

УДК 519.23

## ПОЛУЧЕНИЕ РАЗРЕЖЕННЫХ РЕШЕНИЙ С ИСПОЛЬЗОВАНИЕМ D-ОПТИМАЛЬНОГО РАЗБИЕНИЯ ИСХОДНОЙ ВЫБОРКИ НА ОБУЧАЮЩУЮ И ТЕСТОВУЮ ЧАСТИ И КРИТЕРИЯ РЕГУЛЯРНОСТИ

**А. А. Попов, Ш. А. Бобоев**

*Новосибирский государственный технический университет*  
*a.popov@corp.nstu.ru, shboboev@mail.ru*

В работе рассмотрены способы получения разреженного решения на основе метода LS-SVM. Для получения разреженного решения используются обучающие выборки, формируемые по схеме D-оптимального планирования эксперимента. Приводится также алгоритм формирования тестовой выборки с использованием критерия регулярности. Для проверки эффективности предлагаемых методов формирования обучающей выборки проведен вычислительный эксперимент с использованием модельных данных. Повышение обобщающих свойств моделей, полученных по LS-SVM, проводилось через подбор параметра гауссовой ядерной функции при минимизации ошибки прогноза. Параметр ядерной функции подбирался по минимуму внешних критериев. В качестве внешних критериев оценки качества моделей использовались критерий скользящего контроля и критерий регулярности. Эффективность получаемых решений оценивалась по среднеквадратичной ошибке. Дисперсия помехи (уровень шума) при моделировании данных задавалась в процентах от мощности сигнала. По итогам полученных результатов сделаны выводы, что для получения разреженного решения можно использовать обучающую выборку, полученную с использованием D-оптимального разбиения, а также дополнительную оптимизацию состава точек (алгоритм замены), которая в ряде случаев позволяет получать лучшие решения.

*Ключевые слова:* регрессионная модель, метод LS-SVM, ядерная функция, квадратичная функция потерь, D-оптимальный план, критерий скользящего контроля, среднеквадратичная ошибка.

## OBTAINING OF SPARSE SOLUTIONS WITH D-OPTIMAL PARTITIONING OF ORIGINAL SAMPLE INTO TRAINING AND TEST PARTS AND REGULARITY CRITERION

**A. A. Popov, Sh. A. Boboev**

*Novosibirsk State Technical University*  
*a.popov@corp.nstu.ru, shboboev@mail.ru*

The article considers the methods of obtaining sparse solutions based on the LS-SVM method. The training samples generated according to the D-optimal experiment design are used for obtaining a sparse solution. An algorithm for forming a test sample using the regularity criterion is also presented. A computational experiment is conducted to check the efficiency of the proposed methods of forming a training sample. The artificial data is used for this purpose. The improvement of the generalizing properties of the models obtained by LS-SVM is carried out through the selection of the Gaussian kernel function parameter while minimizing the prediction error. The kernel function parameter is chosen according to the minimum of external criteria. The criteria of cross-validation and regularity are used as external criteria for assessing the quality of models. The effectiveness of the obtained solutions is estimated by mean square error. The noise variance (noise level) is set as a percentage of the signal power during the data modeling. According to the results, conclusions are made that to receive a sparse solution it is possible to use a training sample, obtained with the use of D-optimal split. The additional optimization of the composition of points (exchange algorithm) is possible to use as well, which in some cases allows getting the best solutions.

*Keywords:* regression model, LS-SVM method, kernel function, quadratic loss function, D-optimal design, cross-validation criterion, mean square error.

**Введение.** В настоящее время для построения регрессии предложено несколько модификаций алгоритма опорных векторов [1–2]. При этом все большую популярность приобретает алгоритм опорных векторов на основе квадратичной функции потерь (LS-SVM) [1]. Однако его применение требует решения задачи по настройке его ряда внутренних параметров. Эвристический подход по априорному их выбору описан в [3]. Подбор этих параметров можно также осуществлять через минимизацию критерия скользящего контроля [4–8].

При использовании классического SVM имеется возможность построения разреженных решений. Разреженное решение характеризуется тем, что в аддитивном его разложении по ядерным функциям задействуются не все точки выборки. Это достигается за счет использования функции потерь  $\varepsilon$ -нечувствительности Вапника. Построение разреженных решений особенно важно в случае выборок больших объемов. Стоит заметить, что при использовании линейной и полиномиальной ядерных функций проблем с компактным представлением модели не возникает. Для этих функций всегда можно вернуться в исходное пространство данных и в результате получить сжатую форму представления модели. В то же время при использовании, например, ядерной функции Гаусса сделать это не удастся. Модель в этом случае записывается в виде суммы ядер. Максимальное число ядер может быть равно числу наблюдений в выборке. Можно поставить задачу по сокращению числа слагаемых в этой сумме ядер. При ее решении мы будем получать так называемое разреженное решение.

