2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)
Mexico City, Mexico November 10-12, 2021

A Generalized Lagrange Multiplier Method for Support Vector Regression

Sara Rodríguez-Reyes, Pablo Benavides-Herrera, Gregorio Álvarez, Riemann Ruiz-Cruz, and Juan Diego Sánchez-Torres*
Department of Mathematics and Physics
ITESO, Tlaquepaque, Jalisco, México
Email: dsanchez@iteso.mx*

***Abstract*—This paper presents an approach to support vector regression that extends the classic Vapnik's formulation. After recalling that the classic formulation contains a Lasso regularization structure in its dual form, we propose an generalized Lagrangian function with additional terms to include the Ridge regularization in the dual problem. In this form, by including both regularization methods, the resulting dual problem with the generalized Lagrangian comprises an elastic net regularization structure. Hence, as an immediate consequence, the classical formulation turns out to be a particular case of the current proposal. Finally, to demonstrate the capabilities of this approach, the document includes examples of predicting some benchmark problems.**

***Index Terms*—Kernel-based methods, Support vector regression, Extended Lagrangian, Convex optimization**

I. Introduction

The support vector regression (SVR) has shown to be a powerful method for proposing empirical models for predicting continuous variables [1]–[5]. The interpretability, the formulation as a convex optimization problem [6], the use of kernels [7], [8], and its relationships to other models make the SVR a robust and reliable method for several industrial and research problems.

A well-known fact about the classic formulation of SVR is that it exhibits a Lasso regularization [9] in its dual optimization problem [4], [10]. This event coincides with the existence of Lagrange multipliers equal to zero and the appearance of support values and vectors. Besides, the support vector methods and the Lasso regularization present substantial equivalences [11]. On the other hand, the simultaneous use of two different regularization schemes provides desirable models' characteristics [12]–[14]. A remarkable case of this approach is the elastic net, where the Ridge regularization [15] works together with the Lasso [16]. Moreover, similarly to the previous case, the support vector models with two regularizations present important equivalences to the elastic net regularization [17].

This paper proposes a new SVR by introducing a Ridge regularization term in the dual through the definition of a generalized Lagrangian function. In this form, the current proposal considers the advantages of the simultaneous use of two different regularization structures while keeping the formality with the generalized Lagrangian approach.

Section II details the SVR based on the extended Lagrangian. Section III presents a case study implementation for this approach, and a comparison with other benchmark algorithms. Last, Section IV provides conclusions for this paper.

II. SVR Based on an Generalized Lagrangian

A. Classical Support Vector Regression

For the case of SVR, let the set $D = (x_1, y_1), \dots, (x_N, y_N)$, where $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}$. Let $\varphi : X \to \mathcal{F}$ be the function that makes each input point x correspond a point in the feature space $\mathcal{F}$, where $\mathcal{F}$ is a Hilbert space. This feature space can be of high dimension or even infinite. However, is common to define $X = \mathbb{R}^n$ and $\mathcal{F} = \mathbb{R}^m$. In this form, the approximating function, namely the model, has the form $\hat{y}_k = f(x_k) = w^T \varphi(x_k) + b$ with $w \in \mathbb{R}^m$ and $b \in \mathbb{R}$.

The following problem statement considers such a regression problem as a convex optimization problem.

$$\begin{aligned} \min_{w,b,\xi,\xi^*} \mathcal{P}_\epsilon(w, b, \xi, \xi^*) &= \frac{1}{2} w^T w + C \sum_{k=1}^{N} \left(\xi_k^p + \xi_k^{*p}\right) \\ \text{s.t. } & y_k - w^T \varphi(x_k) - b \le \epsilon + \xi_k, \quad k = 1, \dots, N \\ & w^T \varphi(x_k) + b - y_k \le \epsilon + \xi_k^*, \quad k = 1, \dots, N \\ & \xi_k, \xi_k^* \ge 0, \quad k = 1, \dots, N \end{aligned} \tag{1}$$

where $\varphi(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$ and the regularization parameter $c > 0$ determines the balance between the regularity of f and the quantity up to which we tolerate deviations more significant than ϵ. Consider ξ_k and ξ_k^* as slack variables that control the error between the prediction $\hat{y}_k$ and the k-th sample y_k. The number p is either 1 or 2. If $p = 1$, the support vector regressor is called *L_1 soft-margin supoport vector regressor (L_1 SVR)* and $p = 2$, the *L_2 soft-margin support vector regressor (L_2 SVR)* [4].

