УДК 517.929
MSC 34K20

Вестник СПбГУ. Прикладная математика. Информатика... 2023. Т. 19. Вып. 2

О влиянии центральной тенденции на характер плотности распределения максимальной энтропии в машинном обучении

А. В. Квснов[1], *А. А. Бараненко*[2], *Е. Ю. Бутырский*[3], *У. П. Зараник*[3]

[1] Санкт-Петербургский государственный университет Петра Великого, Российская Федерация, 195251, Санкт-Петербург, ул. Политехническая, 29
[2] Военный учебно-научный центр Военно-Морского флота «Военно-морская академия имени Адмирала Флота Советского Союза Н. Г. Кузнецова» Российская Федерация, 197045, Санкт-Петербург, Ушаковская наб, 17/1
[3] Санкт-Петербургский государственный университет, Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

Для цитирования: *Квснов А. В., Бараненко А. А., Бутырский Е. Ю., Зараник У. П.* О влиянии центральной тенденции на характер плотности распределения максимальной энтропии в машинном обучении // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2023. Т. 19. Вып. 2. С. 176–184. https://doi.org/10.21638/11701/spbu10.2023.204

Принцип максимальной энтропии (МЭ) обладает рядом преимуществ, позволяющих применять его в машинном обучении. Плотность распределения максимальной энтропии (ПРМЭ) требует решения задачи вариационного исчисления на априорном распределении, где в качестве параметра может быть использована центральная тенденция, которая в пространстве Лебега описывается обобщенным средним по Гельдеру. В работе показана эволюция плотности распределения МЭ в зависимости от заданной нормы среднего. Минимум расхождения Кульбака — Лейблера между ПРМЭ и априорной плотностью достигается на гармоническом среднем, что эффективно для сокращения размерности обучающей выборки. В то же время это приводит к ухудшению функции потерь в условиях машинного обучения по прецедентам.

Ключевые слова: принцип максимальной энтропии, распределение максимальной энтропии, центральная тенденция, обобщенное среднее, машинное обучение.

1. Введение. Принцип максимальной энтропии (МЭ) для многомерных плотностей вероятности является одним из эффективных методов обработки данных в машинном обучении. Как отмечал основоположник этого принципа, «необходимо отыскать такое распределение с максимальной энтропией, которая была бы максимально неинформативна и тем самым обеспечивала его параметрическую оценку» в работах [1, 2]. Тем не менее остается открытым вопрос о выборе априорных данных (в том числе центральных моментов) для описания «параметризированной» статистики плотности распределения максимальной энтропии (ПРМЭ) для задач машинного обучения по прецедентам («обучение с учителем»). Требуется получить такую ПРМЭ, которая бы, с одной стороны, редуцировала размерность обучающей выборки, а с другой — минимизировала функцию потерь.

В ряде работ анализ ПРМЭ строится в контексте энтропийной оптимизации стохастических матриц-проекторов (ERP-метод) [3–5]. На основе равномерного многообразия (Uniform Manifold Sampling — UMS) предложен метод по уточнению необходимой выборки для произвольной размерности [6]. В отдельных случаях показывают

хорошие результаты сингулярное разложение по параметрам выборки [7] и метод на основе минимаксных решающих правил [8]. В то же время основной инструмент снижения размерности — метод главных компонент (Principle Component Analysis — PCA) — ограничен в использовании из-за сложности расчета собственных значений для больших размерностей [9]. В настоящей работе предлагается оценить эволюцию МЭ при известном условии средней тенденции в пространстве Лебега L^{ℓ}. Норма ℓ выполняет функцию регулятора, обеспечивающего решение вариационной задачи на пространстве многомерных выборок из распределения.

2. Постановка задачи. Пусть известна обучающая выборка на вероятностном пространстве $\mathbb{B} \subseteq \{x \subset \{C_n^m\}, \mathfrak{F}, \mathbb{P}\}$, где $x \subset \{C_n^m\}$ — множество возможных исходов выборки из сочетания C_n^m; $\mathfrak{F}$ — сигма-алгебра; $\mathbb{P}$ — метрическое распределение. Введем процедуру обучения согласно условию [10]

$$Y \equiv A(X) \quad \forall A : \mathbb{R}^{m\times n} \to \mathbb{R}^{n\times 1},$$

где $A \subset \mathbb{B}$ — линейный оператор на вероятностном пространстве $\mathbb{B}$; $X \in \mathbb{R}^{m\times n}$ — массив обучающих данных для n наблюдений и m признаков; $Y \in \mathbb{N}^{k\times 1}$ — вектор отметок для прецедентов.

