Noname manuscript No.
(will be inserted by the editor)

Analysis of the properties of probabilistic models in learning problems with an expert

Andrey V. Grabovoy · Alexandra I. Bazarova · Vadim V. Strijov

Received: date / Accepted: date

Abstract This paper is devoted to the construction of interpretable models in machine learning. It solves the problem of approximating a given set of figures on a contour image. Assumptions are introduced that the figures are curves of the second order. When approximating figures, information about the type, location and form of curves, as well as the set of their possible transformations, is used. Such information is called *expert*, and the machine learning method based on expert information is called *expert learning*. It is assumed that the set of shapes is approximated by a set of *local models*. Each local model, based on expert information, approximates one shape in the contour image. To build models, it is proposed to map second-order curves into a feature space where each local model is a linear model. Thus, higher-order curves are recognized using the composition of linear models. As an applied problem, the problem of approximating the iris of the eye on a contour image is considered.

Keywords Mixture of experts · Expert learning · Linear models · Interpretabe models

The reported study was funded by Russian Foundation for Basic Research according to the research projects 20-37-90050, 19-07-01155 and National Technological Initiative project 13/1251/2018.

Andrey V. Grabovoy
Moscow Institute of Physics and Technology
9 Institutskiy per., Dolgoprudny, Moscow Region, 141701, Russian Federation
E-mail: grabovoy.av@phystech.edu

Alexandra I. Bazarova
Moscow Institute of Physics and Technology
9 Institutskiy per., Dolgoprudny, Moscow Region, 141701, Russian Federation
E-mail: bazarova.ai@phystech.edu

Vadim V. Strijov
Moscow Institute of Physics and Technology
9 Institutskiy per., Dolgoprudny, Moscow Region, 141701, Russian Federation
E-mail: strijov@phystech.edu

1 Introduction

Interpretable model building in machine learning (Ribeiro et al. 2016) is one of the key challenges. Modern solutions of the image classification problem based on deep learning networks ResNet, VGG, Intercept (Kaiming et al. 2016) are poorly interpreted models. The papers (Han et al. 2020; Akhtar et al. 2018) show that deep learning networks are sensitive even to small noise in the data, which is due to their uninterpretability.

In this paper, we propose a *training with an expert* method. This method assumes the use of subject knowledge of experts to improve the quality of approximation, as well as to obtain interpretable machine learning models. The subject knowledge of experts about the sample will be called *expert information*. It is assumed that the use of expert information allows the sample to be approximated by simple interpretable models, such as linear models. Machine learning methods that take expert knowledge into account when building models are called *expert learning*.

This paper solves the problem of approximating second-order curves on a contour image. Second-order curves are selected for analysis, since they are easily described by linear models. In this case, these figures need to be restored in such applied problems as the problem of recognizing the iris of the eye (Matveev 2010; Matveev et al. 2014; Bowyer et al. 2010), the problem of describing the particle track in the hadron collider (Salamani et al. 2018). Expert information about a second-order curve allows you to map points on a plane into a new feature description, where each curve is approximated by one linear model. A model that approximates one curve is called a *local model*. To approximate the entire contour image, it is required to approximate several second-order curves using several local models. In this paper, the following restrictions on images are introduced: a) the image consists only of second-order curves; b) the image is approximated by a small number of second-order curves; c) the number and type of curves in the image is known.

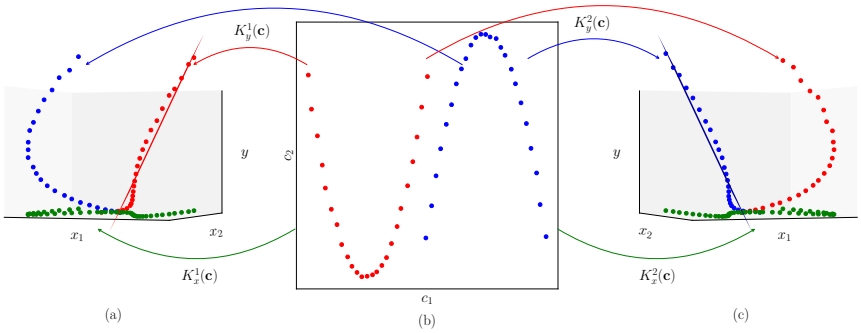
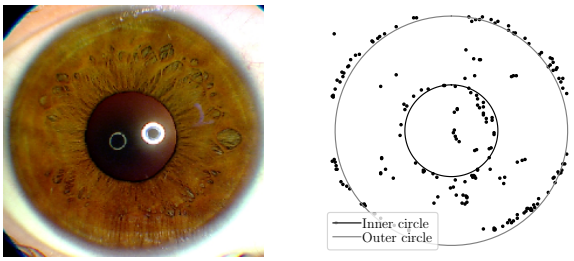