При использовании метода SVM с квадратичной функцией потерь для получения разреженных решений приходится прибегать к специальным приемам. Например, для их получения можно воспользоваться подходом, предложенным в работе [9]. Основная идея при этом состоит в отбрасывании точек выборки, для которых параметры в аддитивном разложении решения по ядерным функциям имеют малые по модулю значения. Другим подходом в получении разреженных решений может быть разбиение имеющейся выборки на обучающую и тестовую части. При этом обучающая часть и составит выборку, по которой будут вычисляться решения. По отношению к полной выборке эти решения будут разреженными. Получаемая тестовая выборка может быть использована для вычисления на ней критериев качества решений, по которым можно вести настройку параметров алгоритма. Эти критерии, связанные с точностью прогноза на тестовой выборке, относятся к классу внешних критериев [10–12]. Формирование обучающей и тестовой части выборки можно проводить по процедуре планирования эксперимента [13–14].

Рассмотрим возможность получения разреженных решений посредством формирования обучающей части выборки с привлечением методов оптимального планирования эксперимента.

### Разреженное решение по методу LS-SVM

Предположим, у нас имеется обучающая выборка  $D_n = \{(x_k, y_k) : x_k \in X, y_k \in Y; k = 1, \dots, n\}$  объема  $n$  наблюдений вида:

$$y_k = m(x_k) + e_k, k = 1, \dots, n,$$

где ошибка  $e_k \in R$  имеет характеристики  $E[e_k | x = x_k] = 0$  и  $Var[e_k] = \sigma^2 < \infty$ ,  $m(x)$  – неизвестная действительная гладкая функция и  $E[y_k | x = x_k] = m(x_k)$ . Нас будет интересовать аппроксимация  $m(x)$  в виде  $f(x) = \omega^T \varphi(x) + b$ . Предположим, что исходная выборка наблюдений  $W$  разбита на обучающую и тестовую части, и обозначим их через  $A$  и  $B$ . Соответственно, через

$n_A$  обозначим объем обучающей выборки, а через  $n_B$  – объем тестовой выборки. Координаты точек обучающей и тестовой выборок будем обозначать через  $x$  и  $z$ . Для получения решения по методу LS-SVM составляем СЛАУ вида:

$$\begin{bmatrix} 0 & 1_{n_A}^T \\ 1_{n_A} & \Omega_A + \frac{1}{\gamma} I_{n_A} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{\alpha}_A \end{bmatrix} = \begin{bmatrix} 0 \\ y_A \end{bmatrix}, \quad (1)$$

где  $y_A = (y_1, \dots, y_{n_A})^T$  – точки из обучающей выборки,  $1_{n_A} = (1, \dots, 1)^T$ ,  $\hat{\alpha}_A = (\hat{\alpha}_1, \dots, \hat{\alpha}_{n_A})$  и  $\Omega_{kl} = \varphi(x_k)^T \varphi(x_l)$ ,  $k, l = 1, \dots, n_A$ ;  $\gamma$  – параметр регуляризации.

Решив СЛАУ вида (1), получим результирующую разреженную LS-SVM модель, имеющую вид:

$$\hat{y}(x) = \sum_{k=1}^{n_A} \hat{\alpha}_k K(x, x_k) + \hat{b},$$

где  $K(x, x_k)$  – ядро скалярного произведения,

$$\hat{b} = \frac{1_{n_A}^T \left( \Omega_A + \frac{1}{\gamma} I_{n_A} \right)^{-1} y_A}{1_{n_A}^T \left( \Omega_A + \frac{1}{\gamma} I_{n_A} \right)^{-1} 1_{n_A}},$$

$$\hat{\alpha}_A = \left( \Omega_A + \frac{1}{\gamma} I_{n_A} \right)^{-1} (y_A - 1_{n_A} \hat{b}).$$

### Выбор обучающей части выборки для метода LS-SVM с помощью $D$ -оптимального планирования эксперимента

Для построения дискретного  $D$ -оптимального плана будем использовать описанные в работе [15] алгоритмы.