Рассмотрим преобразование (отображение) T, осуществляющее максимизацию энтропии обучающей выборки $X \subset \{x_i \times x_j\}_{i\in n, j\in m}$:

$$T : X \to \mathcal{H}(X),\ X \in \mathbb{R}^{N\times M},\ N \gg M, \tag{1}$$

здесь $\mathcal{H}(X) = \sum\limits_{i\in n, j\in m} p_{i,j}(x) \log p_{i,j}(x)$ задает многомерное распределение МЭ относительно аргумента $X : \{x\}_{i=1,j=1}^{n,m}$.

Последнее условие в (1) $N \gg M$ обозначает, что количество признаков намного меньше количества наблюдений для выборки.

Согласно принципу неопределенности Лапласа,

$$\forall p_{i,j}(x) \in unif(\overline{x}, \sigma_x^2) : \mathcal{H}(X) \to \max.$$

При известной априорной центральной тенденции ПРМЭ, строго говоря, может подчиняться произвольному закону распределения [11]. Поэтому оценка этого распределения МЭ задает критерии (параметры распределения, его границы), по которым мы прогнозируем различные сценарии «обучения по прецедентам» [12].

3. Эволюция МЭ в пространстве средних. Для произвольного распределения $\widetilde{p}(x)$ среднее M, вообще говоря, задается по Колмогорову:

$$\mathcal{M}(x_1, x_2, \ldots, x) = \varphi^{-1}\left\{\frac{1}{n}\sum_{k=1}^{n}\varphi(x_k)\right\},$$

где φ — непрерывная строго монотонная функция; φ_1 — функция, обратная к φ. Поскольку на вероятностном пространстве задано множество изолированных точек $x \subset \{C_n^m\}$, то введем обобщенное среднее относительно его отображения в пространстве экстремальных распределений.

Определение. Плотность распределения $p(x) : X \to X\ \forall x \in X$ с обобщенным средним (среднее по Гельдеру — *Hölder mean*) $\mathcal{M}_{\ell} \overset{Hel}{=} \left(\frac{1}{mn}\sum\limits_{j=1}^{m}\sum\limits_{i=1}^{n} x_{ij}^{\ell}\right)^{\ell^{-1}}$ для

$\ell \in (-\infty, +\infty)$ и $\ell \in \mathbb{Z}$ в пространстве экстремальных распределений $T : X \to \mathcal{H}(X)$ есть функция

$$\mathcal{H}(X) = T(x|\mathcal{M}_\ell) \quad \forall x \sim p(x).$$

Используя вариационное исчисление, утверждаем, что знание априорного среднего $\mathcal{M}_\ell$ позволяет достичь такого распределения $\mathcal{H}(\cdot)$, при котором линейный оператор

$$\mathcal{A} : C_{(\ell)}\{X \to Y\} \longrightarrow \min \ \ \forall \ell \in (-\infty, +\infty),$$

где $\mathcal{C}_{(\ell)}$ — функция потерь для оператора $\mathcal{A}$.

Теорема. *Если $p(x)$ — плотность распределения выборки*

$$(x_{11}, \ldots, x_{1m}, \ldots, x_{n1}, \ldots, x_{nm}) \subseteq X$$

с обобщенным средним по Гельдеру M_ℓ $\forall \ell \subset \{(-\infty, +\infty) \vee \mathbb{Z}\}$, тогда распределение максимальной энтропии $\mathcal{H}(X) \Leftrightarrow T\{p(x)\}$ $\forall X \in \mathbb{R}^{n \times m}$ будет удовлетворять условию

$$\begin{cases} \sup\{\mathcal{H}(X)\} & \xrightarrow{\ell \to -\infty} \quad p(x), \\ \inf\{\mathcal{H}(X)\} & \xrightarrow{\ell \to +\infty} \quad \max[G_p(x)], \end{cases}$$