Fig. 1: Example: a) expert information of the first expert; b) baseline data; c) expert information of the second expert

Figure 1 shows an example of second-order curves, as well as expert information on curves. Figure 1 a shows the expert information of the first expert. Using this information, the first curve is fitted with a linear model and the second curve is noise. Figure 1 b shows the expert information of the second expert. Using this information, the second curve is fitted with a linear model and the first curve is noise.

When approximating several curves on one contour image, a multi-model is built. An example of multi-models is a random forest (Chen et al. 2012), tree boosting (Chen et al. 2016), a mixture of experts (Yuksel et al. 2012). In this paper, a mixture of experts is considered as a multi-model. Expert mixture is a multi-model that linearly weights local models that approximate a portion of the sample. The values of the weighting coefficients depend on the object for which the prediction is made. To solve the problem of a mixture of experts, a variational EM-algorithm (Dempster et al. 1997; Bishop 2010; Peng et al. 1996) is used. The mixture of experts has many uses in a number of applications. In the paper (Estabrooks et al. 2001), the text classification problem is solved. In the papers (Cheung et al. 1995; Weigend et al. 2000; Cao 2003; Mossavat et al. 2010; Sminchisescu C et al. 2007; Tuerk 2001; Yumlu et al. 2003), a mixture of experts is used to predict time series for speech recognition, daily human activity, and prediction of the value of securities. In the paper (Ebrahimpour et al. 2009), a mixture of experts was considered to solve the problem of recognizing handwritten numbers in images.

As an example, the problem of approximation of the iris image is considered. Figure 2a shows an example of the image that needs to be approximated. In this paper, we consider a processed image, which is given in outline form, an example of such an image is shown in Figure 2b. Figure 2b shows two local circle models that approximate the iris of the eye. Circumferences are a simple example of a second order curve.



(a) (b)

Fig. 2: An example of the image of the iris of the eye and its outline representation: a) the image of the iris of the eye; b) contour image of the iris and approximating the given image of the circumferences

For the problem of approximating the iris of the eye, the following expert information is used: the iris of the eye is approximated by two concentric circumferences. Expert information is used to construct a feature description of plane points, as well as to build an optimization function. The part of the error function for optimization that uses expert information is called a regularizer. Thus, the information that the image of the circumferences is specified by the feature description, and the information that the concentric circumferences are specified using a special regularizer.

In a computational experiment, the quality of the approximation of the contour image is analyzed depending on the specified expert information and on the noise level in the synthetically generated data. The analysis of the quality of the approximation of the iris is carried out, depending on the amount of expert information that was used to build the model. Note that each approximated image is a separate set of points that need to be approximated.

2 Statement of the problem of finding the parameters of the second-order curves in the image

Binary image is set:

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

where 1 corresponds to the black point of the image, and 0 corresponds to the white point of the background. From the image $\mathbf{M}$, a sample$\mathbf{C}$ is constructed, the elements of which are the coordinates (x_i, y_i) of black points:

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

The expert assumes that the image consists of a second-order curve Ω. Let for a set of points $\mathbf{C} \in \mathbb{R}^{N \times 2}$ that form a curve Ω, expert information about the figure $E(\Omega)$is given. The set $E(\Omega)$ consists of the shape Ω expected by the expert and the set of its admissible transformations. Based on the expert description, let us introduce mappings into a new problem for approximation:

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

where K_x mapping objects to the attribute description of objects, $n$is the number of features, and K_y is a mapping to a target variable for an object. Applying the mappings $K_x, K_y$for the sample $\mathbf{C}$ element by element we obtain:

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

where $\mathbf{c} = (x_i, y_i)$ is a sample point $\mathbf{C}$.

