{0\mathsf{T}},$$

$$\frac{1}{\beta}=\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\left[y_i^2-2y_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k+\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]\mathsf{E}z_{ik},$$

$$\mathbf{w}_k^0=\left[\mathbf{A}_k^{-1}+(K-1)\,\boldsymbol{\Xi}\right]^{-1}\left(\mathbf{A}_k^{-1}\mathsf{E}\mathbf{w}_k+\boldsymbol{\Xi}\sum_{k'=1,\ k'\neq k}^{K}\mathbf{w}_{k'}^0\right),$$

$$\mathbf{V}=\arg\max_{\mathbf{V}}\mathsf{E}_{q^s}\log p\big(\mathbf{y},\mathbf{Z},\mathbf{W}|\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\boldsymbol{\Xi},\beta\big).$$

Вычислительный эксперимент делится на следующие этапы:

1. Анализ синтетических данных с разным типом шума в изображении;
2. Анализ изменения параметров локальных моделей во время обучения;
3. Анализ мультимоделей в зависимости от уровня шума в изображении;
4. Анализ качества модели на реальных данные.

Гиперпараметры заданы следующим образом:

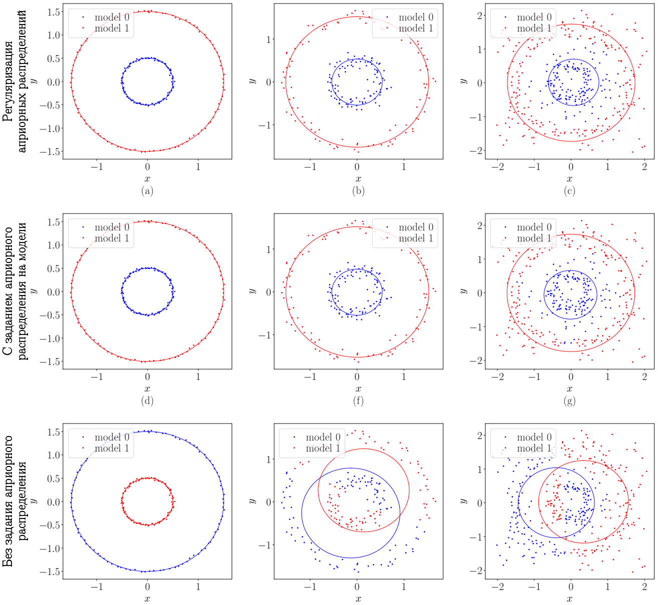
1. Априорные распределения на параметры локальных моделей в эксперименте задано следующим образом:

$$p^1\left(\mathbf{w}_1\right) \sim \mathcal{N}\left(\mathbf{w}_1^0, \mathbf{I}\right), \quad p^2\left(\mathbf{w}_2\right) \sim \mathcal{N}\left(\mathbf{w}_2^0, \mathbf{I}\right),$$

   где $\mathbf{w}_1^0 = [0, 0, 0.1]$, $\mathbf{w}_2^0 = [0, 0, 2]$.
2. Параметр регуляризации:

$$\boldsymbol{\Xi} = \begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

   что указывает на концентричность окружностей.

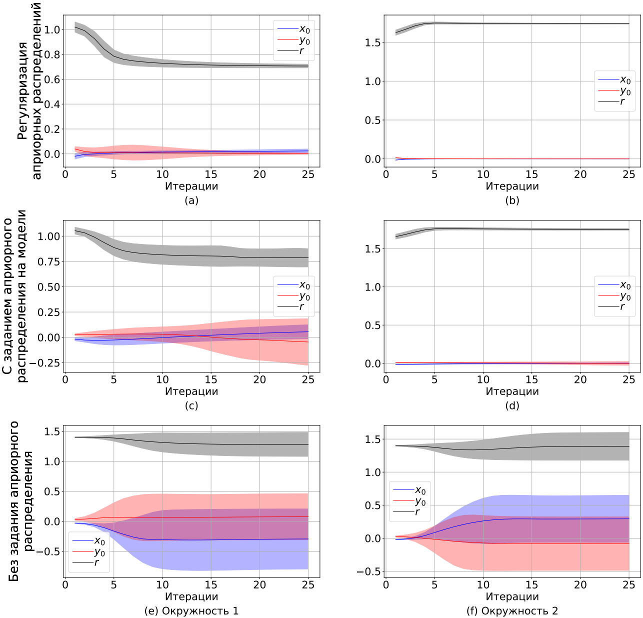


(a) (b) (c)

(d) (f) (g)

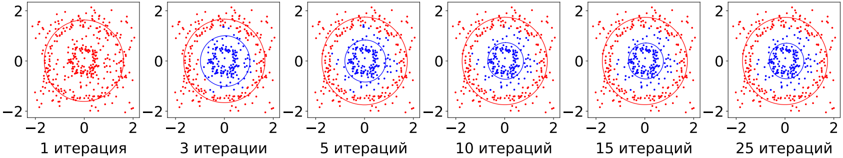
(e) Без шума (h) Шум вблизи окружности (j) Равномерный шум

