Серия «Математика»
2023. Т. 43. С. 110—121

Онлайн-доступ к журналу:
http://mathizv.isu.ru

ИЗВЕСТИЯ
Иркутского
государственного
университета

Research article

УДК 519.246
MSC 68T10, 62H30
DOI https://doi.org/10.26516/1997-7670.2023.43.110

On the Properties of Bias-Variance Decomposition for kNN Regression

Victor M. Nedel'ko[1]✉

[1] Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russian Federation
✉ nedelko@math.nsc.ru

Abstract. When choosing the optimal complexity of the method for constructing decision functions, an important tool is the decomposition of the quality criterion into bias and variance.

It is generally assumed (and in practice this is most often true) that with increasing complexity of the method, the bias component monotonically decreases, and the variance component increases. The conducted research shows that in some cases this behavior is violated.

In this paper, we obtain an expression for the variance component for the kNN method for the linear regression problem in the formulation when the "explanatory" features are random variables. In contrast to the well-known result obtained for non-random "explanatory" variables, in the considered case, the variance may increase with the growth of k.

Keywords: bias-variance decomposition, machine learning, k-nearest neighbors algorithm, overfitting

Acknowledgements: The study was carried out within the framework of the state contract of the Sobolev Institute of Mathematics (project no FWNF-2022-0015).

For citation: Nedel'ko V. M. On the Properties of Bias-Variance Decomposition for kNN Regression. *The Bulletin of Irkutsk State University. Series Mathematics*, 2023, vol. 43, pp. 110–121.
https://doi.org/10.26516/1997-7670.2023.43.110

Научная статья

О свойствах разложения функции потерь на смещение и разброс для метода kNN

В. М. Неделько[1]✉

[1] Институт математики им. С. Л. Соболева СО РАН, Новосибирск, Российская Федерация,
✉ nedelko@math.nsc.ru

Аннотация. Для метода ближайших соседей (kNN) существует общеизвестное аналитическое выражение для разложения ошибки регрессионной модели на смещение и разброс. Однако данное выражение справедливо только для классической постановки задачи регрессионного анализа, в которой случайной является только целевая переменная, а «объясняющие» переменные неслучайны. Получены аналитические выражения для разложения для некоторых постановок, когда все переменные являются случайными. В отличие от классической постановки в полученных выражениях компонента разброса демонстрирует различное поведение при разной размерности пространства, в частности, при размерности 1 разброс практически линейно увеличивается с ростом k, т.е. уменьшается с ростом сложности. Подобное поведение разложения на смещение и разброс является нежелательным его свойством при использовании для объяснения структуры ошибок обучения. В связи с этим представляется целесообразным использовать разложение ошибки на погрешность аппроксимации и статистическую погрешность. Компоненты последнего разложения всегда монотонны.

Ключевые слова: разложение на смещение и разброс, машинное обучение, метод k-ближайших соседей, проблема переобучения

Благодарности: Работа выполнена в рамках госзадания Института математики им. С.Л. Соболева СО РАН, проект FWNF-2022-0015.

Ссылка для цитирования: Nedel'ko V. M. On the Properties of Bias-Variance Decomposition for kNN Regression // Известия Иркутского государственного университета. Серия Математика. 2023. Т. 43. С. 110–121.
https://doi.org/10.26516/1997-7670.2023.43.110

1. Introduction

When analyzing the effectiveness of methods for constructing decision functions [18], the error decomposition into bias and variance [19] is widely used.

In the classical version of the decomposition, the variance is understood as the average variance of the predicted values, and the bias is called the average square of the difference between the average forecast and the optimal forecast. Averaging is performed over the space of variable values and over samples (from which the decision function is constructed).

This decomposition is widely used to explain the characteristic shape of the learning curve (the dependence of the quality [17] of the solution on the complexity of the method [8]), and for justification [13; 14] of ensemble methods [3; 6; 7], such as a random forest. In most of the examples [5] the dependence of the decomposition components on complexity has a

characteristic qualitative form: the bias decreases monotonically, the variance increases monotonically, the sum of these values has a characteristic minimum. However, it is noted that in practice, the bias with a sufficiently high complexity can also begin to grow [1], and the variance can decrease [2; 11; 12; 20].