Applying the mappings (2) to the original set of points $\mathbf{C}$, we obtain the sample

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

We get that the original problem of curve approximation Ω is reduced to approximation of the sample $\mathfrak{D}$. In this paper, it is assumed that the sample $\mathfrak{D}$ is approximated by a linear model:

$$g(\mathbf{x}, \mathbf{w}) = \mathbf{x}^{\mathsf{T}}\mathbf{w}, \tag{4}$$

where $\mathbf{w}$ vector, the parameter to be found.

To find the optimal vector of parameters $\hat{\mathbf{w}}$, it is required to solve the following optimization problem:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathbb{R}^n} \sum_{(\mathbf{x},y)\in\mathfrak{D}} \|g(\mathbf{x}, \mathbf{w}) - y\|_2^2. \tag{5}$$

Thus, the problem of approximating the original curve Ω is reduced to solving the problem of linear regression, i.e. finding the components of the vector $\hat{\mathbf{w}}$ connecting the resulting $\mathbf{x}$ and y.

In the case when on the image K the second-order curves $\Omega_1, \ldots, \Omega_K$, for each of which there is expert information $E_k = E(\Omega_k)$, $k \in \{1, \ldots, K\}$, the problem of constructing a multi-model called a mixture of K experts is posed.

Definition 1 We call the multimodel f a mixture of K experts

$$f = \sum_{k=1}^{K} \pi_k(\mathbf{x}, \mathbf{V}) g_k(\mathbf{w}_k), \quad \pi_k(\mathbf{x}, \mathbf{V}) : \mathbb{R}^{n\times|\mathbf{V}|} \to [0,\, 1], \quad \sum_{k=1}^{K} \pi_k(\mathbf{x}, \mathbf{V}) = 1, \tag{6}$$

where g_k is a local model called by the expert, $\mathbf{x}$is an attribute description of an object, $\pi_k$is a gateway function, the vector $\mathbf{w}_k$ are local model parameters, the vector $\mathbf{V}$ are gateway function parameters. In this paper, g_k is a linear model.

For each second-order curve, mappings (1) are given. For convenience, we introduce the following notation: $K_x^k(\mathbf{c}) = K_x(\Omega_k, \mathbf{c})$ and $K_y^k(\mathbf{c}) = K_y(\Omega_k, \mathbf{c})$. Then, using local linear models, we construct a universal multi-model describing the curves $\Omega_1, \ldots, \Omega_K$ on the image $\mathbf{M}$:

$$f = \sum_{\mathbf{c}\in\mathbf{C}} \sum_{k=1}^{K} \pi_k(\mathbf{c}, \mathbf{V}) g_k(K_x^k(\mathbf{c}), \mathbf{w}_k), \tag{7}$$

where π_k is the gateway function. In this paper, we consider a simple case, where $\mathbf{x} = K_x^1(\mathbf{c}) = \cdots = K_x^K(\mathbf{c})$, then the expression (7) is rewritten in the following simple form:

$$f = \sum_{\mathbf{c}\in\mathbf{C}} \sum_{k=1}^{K} \pi_k(\mathbf{x}, \mathbf{V}) g_k(\mathbf{x}, \mathbf{w}_k), \tag{8}$$

where the gateway function π_k has the following form:

$$\pi_k(\mathbf{x}, \mathbf{V}) : \mathbb{R}^{n\times|\mathbf{V}|} \to [0,\, 1], \quad \sum_{k=1}^{K} \pi_k(\mathbf{x}, \mathbf{V}) = 1, \tag{9}$$

where $\mathbf{V}$ are the gateway function parameters, and g_k is a local model.

In this paper

$$\boldsymbol{\pi}(\mathbf{x}, \mathbf{V}) = \text{softmax}\big(\mathbf{V}_1^\mathsf{T}\boldsymbol{\sigma}(\mathbf{V}_2^\mathsf{T}\mathbf{x})\big), \tag{10}$$

where $\mathbf{V} = \{\mathbf{V}_1, \mathbf{V}_2\}$ are the gateway function parameters, $\mathbf{V}_1 \in \mathbb{R}^{p\times k}$, $\mathbf{V}_2 \in \mathbb{R}^{n\times p}$.