где $G_p(\cdot)$ — *распределение Гиббса.*

Доказательство. Доказательство теоремы осуществляется по методу множителей Лагранжа. Более подробно этот подход описан в работах [11, 13]. Рассмотрим дискретное выборочное распределение $p^*(X) : \{x_i\}_{i=1}^n \ \forall x \in X$. Зададим функцию ограничения $\varphi(x)$ для условия выборочного распределения $p^*(X)$:

- поскольку $X \in \mathbb{R}$, можно показать существование эмпирического среднего $\mathcal{M} \stackrel{def}{=} \mathbb{E}[X^\ell]$ $\forall \ell \in \mathbb{Q}$ известной нормой Хельма $\|\mathcal{M}\|_{\ell \in (-\infty, +\infty)}$;

- принцип нормирования распределения $p^*(X)$ позволяет установить, что

$$p^*(X|\mathcal{M}_{\ell \in (-\infty, +\infty)}) : \sum_{i \in \mathbb{R}} p^*(x_i) = 1.$$

Если $\exists \mathcal{H}\{p^*(X)\} \Leftrightarrow p^*(x_i)$ $\forall X \in \mathbb{R}$, тогда границы сходимости $p^*(x)$ к распределению $\mathcal{H}\{p^*(X)\} \equiv \mathcal{H}[\cdot]$ можно определить как

$$\sup[p^*(x_1, \ldots, x_n)] \xrightarrow{n \to +\infty} \mathcal{H}(x) : \left\{ \|\mathcal{M}\|_{l=\infty} \longrightarrow M \vee \sum p^*(x) = \sum_{x \in \mathbb{R}} \mathcal{H}(x) = 1 \right\},$$

$$\inf[p^*(x_1, \ldots, x_n)] \xrightarrow{n \to -\infty} \mathcal{H}(x) : \left\{ \|\mathcal{M}\|_{l=-\infty} \longrightarrow M \vee \sum p^*(x) = \sum_{x \in \mathbb{R}} \mathcal{H}(x) = 1 \right\},$$

здесь $M \sim \mathbb{E}[(x - \mathbb{E}x)]$ — априорное теоретическое среднее значение математического ожидания E.

В соответствии с определением можно установить энтропию

$$\mathcal{H}(p_1, \ldots, p_n) \sim H(x|p^*(x))$$

для функции распределения $p^* \subseteq \{p_1, \ldots, p_n\}$. Принцип максимума указывает, что

$$\forall \{p_1, \ldots, p_n\} \in \mathbb{R}, \mathcal{H}(p_1, \ldots, p_n | p_i \neq 0) \stackrel{def}{=} -\sum_{i=1}^n p_i ln p_i(x).$$

Для известных ограничений $\varphi_{\xi=2}(x)$ величину энтропии для множества $\{p_1,\dots,p_n\}$ из (1) находят по методу множителей Лагранжа (Lagrangian multipliers) [11] как функцию вида

$$L(x,\lambda) = f(x) + \sum_{\xi} \lambda_\xi \varphi_\xi(x), \tag{2}$$

где $\lambda = (\lambda_1,\dots\lambda_\xi)$ — множители Лагранжа; $\varphi_{\xi\Leftrightarrow\Sigma}(x) = \sum\limits_{i=1}^{n} x_i - 1$ — ограничение согласно принципу нормирования плотности вероятности; $\varphi_{\xi\Leftrightarrow\ell}(x) = \mathbb{E}[(x-\mathbb{E}x)] - \mathcal{M} = \sum\limits_{i=1}^{n}(x_i - M)p_i(x) - \mathcal{M}$ — ограничение по условию существования эмпирического среднего $M := \mathbb{E}[X^\ell]$.