In [15] some formal statements are proved that with increasing complexity, the bias can grow, and the variance decrease.

In this paper, we investigate the behavior of the error decomposition for the k nearest neighbors (kNN) method. This method has a relatively low relevance in solving applied problems (although it is often used as an ensemble component), however, it is one of the rare examples when an analytical expression is obtained for the error decomposition.

At first glance, it may seem that this issue has been fully investigated. Indeed, even on Wikipedia there is an analytical expression [4] for the bias-variance decomposition for kNN.

However, this expression is obtained for the "classical" formulation of the regression problem, when the "explanatory" variables are not random. But in real problems, the available data can most often be adequately interpreted as a random sample. In this case, all variables are random, not just the target one.

In the case of random "explanatory" variables, the expression remains valid, but it ceases to be a decomposition into bias and variance, since one of the terms ceases to be a constant and contributes to both the bias and the variance.

In this paper, we will obtain an analytical expression for decomposition of the error for the kNN method for the case of one-dimensional regression, when all variables are random variables. It turns out that in this formulation, the qualitative behavior of the variance component turns out to be fundamentally different.

2. Problem statement

2.1. Regression problem

Let X be the space of values of variables used for forecasting, and Y be the set of values of the predicted variable.

All variables are random variables with some joint distribution function.

Decision function is a mapping $f : X \to Y$.

The decision function is constructed based on some training sample of size N

$$S_N = \left((x^\omega, y^\omega), \omega = \overline{1, N}\right).$$

The quality of the decision made [16] is evaluated by a given loss function: $L : Y \times Y \to [0, \infty)$. In this paper, we will consider the square of the deviation as a loss function, i.e. $L(y, f(x)) = (y - f(x))^2$.

For the decision function as a whole, the quality criterion will be MSE, i.e.

$$R(f(\cdot)) = \mathsf{E}_{x,y}(y - f(x))^2.$$

By this criterion, the optimal solution will be a regression function, i.e. a conditional mathematical expectation.

2.2. The classical formulation of the regression analysis problem

In the classical statement of regression ploblem, the values of X are not random. Only the target variable is random, which is represented as

$$y(x) = \hat{f}(x) + \delta, \tag{2.1}$$

where $\hat{f}(x)$ is some unknown function, and δ is a random variable with zero mean and variance σ^2.

3. Bias and variance

3.1. Decomposition for MSE

For arbitrary independent random variables u and v (if the corresponding moments exist), the identity holds

$$\mathsf{E}(u - v)^2 = \mathsf{D}u + (\mathsf{E}u - \mathsf{E}v)^2 + \mathsf{D}v,$$

where D denotes variance, i.e. $\mathsf{D}u \equiv \mathsf{E}u^2 - (\mathsf{E}u)^2$.

Let's fix a point x of the feature space and substitute $u = y|x$, $v = f(x)$. Since $f(x)$ is constructed on a random sample, v is a random variable. Then we get

$$\mathsf{E}_{S_N,y|x}(y - f(x))^2 =$$

$$\mathsf{D}_{y|x}y + (\mathsf{E}_{y|x}y - \mathsf{E}_{S_N}f(x))^2 + \mathsf{D}_{S_N}f(x). \tag{3.1}$$

The notation $\mathsf{E}_{S_N,y|x}$ means that the expectation is taken over all samples of size N and over the conditional distribution on the target variable y at the point x. So, a subscript at operators E or D indicates the domain for averaging.

We obtain that 3.1 in this formulation is the decomposition of MSE into "noise", bias and variance.

Note that this decomposition is done for each point x. If necessary, 3.1 can be additionally averaged over X.

3.2. Decomposition for kNN when explanatory features are not random

A number of sources (e.g. [4]) provide the following decomposition formula for the kNN method

$$\mathsf{E}_{S_N,y|x}(y - f(x))^2 = \left(f(x) - \frac{1}{k}\sum_{i=1}^{k}\hat{f}(\xi_i(x))\right)^2 + \frac{\sigma^2}{k} + \sigma^2, \qquad (3.2)$$

where $\xi_i(x)$ is the coordinates of the i–th "neighbor" of a point x.

The second term in this decomposition is proposed to be interpreted as a variance.

This expression is obtained for the case when the coordinates of x^ω in the training sample are fixed, i.e. for the statement 2.1.