To find the optimal parameters of the multi-model, it is necessary to solve the following optimization problem:

$$\mathcal{L} = \sum_{(\mathbf{x},y)\in\mathfrak{D}}\sum_{k=1}^{K}\pi_k(\mathbf{x}, \mathbf{V})(y - \mathbf{w}_k^\mathsf{T}\mathbf{x})^2 + R\big(\mathbf{V}, \mathbf{W}, E(\Omega)\big) \to \min_{\mathbf{V},\mathbf{W}}, \tag{11}$$

where $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_k]$ are parameters of local models, $R\big(\mathbf{V}, \mathbf{W}, E(\Omega)\big)$ is regularization parameters, based on expert information.

3 Building an attribute description of figures

Unified space for second-order curves. An arbitrary second-order curve, the main axis of which is not parallel to the ordinate axis, is given by the following expression:

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

where the coefficients B', C' are subject to restrictions that depend on the type of the curve. The expression (2) takes the following form:

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

whence we obtain the linear regression problem for recovering the parameters B', C', D', E', F' from the selected sample.

Circumference. As a special case of a second-order curve, we consider the circumference. Let (x_0, y_0) be the center of the circumference to be found on the binary image $\mathbf{M}$, and r be its radius. The sample elements $(x_i, y_i) \in \mathbf{C}$ are the locus of points, which is approximated by the equation of the circumference:

$$(x_i - x_0)^2 + (y_i - y_0)^2 = r^2. \tag{12}$$

Expanding the brackets, we get:

$$(2x_0)\cdot x_i + (2y_0)\cdot y_i + (r^2 - x_0^2 - y_0^2)\cdot 1 = x_i^2 + y_i^2. \tag{13}$$

Then the presentations (2) take the following form:

$$K_x(\mathbf{c}_i) = [x_i,\, y_i,\, 1] = \mathbf{x},\; K_y(\mathbf{c}_i) = x_i^2 + y_i^2 = y. \tag{14}$$

Assign the linear regression problem (3). Vector components $\mathbf{w} = [w_0,\, w_1,\, w_2]^\mathsf{T}$, binding $\mathbf{x}$ and y, restore the parameters of the circumference:

$$x_0 = \frac{w_0}{2},\; y_0 = \frac{w_1}{2},\; r = \sqrt{w_3 + x_0^2 + y_0^2}. \tag{15}$$

4 Composition of figures

o construct a composition of figures, we will use the expression (11), which takes the following form:

$$\mathcal{L} = \sum_{\mathbf{c}\in\mathbf{C}} \sum_{k=1}^{K} \pi_k(\mathbf{c}, \mathbf{V}) \left(K_y^k(\mathbf{c}) - \mathbf{w}_k^{\mathsf{T}} K_x^k(\mathbf{c})\right)^2 + R\big(\mathbf{V}, \mathbf{W}, E(\Omega)\big) \to \min_{\mathbf{V},\mathbf{W}}, \quad (16)$$

where K_x^k, K_y^k expert representation of the k-th expert. Assuming that all curves in the image are described by one attribute description $\mathbf{x} = K_x^1(\mathbf{c}) = \cdots = K_x^K(\mathbf{c}), x = K_y^1(\mathbf{c}) = \cdots = K_y^K(\mathbf{c})$, we get the following optimization problem:

$$\mathcal{L} = \sum_{(\mathbf{x},y)\in\mathfrak{D}} \sum_{k=1}^{K} \pi_k(\mathbf{x}, \mathbf{V}) \left(y - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}\right)^2 + R\big(\mathbf{V}, \mathbf{W}, E(\Omega)\big) \to \min_{\mathbf{V},\mathbf{W}}, \quad (17)$$

As a regularizer R, additional restrictions on the vectors of model parameters are considered. To solve the optimization problem (17) it is proposed to use the EM-algorithm.