Подставив ограничения $\varphi_{\xi\Leftrightarrow\Sigma}(x)$ и $\varphi_{\xi\Leftrightarrow\ell}(x)$ в выражение (2) и задавшись условиями $\lambda_{\xi=1} \Leftrightarrow \lambda^\Sigma$ и $\lambda_{\xi=2} \Leftrightarrow \lambda^\ell$, имеем уравнение

$$L(p_1,\dots,p_n,\lambda^\Sigma,\lambda^\ell) = \mathcal{H}(p_1,\dots,p_n) + \lambda^\Sigma\left[\sum_{i=1}^{n} p_i(x) - 1\right] + \lambda^\ell\left[\sum_{i=1}^{n}(x_i - M)p_i(x) - \mathcal{M}\right].$$

Составим систему из n уравнений и приравняем к нулю частные производные:

$$\begin{cases} \frac{\partial}{\partial p_1} L(p_1,\dots,p_n,\lambda^\Sigma,\lambda^\ell) &= -(\ln p_1 + 1) + \lambda^\Sigma + \lambda^\ell(x_1 - M) = 0, \\ \dots \\ \frac{\partial}{\partial p_n} L(p_1,\dots,p_n,\lambda^\Sigma,\lambda^\ell) &= -(\ln p_n + 1) + \lambda^\Sigma + \lambda^\ell(x_n - M) = 0, \\ 1 = \sum\limits_{i=1}^{n} p_i(x), \\ \mathcal{M} \overset{def}{=} \sum\limits_{i=1}^{n}(x_i - M)p_i(x). \end{cases} \tag{3}$$

Найдем разность между первым и i-м выражениями в системе уравнений (3):

$$p_i = p_1 e^{\lambda^\ell}(x_i - x_1). \tag{4}$$

Объединим (4) с условием для суммы плотности вероятности $p \in \mathbb{P}^n$:

$$\begin{cases} p_i &= p_1 e^{\lambda^\ell(x_i - x_1)}, \\ 1 &= \sum\limits_{i=1}^{n} p_1 e^{\lambda^\ell(x_i - x_1)} = p_1 e^{-\lambda^\ell x_1} \sum\limits_{i=1}^{n} e^{\lambda^\ell x_i}. \end{cases} \tag{5}$$

Из последнего выражения в системе уравнений (5) находим p_i и после подстановки в первое выражение (5) получим, что

$$p_i \frac{e^{\lambda^\ell x_1}}{\sum\limits_{i=1}^{n} e^{\lambda^\ell x_i}} e^{\lambda^\ell(x_i - x_1)} = \frac{e^{\lambda^\ell x_1}}{\sum\limits_{i=1}^{n} e^{\lambda^\ell x_i}}. \tag{6}$$

Из уравнений (3) для частных производных $\frac{\partial}{\partial p_i} L$ имеем $p_i = e^{[\lambda^\Sigma + \lambda^\gamma \lambda^\ell(x_i - M - 1)]}$, которое приравняем к выражению (6). В итоге находим, что

$$e^{[\lambda^\Sigma - \lambda^\ell M - 1]} = \left(\sum_{i=1}^{n} e^{\lambda^\ell x_i}\right)^{-1}.$$

Отсюда имеем уравнение

$$\sum_{i=1}^{n} e^{\lambda^{\ell} x_i} = e^{[1-\lambda^{\Sigma}-\lambda^{\ell} M]}. \tag{7}$$

Воспользуемся выражением для эксцесса из системы уравнений (3) и подставим вместо величины $p_i(x)$ значения из (6):

$$\mathcal{M}=\sum_{i=1}^{n}(x_i-M)p_i(x)=\sum_{i=1}^{n}(x_i-M)e^{\lambda^{\ell} x_i}\left(\sum_{i=1}^{n} e^{\lambda^{\ell} x_i}\right)^{-1}=\left(\sum_{i=1}^{n}(x_i-M)e^{\lambda^{\ell} x_i}\right)\left(\sum e^{\lambda^{\ell} x_i}\right)^{-1}.$$

Поскольку в правой части последнего выражения $\sum\limits_{i=1}^{n} e^{\lambda^{\ell} x_i} \neq 0$ в силу того, что $x_k > 0$, получим равенство

$$\frac{\left(\sum\limits_{i=1}^{n}(x_i-M)e^{\lambda^{\ell} x_i}\right)-\mathcal{M}\sum\limits_{i=1}^{n} e^{\lambda^{\ell} x_i}}{\sum\limits_{i=1}^{n} e^{\lambda^{\ell} x_i}}=0,$$

откуда вытекает, что

$$\sum_{i=1}^{n}\left[(x_i-M)-\mathcal{M}\right]e^{\lambda^{\ell} x_i}=0. \tag{8}$$