Note that the conventional definition of the decomposition components assumes complete averaging over random samples.

The variance component in 3.2 decreases monotonically with the growth of k, i.e. it increases with increasing complexity, since the complexity characteristic for kNN is opposite to k and can be, for example, $\frac{1}{k}$.

4. Decomposition for kNN for random features

4.1. Single dimension

Let's consider the problem of regression recovery in the one-dimensional case.

Let $X = [0, 1]$ and $y = x + \delta$, where δ is the so-called "noise", i.e. an independent random variable with variance σ^2 and zero mean. We consider the model $\hat{f}(x) = x$ as a linear regression model (single dimensional) without loss of generality because any coefficient at x will give only a constant factor in expressions.

Suppose that x is a random variable with a uniform distribution, i.e. $x \sim U(0, 1)$. This assumption fundamentally distinguishes this problem statement from 2.1.

Proposition 1. *For the inner points of the segment X, there is:*

$$\mathsf{E}_{S_N,y|x}(y - f(x))^2 \approx \frac{(k+1)(k+2)}{12N^2k} + \frac{\sigma^2}{k} + \sigma^2. \qquad (4.1)$$

The decomposition is asymptotically exact as $N \to \infty$.

Here, the first two terms are variance, the last term is the noise. The bias is approximately 0.

Proof. Let's fix an arbitrary inner point x from the segment X.

Since the sample is random, the coordinates of the neighboring points $\xi_i(x)$ are random variables. With the growth of N, the distributions for $\xi_i(x)$ become (approximately) symmetric with respect to x. It follows that $\mathsf{E}\hat{f}(\xi_i(x)) \approx \hat{f}(x) = x$.

If x is significantly more distant from the edges of the segment X than by $\frac{k}{N}$, then the distributions are almost symmetric and the equality can be considered almost exact.

Thus, the bias component really tends to 0.

Let's calculate the variance component

$$\mathsf{D}f(x) = \mathsf{D}\left[\frac{1}{k}\sum_{i=1}^{k}\xi_i(x)\right] + \frac{\sigma^2}{k}.$$

The values of $\xi_i(x)$ are not independent, but they are uncorrelated (due to the symmetry of conditional distributions), so the variance of the sum is equal to the sum of the variances.

It remains to find $\mathsf{D}\xi_i(x)$.

Consider in X a certain segment of length Δ. The probability $P_\Delta(m)$ that m points from the sample will fall into this segment is

$$P_\Delta(m) = C_N^m \Delta^m (1-\Delta)^{N-m} \approx \frac{(N\Delta)^m}{m!} e^{-N\Delta}.$$

Let $\zeta_i = 2|\xi_i(x) - x|$. Then

$$\mathsf{D}\xi_i(x) = \mathsf{E}(|\xi_i(x) - \mathsf{E}|\xi_i(x))^2 = \mathsf{E}(|\xi_i(x) - x)^2 = \frac{1}{4}\mathsf{E}\zeta_i^2.$$

Let $F_i(\Delta) = \mathsf{P}(\zeta_i < \Delta)$ be a distribution function for $\zeta_i(x)$.

Note that

$$F_i(\Delta) = \sum_{m=i}^{N} P_\Delta(m) = 1 - \sum_{m=0}^{i-1} P_\Delta(m).$$

We have

$$\mathsf{E}\zeta_i^2 = \int_0^\infty \Delta^2 dF_i(\Delta) = \int_0^\infty 2\Delta(1 - F_i(\Delta))d\Delta =$$

$$\int_0^\infty 2\Delta \sum_{m=0}^{i-1} P_\Delta(m)d\Delta = 2\sum_{m=0}^{i-1}\int_0^\infty \frac{m+1}{N} P_\Delta(m+1)d\Delta =$$

$$2\sum_{m=1}^{i}\frac{m}{N}\int_0^\infty P_\Delta(m)d\Delta = \frac{i(i+1)}{N^2}.$$