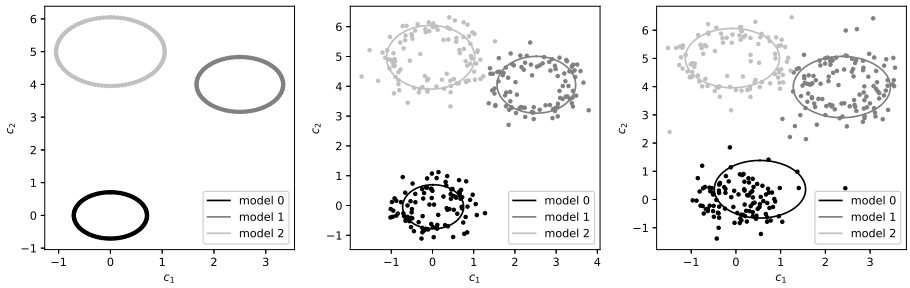
5 Computing experiment

A computational experiment was carried out to analyze the quality of models of second-order curves in the image. The experiment is divided into several parts. The first part is an experiment with several circumferences in the image. The second part analyzes the convergence of the method depending on the noise level in the data and on the specified expert information. In the third part, an experiment is conducted to approximate the iris of the eye.

5.1 Experiment with circumstances

In this part of the experiment, an example of training a multi-model is shown to approximate several second-order figures simultaneously. A synthetic sample is used as data, which is obtained by generating three arbitrary non-intersecting circumferences, as well as adding noise to these circumferences. Noise was added to the radius of the circle for each point, and random points were added to the sample that do not belong to circumstances.

Figure 3 shows the result of building an ensemble of locally approximating models that approximate the sample. Each local model approximates one circumference, and when adding different noise, the quality of the approximation will drop. Figure 4 shows a graph of the dependence of the radius of the circumferences r and their centers (x_0, y_0) on the iteration number.



(a) (b) (c)

Fig. 3: Multi-model depending on different prior assumptions and noise level. From left to right: circumferences without noise; noise in the radius of the circle; noise in the radius of a circle as well as arbitrary points throughout the image.

5.2 Experiment with different noise levels and variance of the prior distribution

In this part of the experiment, we analyze the quality of approximation S on the noise level β in the data and on the parameter of a priori distributions γ. The sample is obtained as followsfirst, two vectors of parameters are randomly selected $\mathbf{w}_1^{\text{true}}$ and $\mathbf{w}_2^{\text{true}}$ are coefficients of two parabolas. The vectors $\mathbf{w}_1^{\text{true}}$ and $\mathbf{w}_2^{\text{true}}$ are used to generate points x_i and y_i with normal noise added $\varepsilon \sim \mathcal{N}\big(0, \beta\big)$. When training a multi-model, the prior distribution of parameters is considered $\mathbf{w}_1 \sim \mathcal{N}\big(\mathbf{w}_1^{\text{true}}, \gamma\mathbf{I}\big), \mathbf{w}_2 \sim \mathcal{N}\big(\mathbf{w}_2^{\text{true}}, \gamma\mathbf{I}\big)$.

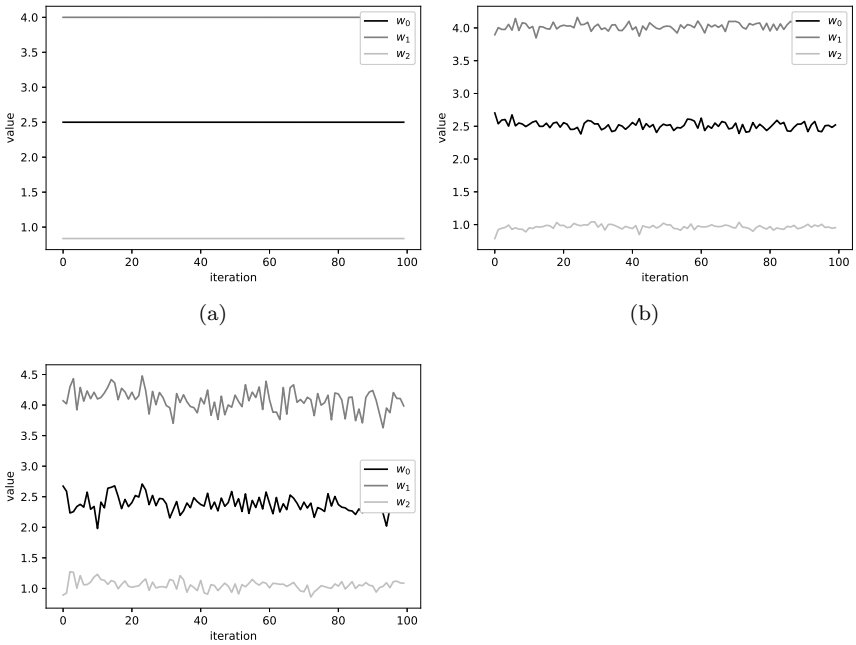
The following quality criterion is considered:

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

where $\mathbf{w}_1^{\text{pred}}$ approximation of the vector of parameters of the first local model, and $\mathbf{w}_2^{\text{pred}}$ approximation of the vector of parameters of the second local model.