Remark 1. *For the present work, only the case L_1 will be considered since it can be easily proven that for the aim of this paper, the L_2 provides an equivalent result.*

Theorem 1. *The primal problem* (1) *with the Lagrangian* $\mathcal{L}(w, b, \xi_k, \xi_k^*; \alpha_k, \alpha_k^*, \eta_k, \eta_k^*) = \frac{1}{2} w^T w +$

978-1-6654-0029-9/21/$31.00 ©2021 IEEE

$C\sum_{k=1}^{N}\left(\xi_k+\xi_k^*\right)-\sum_{k=1}^{N}\alpha_k\left(\xi_k-y_k+w^T\ \varphi(x_k)+b\right)-\sum_{i=k}^{N}\alpha_k^*\left(\xi_k^*+y_k-w^T\varphi(x_k)-b\right)-\sum_{k=1}^{N}\eta_k\xi_k-\sum_{i=k}^{N}\eta_k^*\xi_k^*$, *with* $\alpha_k,\alpha_k^*,\eta_k,\eta_k^*\geq 0$ *results in the following dual problem:*

$$\begin{aligned}\max_{\alpha_k,\alpha_k^*}\mathcal{D}\left(\alpha,\alpha^*\right)=&-\frac{1}{2}\sum_{k,l=1}^{N}(\alpha_k-\alpha_k^*)(\alpha_l-\alpha_l^*)\varphi^T(x_k)\varphi(x_l)\\&+\sum_{k=1}^{N}(\alpha_k-\alpha_k^*)y_k-\epsilon\sum_{k=1}^{N}\left(\alpha_k+\alpha_k^*\right)\\ \textit{s.t.}\ &\sum_{k=1}^{N}\left(\alpha_k-\alpha_k^*\right)=0\\&\alpha_k,\alpha_k^*\in\left[0,C\right],k=1,\ldots,N\end{aligned}\tag{2}$$

Proof. See [3]–[5]. □

Defining $\beta_k=\alpha_k-\alpha_k^*$. Then, $\beta_k\in[-C,C]$ Similarly, defining $|\beta_k|=\alpha_k+\alpha_k^*$, where $|\beta_k|\in[0,C]$. Reformulating the dual problem in terms of β_k in a matrix form:

$$\begin{aligned}\max_{\beta}\mathcal{D}\left(\beta\right)=&-\frac{1}{2}\beta^TK\beta+y^T\beta-\epsilon\|\beta\|_1\\ \text{s.t. }&\beta^T1_v=0\\&|\beta|\preceq C\end{aligned}\tag{3}$$

Remark 2. *The equation* (3) *the connection between the LASSO and the SVR due to the appearance of a term with the* L_1 *norm [4], [10].*

B. A GLMM for the L_1^{ϵ}*-SVR*

With the aim of proposing a new type of ϵ-SVR, consider the primal problem (1) with the following Lagrangian based on the generalized Lagrange multiplier method (GLMM) [18]:

$$\begin{aligned}\mathcal{L}(w,b,\xi_k,\xi_k^*;\alpha_k,\alpha_k^*,\eta_k,\eta_k^*)=&\frac{1}{2}w^Tw+C\sum_{k=1}^{N}\left(\xi_k+\xi_k^*\right)\\&-\sum_{k=1}^{N}\alpha_k\left(\xi_k-y_k+w^T\ \varphi(x_k)+b\right)\\&-\sum_{i=k}^{N}\alpha_k^*\left(\xi_k^*+y_k-w^T\varphi(x_k)-b\right)\\&-\sum_{k=1}^{N}\eta_k\xi_k-\sum_{i=k}^{N}\eta_k^*\xi_k^*\\&-\lambda\left[(1-\epsilon)\sum_{k=1}^{N}\left(\alpha_k+\alpha_k^*\right)+\frac{\epsilon}{2}\sum_{k=1}^{N}\left(\alpha_k+\alpha_k^*\right)^2\right]\end{aligned}\tag{4}$$

Proposition 1. *The function* (4) *fulfills all the conditions of the GLMM [18].*

Proof. The proof follows directly from the definition, see [18]. □

Theorem 2. *The primal problem* (1) *with the Lagrangian* (4) *leads to the following dual problem:*

$$\begin{aligned}\max_{\alpha_k,\alpha_k^*}&\mathcal{D}\left(\alpha,\alpha^*\right)=\\&-\frac{1}{2}\sum_{k,l=1}^{N}(\alpha_k-\alpha_k^*)(\alpha_l-\alpha_l^*)\varphi^T(x_k)\varphi(x_l)\\&+\sum_{k=1}^{N}(\alpha_k-\alpha_k^*)y_k\\&-\lambda\left[(1-\epsilon)\sum_{k=1}^{N}\left(\alpha_k+\alpha_k^*\right)+\frac{\epsilon}{2}\sum_{k=1}^{N}\left(\alpha_k+\alpha_k^*\right)^2\right]\\ \textit{s.t.}\ &\sum_{k=1}^{N}\left(\alpha_k-\alpha_k^*\right)=0\\&\alpha_k,\alpha_k^*\in\left[0,C\right],k=1,\ldots,N.\end{aligned}$$