Now

$$\mathsf{D}\xi_i(x) \approx \frac{i(i+1)}{4N^2}.$$

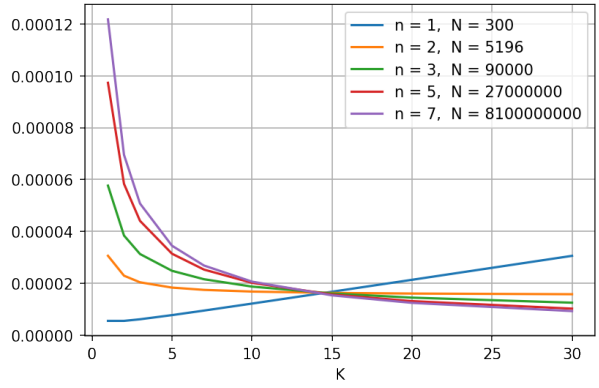

Figure 1. The dependence of the variance component on k for $\sigma = 0$ estimated using the formula 4.2

Using the formula $\sum\limits_{i=1}^{k} i^2 = \frac{k(k+1)(k+2)}{6}$ we get

$$\mathsf{D}\left[\frac{1}{k}\sum_{i=1}^{k}\xi_i(x)\right] = \frac{1}{k^2}\sum_{i=1}^{k}\frac{i(i+1)}{4N^2} = \frac{(k+1)(k+2)}{12N^2k}.$$

So, we obtained the desired expression. □

In contrast to 2.1, the resulting decomposition has a monotonically increasing (close to linear growth) term in the variance component. This term provides the possibility of decreasing variance with increasing complexity.

4.2. Multi-dimensional case

Let $X = [0,1]^n$ and $y = x_1 + \delta$, where $x = (x_1, ..., x_n) \in X$.

We consider the model $\hat{f}(x) = x_1$ as a linear regression model without loss of generality because any linear model may be converted to it via proper transformation of features.

Suppose that x_j are independent random variables, $x_j \sim U(0,1)$.

Proposition 2. *For the inner points of X, there is:*

$$\mathsf{E}_{S_N,y|x}(y - f(x))^2 = \mathsf{D}\left[\frac{1}{k}\sum_{i=1}^{k}\xi_i(x)\right] + \frac{\sigma^2}{k} + \sigma^2,$$

where

$$\mathsf{D}\left[\frac{1}{k}\sum_{i=1}^{k}\xi_i(x)\right] \approx \frac{(N^*)^{-\frac{2}{n}}}{2n^2k^2}\sum_{m=0}^{k-1}\frac{k-m}{m!}\Gamma\left(m+\frac{2}{n}\right) \tag{4.2}$$

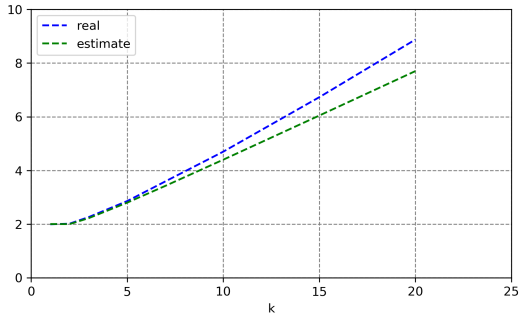

Figure 2. The dependence of the variance component on k for $\sigma = 0$. Curve "real" simulation result at $N = 300$, curve "estimate" estimate using the formula 4.1. For convenience, the values are multiplied by $4N^2$

and

$$N^* = NV_0, \quad V_0 = \frac{\pi^{\frac{n}{2}}}{2^n \Gamma\left(1 + \frac{n}{2}\right)}.$$

The decomposition is asymptotically exact as $N \to \infty$.

Here V_0 is the n-dimensional volume of a Euclidean ball of diameter 1.

If X is an arbitrary finite interval from R^n, then one need to take $N^* = N\frac{V_0}{V}$, where V is the volume (measure) of X.

The proof for proposition 2 is is similar to the proof of proposition 1, but is more cumbersome, so we will omit it.

For $n = 2$ the formula 4.2 get simple

$$\mathsf{D}\left[\frac{1}{k}\sum_{i=1}^{k}\xi_i(x)\right] = \frac{k+1}{4\pi N k}.$$

We can see that the variance for kNN demonstrates different behavior depending on dimensionality. By $n = 1$ the variance increases as k increases. By $n = 2$ the variance tends to a positive constant when $k \to \infty$. By $n > 2$ the variance tends to zero.

Some examples are shown on figure 1.