Таким образом, окончательно систему уравнений из (6)–(8) можно записать следующим образом:

$$\begin{cases} \sum\limits_{i=1}^{n} e^{\lambda^{\ell} x_i} = e^{[1-\lambda^{\Sigma}-\lambda^{\ell} M]}, \\ p_i = e^{\lambda^{\ell} x_i}(\sum\limits_{i=1}^{n} e^{\lambda^{\ell} x_i})^{-1}, \\ \sum\limits_{i=1}^{n}\left[(x_i-M)-M\right]e^{\lambda^{\ell} x_i}=0. \end{cases} \tag{9}$$

Рассмотрим последнее выражение из (9). Для всех отрицательных λ_{-}^{ℓ} можно показать, что

$$\{x_i|M,\mathcal{M}\}:\lambda_{-}^{\ell}\mapsto\varnothing \quad \forall x_i\in\mathbb{R}.$$

В то же время существует решение для $\{x_i|M,\mathcal{M}\}:\lambda_{+}^{\ell}\mapsto\lambda_{+}^{\ell}$. При этом выполняется условие

$$\begin{cases} \mathcal{M}_{+\infty}:\sup\left\{\lambda_{+}^{\ell}|\left(M \overset{g.m.}{\sim} \mathcal{M}\right)\right\} \xrightarrow{p\to\infty} k \\ \mathcal{M}_{-\infty}:\inf\left\{\lambda_{+}^{\ell}|\left(M_{\gamma} \overset{g.m.}{\sim} \mathcal{M}\right)\right\} \xrightarrow{p\to\infty} 0, \end{cases} \tag{10}$$

где $\overset{g.m.}{\sim}$ — соответствие обобщенному среднему; $k \in (0,1)$ — параметр распределения Гиббса согласно первому выражению. Для второго выражения (10) имеем $\lim\limits_{\lambda_{+}^{\ell}\to 0} p_i(x) \equiv 1$, отсюда обобщенное условие представим в виде

$$\begin{cases} \sup\{\mathcal{H}(x|\mathcal{M}_{-\infty})\} \xrightarrow{\text{MaxEnt}} p^*(x), \ \ p\in(0,1), \\ \inf\{\mathcal{H}(x|\mathcal{M}_{+\infty})\} \xrightarrow{\text{MaxEnt}} \max[G_{p^*}(x)] \quad \forall p\subseteq\mathbb{P}. \end{cases}$$

Таким образом, функция $\mathcal{H}(x)$ для $\|\mathcal{M}\|_{\ell\to-\infty}$ стремится к выборочному распределению, в случае $\|\mathcal{M}\|_{\ell\to-\infty}$ нижняя граница достигается при максимизированном распределении Гиббса. Теорема доказана.

Далее приведем основные следствия из теоремы, которые имеют практическое значение и использованы при формировании математических алгоритмов в системах дистанционного зондирования [14–16]. В ряде случаев ПРМЭ может быть применена для распознавания и классификации текстовой информации [17, 18].

Следствие 1. *Минимум РКЛ D^{-}_{KL} для плотности $p(x_{11},\ldots,x_{nm})$ $\forall x \in X$ с априорным обобщенным средним по Гельдеру $\mathcal{M}_\ell$ и соответствующей ей ПРМЭ $\mathcal{H}(\cdot)$ достигается при $\ell \to -\infty$:*

$$D^{-}_{KL}(p(x|\mathcal{M})\|\mathcal{H}\{\cdot\}) = \arg\min_{\mathcal{M}\in\mathbb{N}}\{p(x|\mathcal{M}) : \mathcal{M}_{\ell\to-\infty}\}. \tag{11}$$

Следствие 2. *Максимум РКЛ D^{+}_{KL} для выборки $p(x_{11},\ldots,x_{nm})$ $\forall x \in X$ с известным обобщенным средним по Гельдеру $M_{\ell\in(-\infty,+\infty)}$ и соответствующей ей максимизированной энтропией $\mathcal{H}(\cdot)$ достигается при $\ell \to +\infty$:*

$$D^{+}_{KL}(p(x|\mathcal{M})\|\mathcal{H}\{\cdot\}) = \arg\max_{\mathcal{M}\in\mathbb{N}}\{p(x|\mathcal{M}) : \mathcal{M}_{\ell\to+\infty}\}.$$