Параметры локальных моделей в процессе обучения

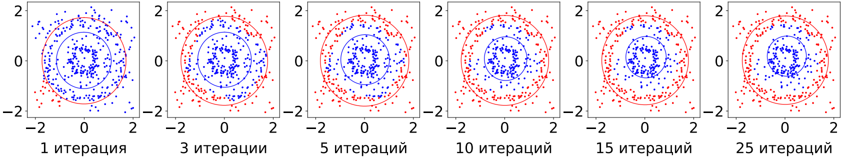


(a) (b) (c) (d) (e) Окружность 1 (f) Окружность 2

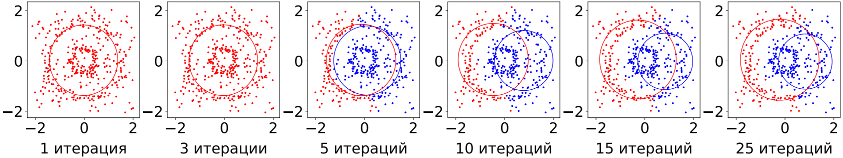
Регуляризация априорных распределений

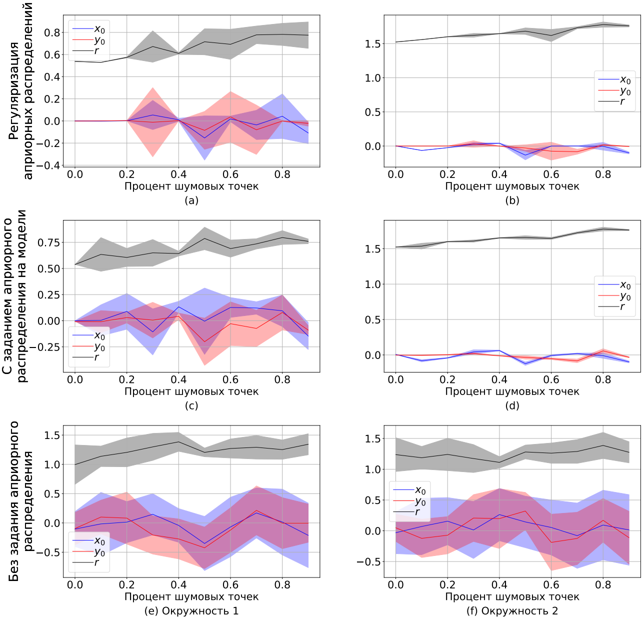


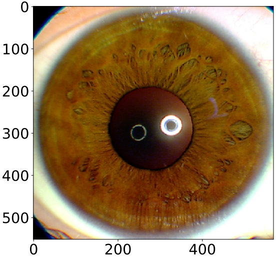
С заданием априорного распределения на модели



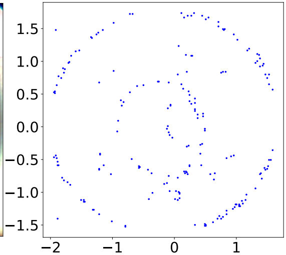
Без задания априорного распределения



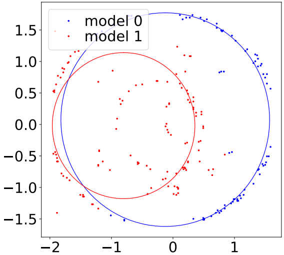

Регуляризация априорных распределений
С заданием априорного распределения на модели
Без задания априорного распределения
x_0
y_0
r
Процент шумовых точек
(a)
(b)
(c)
(d)
(e) Окружность 1
(f) Окружность 2



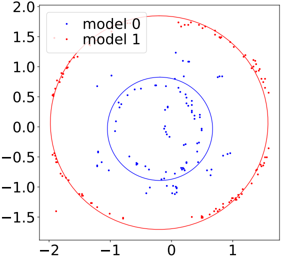
(a) Исходное изображение



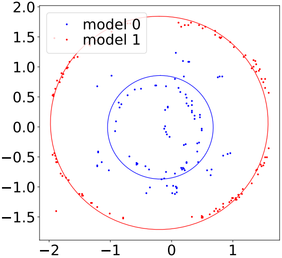
(b) Бинаризованное изображение



(c) Без задания априорного распределения



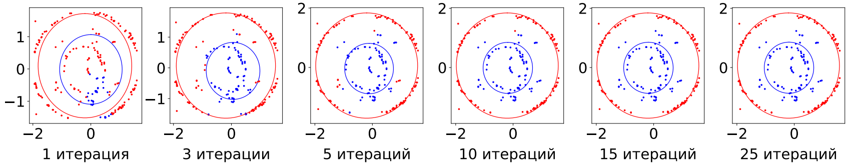
(d) С заданием априорного распределения на модели



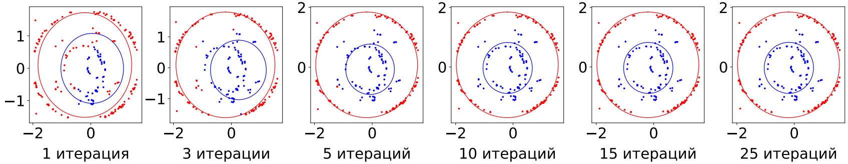
(e) Регуляризация априорных распределений

Обучения на реальных данных

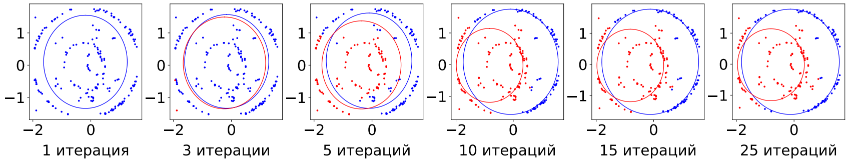
Регуляризация априорных распределений



С заданием априорного распределения на модели



Без задания априорного распределения



Сделано:

1. Предложен метод поиска окружностей на бинарном изображении с различными априорными предположениями.
2. Введено понятие регуляризации априорных распределений для улучшения качества мультимодели.
3. В эксперименте показано, что задания регуляризации позволяет улучшить качество и устойчивость модели.

Планируется:

1. Улучшить мультимодель при помощи задания априорного распределения на шлюзовую функцию.
2. Рассмотреть в качестве локальных моделей не только модели, которые описывают данные, а также модель, которая отвечает за шум в данных.
3. Расширить класс локальных моделей с окружности до произвольной кривой второго порядка.