5. Experimental results

To evaluate the accuracy of the estimate by formula 4.1 for finite N, we performed statistical modeling on synthetic data (single dimension) using the implementation of the kNN method from the scikit-learn library.

The synthetic data were drawn using the model from section 4.1: $y = \hat{f}(x) + \delta$, $\hat{f}(x) = x$, $x \sim U(0, 1)$.

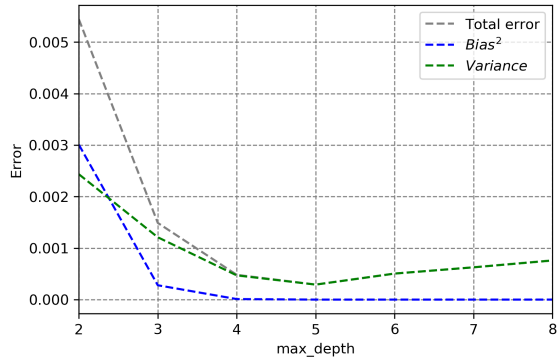

Figure 3. The dependence of the bias and variance on the depth of the tree (the complexity parameter) in the problem of one-dimensional linear regression

In Fig. 2 the simulation result is given together with the theoretical estimate 4.1.

As you can see, the qualitative behavior is consistent, but with the growth of k, the discrepancy increases. The discrepancy is mainly explained by the fact that during modeling, the variance was averaged over the entire segment X. At the same time, at the edges of the segment X, the estimate 4.1 has a significant error due to the fact that the distributions for $\xi_i(x)$ become asymmetric.

In Fig. 3 the result of modeling (on the same regression model) for decision trees is given.

We see that the variance also demonstrates "atypical" behavior, namely, it decreases (on the first half of the plot) with increasing complexity.

6. Discussion

Based on the results obtained, the following conclusions can be drawn: the variance with increasing complexity can not only grow, but also decrease; depending on the assumptions about the randomness or non-randomness of X, qualitatively different results can be obtained in the regression problem.

The discovered facts make the explanation of the learning process, as well as the justification of the effectiveness of collective methods for constructing decision functions, using this decomposition less convincing.

Such undesired properties of bias-variance decomposition encourage to search alternatives. As such alternative might be considered another decomposition of the error: into a measure of adequacy and a measure of stability, that was proposed in [9] [10]. Since the cited references are hardly accessible, one can read brief statement of the concept in [15]. The idea of the approach is to decompose the error into the approximation

error and the statistical error. The components of this decomposition are obviously monotonic. However, this decomposition also has disadvantages, in particular, there is a difficulty in determining the adequacy measure for methods containing regularization by sample size.

It is relevant to study the question to what extent and in what cases the bias can be used as an assessment of the adequacy measure.

The topics for future work might be a generalisation of obtained results to the case of non-uniform distributions, as well as the revealing the relationship between the bias-variance decomposition properties for kNN and other methods.

References

1. d'Ascoli S., Refinetti M., Biroli G., and Krzakala F. Double trouble in double descent: bias and variance(s) in the lazy regime. *In Proceedings of the 37th International Conference on Machine Learning (ICML'20)*. JMLR.org, 2020, article 213, pp. 2280–2290.
2. Belkin M., Hsu D., Ma S., Mandal S. Reconciling modern machine learning practice and the classical bias-variance trade-off. *P*roceedings of the National Academy of Sciences, 2019, vol. 116, no. 32, pp. 15849–1585. https://doi.org/10.1073/pnas.1903070116
3. Berikov V. Semi-supervised classification using multiple clustering and low-rank matrix operations. *Lecture Notes in Computer Science*, 2019, vol. 11548, pp 529–540. https://doi.org/10.1007/978-3-030-22629-9_37
4. Hastie T., Tibshirani R., Friedman H., Jerome. *The Elements of Statistical Learning*. 2009.
5. Heskes T. Bias/variance decompositions for likelihood-based estimators. *Neural Computation*, 1998, vol. 10, no. 6, pp. 1425–1433. http://doi.org/10.1162/089976698300017232
6. Kanevskiy D., Vorontsov K. Cooperative coevolutionary ensemble learning. *Multiple Classifier Systems, 7th International Workshop*, MCS 2007, Prague, Czech Republic, May, 23–25, 2007, pp. 469–478. http://dx.doi.org/10.1007/978-3-540-72523-7_47
7. Kotsiantis S. Bagging and boosting variants for handling classifications problems: A survey. *The Knowledge Engineering Review*, 2014, vol. 29, no. 1, pp. 78–100. http://dx.doi.org/10.1017/S0269888913000313
8. Lbov G., Startseva N. On some a concept of complexity of a strategy of nature in pattern recognition *Data Analysis in Expert Systems. Computational systems*, 1986, vol. 117, pp. 91–102. (in Russian).
9. Lbov G.S. , Startseva N.G. Complexity of distributions in classification problems. *Russ. Acad. Sci., Dokl. Math.*, 1994, vol. 50, no. 2.
10. Lbov G., Startseva N. *Logical Decision Functions and the Problem of Statistical Robustness of Solutions* Novosibirsk: Institute of Mathematics SB RAS, 1999. (in Russian).
11. Nakkiran P., Kaplun G., Bansal Y., Yang T., Barak B., Sutskever I. Deep double descent: where bigger models and more data hurt. *Journal of Statistical Mechanics: Theory and Experiment*, 2021.
12. Neal B., Mittal S., Baratin A., Tantia V., Scicluna M., Lacoste-Julien S., Mitliagkas I. *A Modern Take on the Bias-Variance Tradeoff in Neural Networks*. 2018. https://doi.org/10.48550/arXiv.1810.08591