Figure 6 shows the dependence of the quality criterion S on the noise level β and the a priori distribution parameter γ. The graph shows that at a low noise level β the quality of the approximation does not depend on the parameter γ, and with an increase in the noise β the quality of the approximation S decreases.

Figure 6 shows an example of how the algorithm works with different parameters β and γ. It is seen that in the absence of noise β, both local models approximate the sample. With an increase in the noise level, the quality of the approximation decreases: at $\beta = 0{,}2$, with an increase in γ, the first local model from a parabola goes over to an ellipse; for $\beta = 0{,}4$ as γ increases, the first local model from a parabola goes over to an ellipse, and the second model from a parabola goes over to a hyperbola.



(a) (b)

(c)

Fig. 4: Dependence of the parameters r, x_0 and y_0 on the iteration number for different prior distributions. From left to right: circumferences without noise; noise in the radius of the circle; noise in the radius of a circle as well as arbitrary points throughout the image.

5.3 Iris approximations

An analysis of the quality of the approximation is carried out for the problem of approximating the iris of the eye in the image. The iris of the eye consists of two concentric circumferences, therefore, a multi-model is considered, which consists of two experts: each expert approximates one of the circumstances. In a computational experiment, the quality of the approximation of circumferences is compared in the case of specifying different regularizers R_0, R_1, R_2. Regularizer$R_0\big(\mathbf{V}, \mathbf{W}, E(\Omega)\big) = 0$, that is, there is no regularizer. Regularizer:

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

which romotes near-zero parameters of local models. Regularizer

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

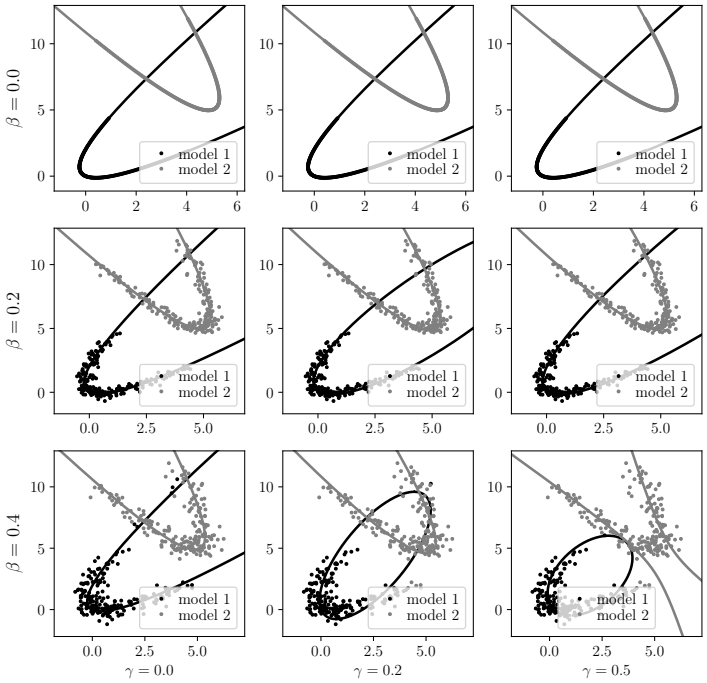


Fig. 5: The result of the approximation for data with different noise levels β and on the variance of the prior distribution γ

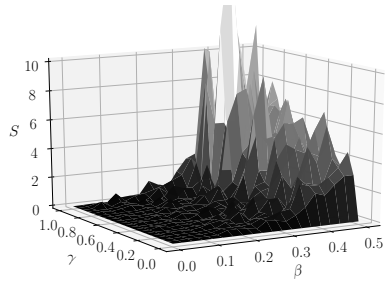


Fig. 6: Dependence of models on the noise level β in the data, as well as on the variance of the prior distribution γ

which promotes the coincidence of the centers of the circumferences and close to zero parameters of the model. Figure 7 shows the result of the eye iris approximation algorithm after 10 iterations. It can be seen that in the absence of a regularizer, one of the circumferences is found incorrectly. If the regularizer R_1 is given, the model approximates both circumferences with good