Через  $G_s$  обозначим матрицу, имеющую размер  $s \times s$  для обучающей выборки объемом в  $s$  наблюдений, который состоит из элементов  $(G_s)_{ij} = K(x_i, x_j) + \frac{1}{\gamma} I_s$ ,  $i, j = 1, \dots, s$ .

Матрица  $G_{s+1}$  на  $s+1$  шаге принимает вид:

$$G_{s+1} = \begin{pmatrix} G_s & F(x_{s+1}) \\ F^T(x_{s+1}) & K(x_{s+1}, x_{s+1}) + \frac{1}{\gamma} \end{pmatrix},$$

где  $F^T(x_{s+1}) = (K(x_1, x_{s+1}), K(x_2, x_{s+1}), \dots, K(x_s, x_{s+1}))$ .

Для вычисления определителя окаймленной матрицы воспользуемся формулой  $|G_{s+1}| = |G_s| * \Delta(x_{s+1})$ , где  $\Delta(x_{s+1}) = [K(x_{s+1}, x_{s+1}) + \frac{1}{\gamma} - F^T(x_{s+1}) G_s^{-1} F(x_{s+1})]$ .

Таким образом, поиск очередной точки, которую будем включать в обучающую вы-

борку, выполняем по следующей схеме:  $x_{s+1} = \mathop{\text{Arg max}}_x \Delta(x)$ . Подробно данный алгоритм разбиения выборки на части приведен в работе [14].

В алгоритме используется матрица  $G_s$ , элементы которой зависят от параметров используемых ядерных функций. Очевидно, что перед проведением процедуры  $D$ -оптимального разбиения, эти параметры необходимо как-то оценить. Можно пойти по традиционному пути, используя критерий контроля по отдельным объектам (Leave-One-Out, LOO).

Важно также, чтобы получаемое разреженное решение не имело какого-либо смещения. В вычислительном эксперименте удобным критерием контроля точности получаемых моделей является среднеквадратичная ошибка предсказания MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^n (u_i - \hat{y}_i)^2,$$

где  $u_i, \hat{y}_i, i=1, \dots, n$  соответственно незашумленное и оцененное по модели значение отклика.

### Получение обучающей части выборки с помощью критерия регулярности

Предлагаемый выше метод получения разреженного решения предполагает использование обучающей выборки, свойства которой обеспечивают получение на ней решения, обеспечивающего наименьшую дисперсию предсказания на тестовой части выборки. Можно пойти по-другому пути, формируя обучающую выборку через минимизацию ошибки прогноза на тестовой выборке. Ошибка прогноза на тестовой выборке в литературе еще известна как критерий регулярности

$$\Delta^2 = (y_B - \hat{y}_B)^T (y_B - \hat{y}_B) / n_B,$$

где  $\hat{y}_B = \Phi_B \hat{\alpha}_A + \hat{b}_A$  – прогнозные значения по модели, оцененной на обучающей выборке. Критерий регулярности может быть также задействован в рассмотренном выше алгоритме разбиения выборки по процедуре  $D$ -оптимального планирования эксперимента на заключительном этапе при уточнении внутренних параметров алгоритма LS-SVM.

Для получения разреженного решения на основе обучающей части выборки и критерия регулярности рассмотрим следующие алгоритмы.

#### Вариант включения:

1. Предварительное оценивание параметра ядерной функции с использованием критерия LOO.
2. Выполняем последовательную схему наращивания объема обучающей выборки до необходимого объема. Добавляемая в обучающую часть точка выбирается по минимуму критерия  $\Delta^2$ . Целесообразно периодически перенастраивать внутренние параметры LS-SVM.

#### Вариант исключения:

1. Предварительное оценивание параметра ядерной функции с использованием критерия LOO.
2. Выполняем последовательную схему наращивания объема тестовой выборки. Добавляемая в тестовую часть точка выбирается по минимуму критерия  $\Delta^2$ . Оставшиеся точки образуют обучающую выборку. Целесообразно периодически перенастраивать внутренние параметры LS-SVM. Новое значение параметров LS-SVM выбираем в том случае, если удастся при нем улучшить значение критерия  $\Delta^2$ .

#### Вариант замены:

1. Предварительное оценивание параметра ядерной функции с использованием критерия LOO.
2. Выполняем  $D$ -оптимальное разбиение исходной выборки на обучающую и тестовую части. Далее выполняем поиск оптимальных значений внутренних параметров LS-SVM

по критерию  $\Delta^2$ .