13. Nedel'ko V. Some aspects of estimating a quality of decision functions construction methods. *Tomsk state university. Journal of control and computer science*, 2013, vol. 3, no. 24, pp. 123–132. (in Russian)
14. Nedel'ko V. Statistical fitting criterion on the basis of cross-validation estimation. *Pattern Recognition and Image Analysis*, 2018, vol. 28, pp. 510–515. http://dx.doi.org/10.1134/S1054661818030148
15. Nedel'ko V. On decompositions of decision function quality measure. *Bulletin of Irkutsk State University. Series Mathematics*, 2020, vol. 33, pp. 64–79. http://dx.doi.org/10.26516/1997-7670.2020.33.64
16. Nedel'ko V. Tight risk bounds for histogram classifier. *Proceedings of IFOST-2016 11th International Forum on Strategic Technology*, 2016, pp. 267–271.
17. Rachakonda A.R., Bhatnagar A. Aratio: Extending area under the roc curve for probabilistic labels. *Pattern Recognition Letters*, 2021, vol. 150, pp. 265–271. http://dx.doi.org/https://doi.org/10.1016/j.patrec.2021.06.023
18. Rudakov K. Mathematical Foundations for Processing High Data Volume. *Machine Learning and Artificial Intelligence, Pattern Recognit. Image Anal*, 2019, vol. 29, pp. 339–343. http://dx.doi.org/10.1134/S1054661819030192
19. Stuart G., Bienenstock E., Doursat R. Neural networks and the bias/variance dilemma. *Neural Computation*, 1992, vol. 4, no. 1, pp. 1–58. http://dx.doi.org/10.1162/neco.1992.4.1.1
20. Yang Z., Yu Y., You C., Steinhardt J., Ma Y. Rethinking bias-variance trade-off for generalization of neural networks. *International Conference on Machine Learning*, 2020, pp. 10767–10777.

Список источников

1. Double trouble in double descent: bias and variance(s) in the lazy regime / S. d'Ascoli, M. Refinetti, G. Biroli, F. Krzakala // Proceedings of the 37th International Conference on Machine Learning (ICML'20). 2020. JMLR.org. Article 213. P. 2280–2290.
2. Reconciling modern machine learning practice and the classical bias-variance trade-off / M. Belkin, D. Hsu, S. Ma, S. Mandal // Proceedings of the National Academy of Sciences. 2019. Т. 116, № 32. С. 15849–15854. https://doi.org/10.1073/pnas.1903070116
3. Berikov V. Semi-supervised classification using multiple clustering and low-rank matrix operations // Lecture Notes in Computer Science. 2019. Т. 11548. С. 529–540. https://doi.org/10.1007/978-3-030-22629-9_37
4. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. 2009.
5. Heskes T. Bias/variance decompositions for likelihood-based estimators // Neural Computation. 1998. Vol. 10, N 6. P. 1425–1433. http://doi.org/10.1162/089976698300017232
6. Kanevskiy D., Vorontsov K. Cooperative coevolutionary ensemble learning // Multiple Classifier Systems, 7th International Workshop, MCS 2007, Prague, Czech Republic, May 23–25. 2007. С. 469–478. http://dx.doi.org/10.1007/978-3-540-72523-7_47
7. Kotsiantis S. Bagging and boosting variants for handling classifications problems: A survey // The Knowledge Engineering Review. 2014. Vol. 29, N 1. P. 78–100. http://dx.doi.org/10.1017/S0269888913000313
8. Лбов Г. С., Старцева Н. Г. Об одном понятии сложности стратегии природы в распознавании образов // Анализ данных в экспертных системах. Новосибирск, 1986. Вып. 117 : Вычислительные системы. С. 91–102.