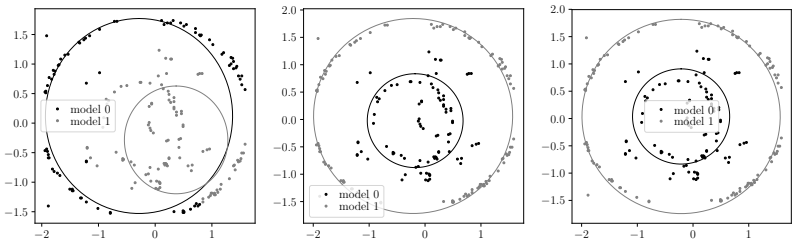


Fig. 7: Visualization of the approximation of the iris: a) if the R_0 regularizer is specified; b) if the R_1 regularizer is specified; b) if the R_2 regularizer is specified

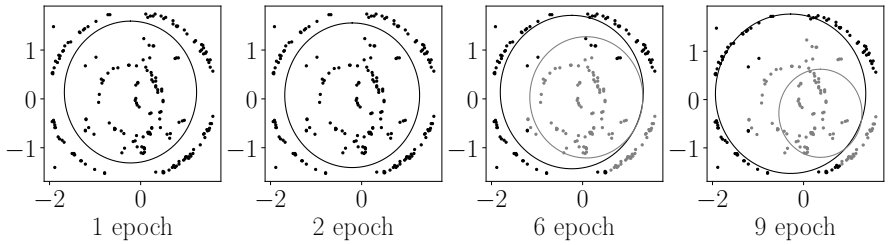


Fig. 8: Visualization of the multi-model convergence process in the case of a regularizer R_0

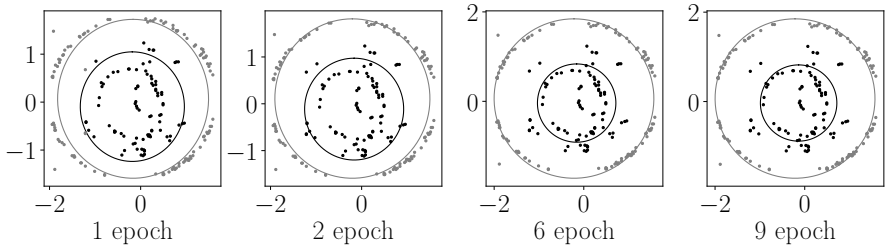


Fig. 9: Visualization of the multi-model convergence process in the case of a regularizer R_1

quality, but the circumferences are not concentric. In case of specifying the regularizer R_2, we get concentric circumferences on the image.

Figure 8–10 shows the process of convergence of multi-models in the case of specifying different regularizers R_0, R_1, R_2. It can be seen that the models with the regularizer type R_1 and R_2 approximate both circumferences, and the multi-model with theR_0 regularizer approximates only the large circumference.

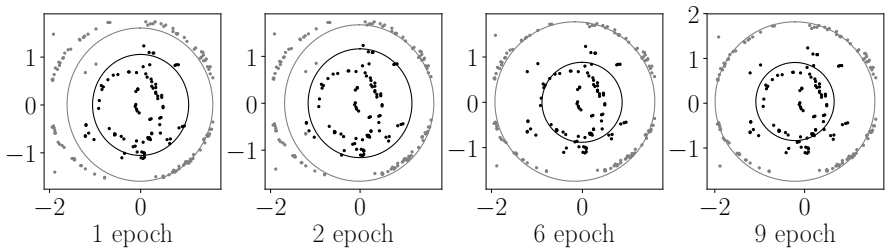


Fig. 10: Visualization of the multi-model convergence process in the case of a regularizer R_2

6 Conclusion

The paper proposes a method for constructing interpretable machine learning models based on expert information. The problem of approximation of the second-order curves: parabola, hyperbola, ellipse is considered as a problem. Approximation of curves of the second order is applied in the problem of approximation of the iris of the eye.