3. Оптимизация состава обучающей выборки. На этом шаге точки из обучающей выборки поочередно заменяем точками тестовой выборки. При этом вычисляем критерий регулярности. В случае если замена привела к улучшению данного критерия, оставляем ее в силе, если удачных замен нет – останов.

4. Выполняем поиск оптимальных значений внутренних параметров LS-SVM по критерию  $\Delta^2$ , обучаясь на выборке, полученной на шаге 3.

Данный алгоритм позволяет уточнить состав точек обучающей выборки, которая была получена по процедуре *D*-оптимального планирования эксперимента.

### Вычислительный эксперимент

Вычислительный эксперимент проводился с целью сравнения качества разреженных решений, получаемых по различным алгоритмам.

Тестовая функция вида:  $m(x) = 7 / e^{(x+0.75)^2} + 3x$ , заданная на отрезке  $[-1; 1]$ , была использована для проведения исследований. В качестве ядерной функции использовалось Гауссово ядро. Нормально распределенные величины были использованы в качестве помехи. Уровень помехи задавался в пределах от 5 % до 25 % от мощности полезного сигнала. Число наблюдений задавалось равным 10, 20, 30 и 50. Параметр регуляризации в LS-SVM фиксировался на уровне 10. Лучшее решение подбиралось по параметру масштаба Гауссового ядра на интервале от  $10^{-5}$  до  $10^0$  по сетке с шагом 0,1.

В таблице приведены значения среднеквадратичной ошибки. В строках «без разбиения» указаны значения MSE для неразреженных решений, полученных на полных выборках. Настройка параметров ядерных функций в этом случае проводилась по критерию LOO.

Таблица

Среднеквадратичная ошибка предсказания при 20 %-м уровне шума

Объем выборки	Вариант разбиения	Размер тестовой части выборки (%)										
		5	10	15	20	25	30	35	40	45	50	
N = 10	без разбиения	0,0367	0,0367	0,0367	0,0367	0,0367	0,0367	0,0367	0,0367	0,0367	0,0367	0,0367
	D-опт. план	0,0944	0,0944	0,0944	0,0944	0,0685	0,0685	0,0192	0,0192	0,0193	0,0193	0,0193
	замена	0,0944	0,0944	0,0944	0,0944	0,0685	0,0685	0,0198	0,0198	0,0298	0,0298	0,0298
	исключение	0,3155	0,3155	0,0944	0,0944	0,0685	0,0685	0,0192	0,0192	0,0193	0,0193	0,0193
	включение	0,3155	0,3155	0,0944	0,0944	0,0685	0,0685	0,0258	0,0258	0,0193	0,0193	0,0193
N = 20	без разбиения	0,0079	0,0079	0,0079	0,0079	0,0079	0,0079	0,0079	0,0079	0,0079	0,0079	0,0079
	D-опт. план	0,0182	0,0277	0,0277	0,0099	0,0138	0,0079	0,0099	0,0119	0,0116	0,0148	0,0148
	замена	0,0170	0,0170	0,0277	0,0182	0,0182	0,0138	0,0182	0,0148	0,0182	0,0182	0,0182
	исключение	0,0110	0,0277	0,0277	0,0099	0,0138	0,0079	0,0099	0,0119	0,0116	0,0148	0,0148
	включение	0,0182	0,0182	0,0182	0,0182	0,0182	0,0182	0,0182	0,0182	0,0182	0,0138	0,0148
N = 30	без разбиения	0,0021	0,0021	0,0021	0,0021	0,0021	0,0021	0,0021	0,0021	0,0021	0,0021	0,0021
	D-опт. план	0,0034	0,0157	0,0024	0,0020	0,0020	0,0020	0,0042	0,0034	0,0034	0,0034	0,0034
	замена	0,0034	0,0157	0,0157	0,0020	0,0020	0,0020	0,0042	0,0042	0,0042	0,0042	0,0042
	исключение	0,0034	0,0157	0,0024	0,0020	0,0020	0,0020	0,0042	0,0034	0,0034	0,0034	0,0034
	включение	0,0068	0,0068	0,0068	0,0068	0,0068	0,0068	0,0068	0,0053	0,0053	0,0053	0,0053
N = 50	без разбиения	0,0021	0,0021	0,0021	0,0021	0,0021	0,0021	0,0021	0,0021	0,0021	0,0021	0,0021
	D-опт. план	0,0028	0,0039	0,0023	0,0039	0,0051	0,0051	0,0068	0,0021	0,0051	0,0021	0,0021
	замена	0,0031	0,0039	0,0100	0,0039	0,0051	0,0068	0,0022	0,0022	0,0022	0,0022	0,0022
	исключение	0,0028	0,0039	0,0023	0,0039	0,0051	0,0051	0,0068	0,0021	0,0051	0,0021	0,0021
	включение	0,0020	0,0022	0,0020	0,0022	0,0022	0,0022	0,0024	0,0023	0,0020	0,0020	0,0022