9. Лбов Г. С., Старцева Н. Г. Сложность распределений в задачах классификации // Доклады РАН. 1994. Т. 338, № 5. С. 592–594.
10. Лбов Г. С., Старцева Н. Г. Логические решающие функции и вопросы статистической устойчивости решений. Новосибирск : Изд-во Ин-та математики, 1999. 211 с.
11. Deep double descent: where bigger models and more data hurt / P. Nakkiran, G. Kaplun, Y. Bansal, T. Yang, B. Barak, I. Sutskever // Journal of Statistical Mechanics: Theory and Experiment. 2021.
12. A Modern Take on the Bias-Variance Tradeoff in Neural Networks / B. Neal, S. Mittal, A. Baratin, V. Tantia, M. Scicluna, S. Lacoste-Julien, I. Mitliagkas // 2018. https://doi.org/10.48550/arXiv.1810.08591
13. Неделько В. М. Некоторые вопросы оценивания качества методов построения решающих функций // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. Т. 24, № 3. С. 123–132.
14. Nedel'ko V. Statistical fitting criterion on the basis of cross-validation estimation // Pattern Recognition and Image Analysis. 2018. Vol. 28. P. 510–515. http://dx.doi.org/10.1134/S1054661818030148
15. Nedel'ko V. On decompositions of decision function quality measure // Известия Иркутского государственного университета. Серия Математика. 2020. Т. 33. С. 64–79. http://dx.doi.org/10.26516/1997-7670.2020.33.64
16. Nedel'ko V. Tight risk bounds for histogram classifier // Proceedings of IFOST-2016 11th International Forum on Strategic Technology. 2016. С. 267–271.
17. Rachakonda A. R., Bhatnagar A. Aratio: Extending area under the roc curve for probabilistic labels // Pattern Recognition Letters. 2021. Vol. 150. P. 265–271. http://dx.doi.org/https://doi.org/10.1016/j.patrec.2021.06.023
18. Rudakov K. Mathematical Foundations for Processing High Data Volume // Machine Learning and Artificial Intelligence, Pattern Recognit. Image Anal. 2019 . Vol. 29. P. 339–343. http://dx.doi.org/10.1134/S1054661819030192
19. Stuart G., Bienenstock E., Doursat R. Neural networks and the bias/variance dilemma // Neural Computation. 1992. Vol. 4, N 1. P. 1–58. http://dx.doi.org/10.1162/neco.1992.4.1.1
20. Rethinking bias-variance trade-off for generalization of neural networks / Z. Yang, Y. Yu, C. You, J. Steinhardt, Y. Ma // International Conference on Machine Learning. 2020. С. 10767–10777.

Об авторах

Неделько Виктор Михайлович , канд. физ.-мат. наук, доц., старший научный сотрудник, Институт математики им. С. Л. Соболева СО РАН, Российская Федерация, 630090, г. Новосибирск, nedelko@math.nsc.ru, https://orcid.org/0000-0001-8217-9304

About the authors

Victor M. Nedel'ko, Cand. Sci. (Phys.Math.), Assoc. Prof., Senior Researcher, Sobolev Institute of Mathematics SB RAS, Novosibirsk, 630090, Russian Federation, nedelko@math.nsc.ru, https://orcid.org/0000-0001-8217-9304

Поступила в редакцию / Received 05.12.2022
Поступила после рецензирования / Revised 16.01.2023
Принята к публикации / Accepted 23.01.2023