An experiment was carried out during which the quality of the approximation of the second-order curves is analyzed depending on the initial noise level in the data, as well as depending on the regularizer. During the experiment, it was shown that with an increase in the noise level in the initial data, the approximation accuracy decreases: with a large noise, the shape of the approximated figure changes from a parabola to a hyperbola. A computational experiment was carried out to approximate the iris of the eye using two concentric circles. The experiment shows that regularization based on expert information improves the quality of the approximation.

Acknowledgements The reported study was funded by Russian Foundation for Basic Research according to the research projects 20-37-90050, 19-07-01155 and National Technological Initiative project 13/1251/2018.

References

1. Akhtar N, Mian A (2018) Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey. IEEE Access 6:14410–14430
2. Bishop C. (2010) Pattern Recognition and Machine Learning. Springer, Berlin
3. Bowyer K, Hollingsworth K, Flynn P (2010) A Survey of Iris Biometrics Research: 2008-2010. Handbook of iris recognition 15–54
4. Cao L (2003) Support vector machines experts for time series forecasting. Neurocomputing 51:321–339
5. Chen T, Guestrin C (2016) XGBoost: A Scalable Tree Boosting System. KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 785–794
6. Chen Xi, Ishwaran H (2012) Random Forests for Genomic Data Analysis. Genomics 6:323–329

7. Cheung Y, Leung W, Xu L (1995) Application of mixture of experts model to financial time series forecasting. In Proc. Int. Conf. Neural Netw. Signal Process. 1–4
8. Dempster A, Laird N, Rubin D (1997) Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39:1–38
9. Estabrooks A, Japkowicz N (2001) A mixture-of-experts framework for text classification. In Proc. Workshop Comput. Natural Lang. Learn., Assoc. Comput. Linguist. 1–8
10. Ebrahimpour R, Moradian R, Esmkhani A, Jafarlou F (2009) Recognition of Persian handwritten digits using characterization loci and mixture of experts. J. Digital Content Technol. Appl. 42–46
11. Han X, Yao M, Debayan D, Hui L, Ji-Liang T, Anil J (2020) Adversarial Attacks and Defenses in Images, Graphs and Text: A Review. International Journal of Automation and Computing 17:151–178
12. Kaiming H, Xiangyu Z, Shaoqing R, Jian S (2016) Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 770–778
13. Matveev I (2010) Detection of iris in image by interrelated maxima of brightness gradient projections. Appl. Comput. Math 9:252–257
14. Matveev I, Simonenko I (2014) Detecting precise iris boundaries by circular shortest path method. Pattern Recognition and Image Analysis 24:304–309
15. Mossavat S, Amft O, Vries B, Petkov P, Kleijn W (2010) A Bayesian hierarchical mixture of experts approach to estimate speech quality. In Proc. 2nd Int. Workshop Qual. Multimedia Exper. 200–205
16. Peng F, Jacobs R, Tanner M (1996) Bayesian inference in mixtures-of-experts and hierarchical mixtures-of-experts models with an application to speech recognition. J. Amer. Stat. Assoc. 91:953–960
17. Ribeiro M, Singh S, Guestrin C (2016) Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 1135–1144
18. Salamani D, Gadatsch S, Golling T, Stewart G, Ghosh A, Rousseau D, Hasib A, Schaarschmidt J (2018) Deep Generative Models for Fast Shower Simulation in ATLAS. 2018 IEEE 14th International Conference on e-Science (e-Science) https://doi.org/10.1109/eScience.2018.00091
19. Sminchisescu C, Kanaujia A, Metaxas D (2007) Discriminative density propagation for visual tracking. IEEE Trans. Pattern Anal. Mach. Intell. 29:2030–2044
20. Tuerk A (2001) The state based mixture of experts HMM with applications to the recognition of spontaneous speech. Ph.D. thesis, University of Cambridge.
21. Weigend A, Shi S (2000) Predicting daily probability distributions of S&P500 returns. J. Forecast 19:375–392
22. Yuksel S, Wilson J, Gader P (2012) Twenty Years of Mixture of Experts. IEEE Transactions on Neural Networks and Learning Systems 8:1177–1193
23. Yumlu M, Gurgen F, Okay N (2003) Financial time series prediction using mixture of experts. In Proc. 18th Int. Symp. Comput. Inf. Sci. 553–560