Proof. The proof follows from the stationary conditions, as follows:

- The first order condition on the parameter w, $\nabla_w\mathcal{L}(w,b,\xi_k,\xi_k^*;\alpha_k,\alpha_k^*,\eta_k,\eta_k^*)=0$, implies $w=\sum_{k=1}^{N}(\alpha_k-\alpha_k^*)\varphi(x_k)$.
- The first order condition on the parameter b, $\frac{\partial}{\partial b}\mathcal{L}(w,b,\xi_k,\xi_k^*;\alpha_k,\alpha_k^*,\eta_k,\eta_k^*)=0$, implies $\sum_{k=1}^{N}(\alpha_k-\alpha_k^*)=0$.
- The first order condition on the parameter ξ_k, $\frac{\partial}{\partial\xi_k}\mathcal{L}(w,b,\xi_k,\xi_k^*;\alpha_k,\alpha_k^*,\eta_k,\eta_k^*)=0$, implies $\alpha_k+\eta_k=C$
- The first order condition on the parameter ξ_k^*, $\frac{\partial}{\partial\xi_k^*}\mathcal{L}(w,b,\xi_k,\xi_k^*;\alpha_k,\alpha_k^*,\eta_k,\eta_k^*)=0$, implies $\alpha_k^*+\eta_k^*=C$

Then, replacing this critical points in the Lagrangian (4). □

The optimal solution must satisfy the Karush Kuhn Tucker (KKT) complementary slackness conditions:

$$\begin{aligned}&\alpha_k\left(\epsilon+\xi_k-y_k+w^T\varphi(x_k)+b\right)=0\\&\alpha_k^*\left(\epsilon+\xi_k^*+y_k-w^T\varphi(x_k)-b\right)=0\\&\eta_k\xi_k=(C-\alpha_k)\xi_k=0\\&\eta_k^*\xi_k^*=(C-\alpha_k^*)\xi_k^*=0\end{aligned}$$

Hence, using the complementary slackness conditions, it follows the calculation of b:

$$\begin{aligned}b=&y_k-w^T\varphi(x_k)-\epsilon,\ \text{such that}\\&\alpha_k\in(0,C)\end{aligned}$$

Finally, defining $\beta_k=\alpha_k-\alpha_k^*$. Then, $\beta_k\in[-C,C]$ Similarly, defining $|\beta_k|=\alpha_k+\alpha_k^*$, where $|\beta_k|\in[0,C]$. Reformulating the dual problem in terms of β_k in a matrix form:

$$\max_{\beta} \mathcal{D}(\beta) = -\frac{1}{2}\beta^T K \beta + y^T \beta$$
$$- \lambda \left[(1-\epsilon)\|\beta\|_1 + \frac{\epsilon}{2}\|\beta\|_2^2 \right] \quad (5)$$
$$\text{s.t. } \beta^T 1_v = 0$$
$$|\beta| \preceq C$$

Remark 3. *It is showed in* (5) *the connection between the LASSO, the Ridge and the* L_1^ϵ*-SVR due to the appearance of a term with the* L_1 *norm and a squared term with the* L_2 *norm. This is enough to show that the* L_1^ϵ*-SVR is in nature a LASSO problem. This new proposal of* ϵ*-SVR based on the* L_1^ϵ*-SVR offers a new structure that proposes an Elastic net regularization keeping the box constraints where* $0 \leq \alpha_k, \alpha_k^* \leq C$ *which makes easier to calculate the* b *parameter.*

Remark 4. *In the dual problem* (5)*, if* $\epsilon = 0$ *and* $\lambda > 0$*, the original formulation* (2) *is recovered. The implies that the solution of* (2) *is a lower bound of the solution of* (5) *i.e., when tuning the hyper-parameters, the worst case scenario for* (5) *is* (2)*.*

III. Implementation

In order to test the performance of the model proposed in this work, two known data sets are used to solve a regression problem. The Extended-Lagrangian-SVR Python implementation is based on the β reformulation in (5). The ϵ, C, γ and λ free hyper-parameters were tuned using the Bayesian optimization algorithm [19] and a Radial Based Kernel.