Поскольку соотношение средних на вероятностном пространстве $\mathbb{B}$ задается через неравенство Йенсена

$$\psi(M[X]) \leqslant M[\psi(X)],$$

где $\psi : \mathbb{R} \to \mathbb{R}$ — выпуклая (вниз) борелевская функция, тогда имеем, что

$$\mathcal{M}_{\ell=-1} \leqslant \mathcal{M}_{\ell=0} \leqslant \mathcal{M}_{\ell=1} \leqslant \mathcal{M}_{\ell=2}.$$

Не выходя за границы рассматриваемых средних $\ell \subseteq \{-1, 0, 1, 2\}$ и выводов из (10) и (11), можно утверждать, что априорное значение $\mathcal{M}_{\ell=-1}$ (т. е. гармонического среднего) «параметризирует» плотность распределения МЭ. С одной стороны, это допускает снижение размерности согласно преобразованию T; с другой — очевидно приводит к ухудшению функции потерь $C_{(\ell)}$.

4. Заключение. В работе описана проблема параметризации плотности распределения МЭ для обучающей выборки. Показано, что выбор априорного фактора может быть решен в пространстве средних величин. В этом случае целесообразен выбор гармонического среднего как величины, обеспечивающей достижение минимального различия расхождения Кульбака — Лейбера между априорным и экстремальным распределениями. Стоит отметить, что ограничения, рассмотренные исключительно для средних по Гельдеру, должны быть обобщены для f-средних по Колмогорову. В таком случае необходимо изучить не частный сценарий дискретной энтропии, а общий случай дифференциальной энтропии по Ренье [8].

Литература

1. *Jaynes E. T.* Prior probabilities // IEEE Transactions on Systems Science and Cybernetics. 1968. Vol. 4. P. 227–251.

2. *Jaynes E. T.* Information theory and statistical mechanics // Physical Review. 1957. Vol. 4. N 106. P. 620–630.

3. *Попков Ю. С.* Асимптотическая эффективность оценок максимальной энтропии // Докл. Российской академии наук. Математика. Информатика. Процессы управления. 2020. Т. 493. С. 104–107.

4. *Попков Ю. С., Дубнов Ю. А., Попков А. Ю.* Энтропийно-рандомизированное проектирование // Автоматика и телемеханика. 2021. Т. 82. № 3. С. 149–168.

5. *Попков Ю. С.* Рандомизация и энтропия в машинном обучении и обработке данных // Докл. Российской академии наук. Математика. Информатика. Процессы управления. 2022. Т. 504. № 1. С. 3–27.

6. *Baggenstoss P. M.* Uniform Manifold Sampling (UMS): Sampling the maximum entropy PDF // IEEE Transactions on Signal Processing. 2017. Vol. 65. N 1. P. 2455–2470.

7. *Mak S., Xie Y.* Maximum entropy low-rank matrix recovery // IEEE Journal of Selected Topics in Signal Processing. 2018. Vol. 12. N 5. P. 886–901.

8. *Mazuelas S., Shen Y., Perez A.* Generalized maximum entropy for supervised classification // IEEE Transactions on Information Theory. 2022. Vol. 68. N 4. P. 2530–2550.

9. *Richards M. A., Scheer J. A., Scheer J., Holm W. A.* Principles of modern radar: basic principles. 1st ed. Atlanta: Institution of Engineering and Technology, 2010. 962 p.

10. *Воронцов К. В.* Комбинаторные оценки качества обучения по прецедентам // Докл. Академии наук. 2004. Т. 394. № 2. С. 175–178.

11. *Niven R. K., Andresen B.* Jaynes' maximum etntropy principle, riemannian metrics and generalised least action bound // Statistical Mechanics. 2010. P. 283–317.

12. *Кваснов А. В.* Применение байесовского программирования в задачах распознавания и классификации источников радиоизлучения // Радиотехника. 2020. Т. 84. № 3(5). С. 5–14.

13. *Кваснов А. В.* Исследование информационной полноты радиолокационных данных в задачах классификации точечных воздушных объектов // Журн. радиоэлектроники. 2021. Т. 11. С. 1–10.

14. *Кваснов А. В.* Повышение информационной полноты классификатора в задачах дистанционного зондирования воздушных точечных объектов // Датчики и системы. 2022. Т. 262. № 3. С. 9–14.