Можно сравнивать между собой неразрезанное решение и наполовину разреженное при 50 %-й тестовой части. Видно, что получаемые разреженные решения при  $D$ -оптимальном разбиении выборки лишь немногим проигрывают неразрезанному по величине MSE. При этом, если использовать вариант разбиения на основе критерия регулярности, то улучшения качества решения с позиции MSE чаще всего не наблюдается. Это позволяет говорить о том, что для получения разреженного решения можно использовать обучающую выборку, полученную с использованием  $D$ -оптимального разбиения ее на части. Вместе с тем включение дополнительной оптимизации состава точек (см. алгоритм «замена») позволяет в ряде случаев получать лучшие решения.

**Заключение.** В работе для получения разреженного решения на основе метода LS-SVM предложены различные способы получения обучающей части выборки. По результатам проведенных вычислительных экспериментов для получения разреженных решений можно рекомендовать использование обучающей выборки, которая может быть получена, в том числе по процедуре  $D$ -оптимального планирования эксперимента. При этом окончательную настройку внутренних параметров метода LS-SVM можно осуществлять, ориентируясь на величину ошибки на тестовой части выборки.

### Литература

1. Suykens J. A. K., Van Gestel T., De Brabanter J., De Moor B., Vandewalle J. Least Square Support Vector Machines. New Jersey ; London ; Singapore ; Hong Kong : World Scientific, 2002. 290 p.
2. Vapnik V. Statistical Learning Theory. New York : John Wiley, 1998. 736 p.
3. Cherkassky V., Ma Y. Practical selection of SVM parameters and noise estimation for SVM regression // *Neural Networks*. 2004. № 17. P. 113–126.
4. Попов А. А., Саутин А. С. Определение параметров алгоритма опорных векторов при решении задачи построения регрессии // *Сб. науч. тр. НГТУ (Новосибирск)*. 2008. № 2 (52). С. 35–40.
5. Popov A. A., Sautin A. S. Selection of support vector machines parameters for regression using nested grids // *The Third International Forum on Strategic Technology (Novosibirsk)*, 2008. P. 329–331.
6. Mao W., Mu X., Zheng Y., Yan G. Leave-one-out cross-validation-based model selection for multi-input multi-output support vector machine // *Neural Computing and Application*. 2014. № 2 (24). P. 441–451.
7. Rivas-Perea P., Cota-Ruiz J., Rosiles J.-G.. A nonlinear least squares quasi-Newton strategy for LP SVR hyper-parameters selection // *International Journal of Machine Learning and Cybernetics*. 2014. № 4 (5). P. 579–597.
8. Gupta A. K., Guntuku S. C., Desu R. K., Balu A. Optimisation of turning parameters by integrating genetic algorithm with support vector regression and artificial neural networks // *The International Journal of Advanced Manufacturing Technology*. 2015. № 1–4 (77). P. 331–339.
9. Suykens A. K., De Brabanter J., Lukas L., Vandewalle J. Weighted least squares support vector machines: robustness and sparse approximation // *Neurocomputing*. 2002. Vol. 48. P. 85–105.
10. Степашко В. С., Кочерга Ю. Л. Методы и критерии решения задач структурной идентификации // *Автоматика*. 1985. № 5. С. 29–37.
11. Сарычев А. П. Усредненный критерий регулярности метода группового учета аргументов в задаче поиска наилучшей регрессии // *Автоматика*. 1990. № 5. С. 28–33.
12. Степашко В. С. Асимптотические свойства внешних критериев выбора моделей // *Автоматика*. 1988. № 6. С. 75–82.
13. Попов А. А. Разбиение выборки для внешних критериев селекции моделей с использованием методов планирования эксперимента // *Заводская лаборатория*. 1997. № 1. С. 49–53.

14. Попов А. А., Бобоев Ш. А. Получение тестовой выборки в методе LS-SVM с использованием оптимального планирования эксперимента // Науч. вестн. Новосиб. гос. технич. ун-та. 2016. № 4. С. 80–99.

15. Попов А. А. Последовательные схемы построения оптимальных планов эксперимента // Сб. науч. тр. НГТУ (Новосибирск). 1995. Вып. 1. С. 39–44.