15. *Кваснов А. В.* Точность определения координат надводной поверхности на основе фотограмметрических измерений снимка с наклонной проекцией // Сенсорные системы. 2022. Т. 36. № 3. С. 262–274.

16. *Marcial M. C. N., Santillan J. R.* A maximum entropy approach for mapping falcata plantations in sen-tinel-2 imagery // IEEE region 10 Conference (TENCON). 2020. P. 596–601. https://doi.org/10.1109/TENCON50793.2020.9293693

17. *Yin C., Xi J., Wang J.* The research of text classification technology based on improved maximum entropy model // First International Conference on Computational Intelligence Theory, Systems and Applications (CCITSA). 2015. P. 142–145. https://doi.org/10.1109/CCITSA.2015.12

18. *Pan W.* Feature selection algorithm based on maximum information coefficient // IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). 2021. P. 2600–2603. https://doi.org/10.1109/IAEAC50856.2021.9390868

Статья поступила в редакцию 10 марта 2023 г.
Статья принята к печати 25 апреля 2023 г.

Контактная информация:

Кваснов Антон Васильевич — канд. техн. наук, доц.; AntonKV@mail.ru

Бараненко Анатолий Афанасьевич — д-р техн. наук, проф.; abar47@yandex.ru

Бутырский Евгений Юрьевич — д-р физ.-мат. наук, проф.; e.butyrsky@spbu.ru

Зараник Ульяна Петровна — канд. физ.-мат. наук, ст. преп.; u.zaranik@spbu.ru

On the infuence of the cental trend on the nature of the density distribution of maximum entropy in machine learning

A. V. Kvasnov[1], *A. A. Baranenko*[2], *E. Y. Butyrsky*[3], *U. P. Zaranik*[3]

[1] Peter the Great St. Petersburg Polytechnic University, 29, Polytekhnicheskaya ul., St. Petersburg, 195251, Russian Federation

[2] Military Educational and Scientific Center of the Navy "Naval Medical Academy" named after N. G. Kuznetsov, 17/1, Ushakovskaia nab., St. Petersburg, 197025, Russian Federation

[3] St. Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation

For citation: Kvasnov A. V., Baranenko A. A., Butyrsky E. Y., Zaranik U. P. On the influence of the central trend on the nature of the density distribution of maximum enropy in machine lerning. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2023, vol. 19, iss. 2, pp. 176–184.
https://doi.org/10.21638/11701/spbu10.2023.204 (In Russian)

The principle of maximum entropy (ME) has a number of advantages that allow it to be used in machine learning. The density distribution of maximum entropy (WEO) requires solving the problem of calculus of variations on the a priori distribution, where the central tendency can be used as a parameter. In Lebesgue space, the central tendency is described by the generalized Gelder average. The paper shows the evolution of the density of the ME distribution depending on the given norm of the average. The minimum Kulback — Leibler divergence between the WEO and the a prior density is achieved at the harmonic mean, which is effective in reducing the dimensionality of the training sample. At the same time, this leads to a deterioration in the function of loss in the conditions of machine learning by precedents.

Keywords: maximum entropy principle, maximum entropy distribution, central trend, generalized average, machine learning.

References

1. Jaynes E. T. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, 1968, vol. 4, pp. 227–251.

2. Jaynes E. T. Information theory and statistical mechanics. *Physical Review*, 1957, vol. 4, no. 106, pp. 620–630.

3. Popkov Y. S. Asimptoticheskaia effektivnost' otsenok maksimal'noi entropii [Asymptotic efficiency of maximum entropy estimates]. *Reports of the Russian Academia of Sciences. Mathematics. Computer Science. Control Processes*, 2020, vol. 493, pp. 104–107. (In Russian)

4. Popkov Y. S., Dubnov Y. A., Popkov A. Y. Entropiino-randomizirovannoe proektirovanie [Entropy-randomized projection]. *Automation and Telemechanics*, 2021, vol. 82, no. 3, pp. 149–168. (In Russian)

5. Popkov Yu. S. Randomizatsiia i entropiia v mashinnom obuchenii i obrabotke dannykh [Randomization and entropy in machine learning and data processing]. *Reports of the Russian Academy of Sciences. Mathematics. Computer Science. Control Processes*, 2022, vol. 504, no. 1, pp. 3–27. (In Russian)

6. Baggenstoss P. M. Uniform Manifold Sampling (UMS): Sampling the Maximum entropy PDF. *IEEE Transactions on Signal Processing*, 2017, vol. 65, no. 1, pp. 2455–2470.

7. Mak S., Xie Y. Maximum entropy low-rank matrix recovery. *IEEE Journal of Selected Topics in Signal Processing*, 2018, vol. 12, no. 5, pp. 886–901.

8. Mazuelas S., Shen Y., Perez A. Generalized maximum entropy for supervised classification. *IEEE Transactions on Information Theory*, 2022, vol. 68, no. 4, pp. 2530–2550.

9. Richards M. A., Scheer J. A., Scheer J., Holm W. A. *Principles of modern radar: basic principles.* 1st ed. Atlanta, Institution of Engineering and Technology Publ., 2010, 962 p.

10. Vorontsov K. V. Kombinatornye otsenki kachestva obucheniia po pretsedentam [Combinatorial bounds for learning performance]. *Reports of the Russian Academia of Sciences*, 2004, vol. 394, no. 2, pp. 175–178. (In Russian)

11. Niven R. K., Andresen B. Jaynes' maximum entropy principle, riemannian metrics and generalised least action bound. *Statistical Mechanics*, 2010, pp. 283–317.

12. Kvasnov A. V. Primenenie baiesovskogo programmirovaniia v zadachakh raspoznavaniia i klassifikatsii istochnikov radioizlucheniia [Application of bayesian programming in recognition and classification of radar emission sources]. *Radio Engineering*, 2020, vol. 84, no. 3(5), pp. 5–14.(In Russian)

13. Kvasnov A. V. Issledovanie informatsionnoi polnoty radiolokatsionnykh dannykh v zadachakh klassifikatsii tochechnykh vozdushnykh ob"ektov [Enhance of information completeness of classifier in remote sensing tasks of aerial point objects]. *Journal of Radio Electronics*, 2021, vol. 11, pp. 1–10.(In Russian)

14. Kvasnov A. V. Povyshenie informatsionnoi polnoty klassifikatora v zadachakh distantsionnogo zondirovaniia vozdushnykh tochechnykh ob"ektov [Enhance of information completeness of classifier in remote sensing tasks of aerial point objects]. *Sensors and Systems*, 2022, vol. 262, no. 3, pp. 9–14. (In Russian)

15. Kvasnov A. V. Tochnost' opredeleniia koordinat nadvodnoi poverkhnosti na osnove fotogrammetricheskikh izmerenii snimka s naklonnoi proektsiei [Accuracy of the coordinates for underlying surface based on photogrammetric measurements of the images with an oblique projection]. *Sensor Systems*, 2022, vol. 36, no. 3, pp. 262–274. (In Russian)

16. Marcial M. C. N., Santillan J. R. A maximum entropy approach for mapping falcata plantations in sentinel-2 imagery. *IEEE region 10 Conference (TENCON)*, 2020, pp. 596–601. https://doi.org/10.1109/TENCON50793.2020.9293693

17. Yin C., Xi J., Wang J. The research of text classification technology based on improved maximum entropy model. *First International Conference on Computational Intelligence Theory, Systems and Applications (CCITSA)*, 2015, pp. 142–145. https://doi.org/ 10.1109/CCITSA.2015.12

18. Pan W. Feature selection algorithm based on maximum information coefficient. *IEEE 5^{th} Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2021, pp. 2600–2603. https://doi.org/10.1109/IAEAC50856.2021.9390868

Received: March 10, 2023.
Accepted: April 25, 2023.

Authors' information:

Anton V. Kvasnov — PhD in Technical Sciences, Associate Professor; AntonKV@mail.ru

Anatoliy A. Baranenko — Dr. Sci. in Technical Sciences, Professor; abar47@yandex.ru

Evgeniy Y. Butyrsky — Dr. Sci. in Physics and Mathematics, Professor; e.butyrsky@spbu.ru

Uliana P. Zaranik — PhD in Physics and Mathematics, Senior Lecturer; u.zaranik@spbu.ru