A. Boston House Dataset

The well known Housing dataset is used, which contains features about different houses in Boston. This dataset can be accessed from the *scikit-learn* library [20]. The objective is to predict the value of prices of the house.

There are 506 samples and 13 feature variables in this dataset, which the respective description can be consulted in [20].

B. Model Evaluation

The model proposed in this work is implemented in Python and is compared against the most popular models that have been used to model this data set. The modeling conditions were based on the compilation of results published on the Kaggle site [21]. The models that were implemented are the multiple linear regression, the random forest, XGBoost, Support Vector Regression (SVR) and the SVR based on Extended Lagrangian proposed by the authors [22]. The partitioning of the data into a training an test datasets was performed using a ratio of 70% and 30%, respectively. Additionally, a model is trained using the symbolic transformer method to perform a feature engineering process before to training [23]. This last model has the purpose of testing if there is an improvement if the performance of the Extended Lagrangian SVR model.

The comparison between the models was made using the performance metrics R^2, MSE and MAE, where the summary of the results obtained is presented in Table I. As can be seen in the table, the proposed model manages to improve MAE compared to the commonly applied models.

Model/Metric	C	γ	ε	λ	R^2	MSE	MAE
Linear Regression					0.7121	30.0539	3.8590
Random Forest					0.8414	16.5543	2.4284
XGBoost					0.8765	12.8934	2.2637
SVR	74.8883	0.06098	0.1569		0.8936	11.1064	2.0167
Extended-SVR	180.9238	0.0692	0.2106	0.101813	0.9141	8.9681	1.9812
Extended-SVR Symb. Transf.	1806.272	0.0317	0.0595	0.0949	0.9313	7.1714	1.9499

TABLE I: Metric Performance Comparison for Boston Dataset

Three performance measures were used for performance evaluation:

- R^2: indicates the model's performance, rather than the loss in absolute terms.
- *Mean Squared Error (MSE):* calculates the squared difference between actual and predicted values.
- *Mean Absolute Error (MAE):* calculates the absolute difference between actual and predicted values.

MAE is a good metric to evaluate model performance against different models, it is a more natural measure of average error. Since dimensioned evaluations and inter-comparisons of average model-performance error is the goal, this metric will be the main focus. The Extended Lagrangian SVR with the Symbolic Transformer [23] is the best-suited model for this example as is has the lowest MAE metric.

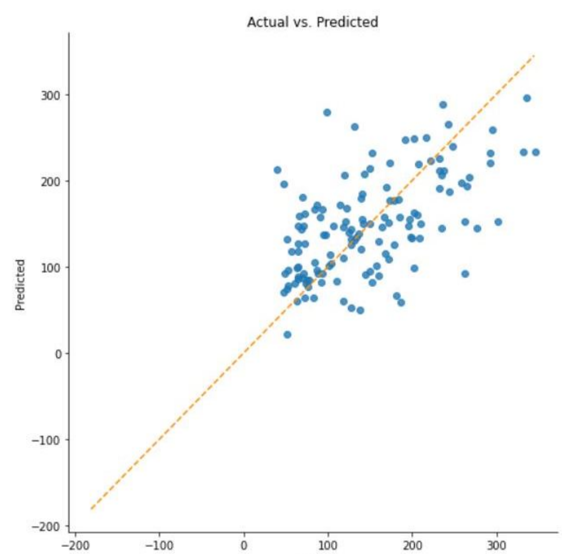

Fig. 1: Visualizing the differences between actual prices and predicted values.

In Fig. 1, there is a fairly strong correlation between the model's predictions and its actual results, it is observed a relatively even spread around the diagonal line. The R^2 metric indicates that the extended Lagrangian SVR using the Symbolic Transformer fits better than the other models compared in Table I.

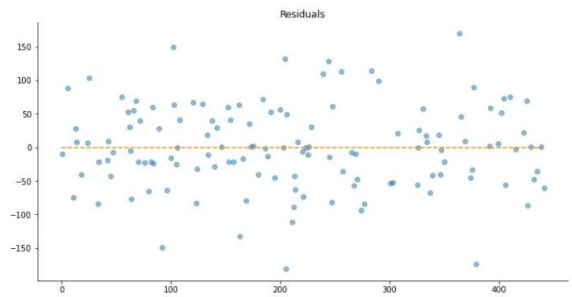

Fig. 2: Checking residuals

In Fig. 2, it can be seen that the residual plot shows a fairly random pattern, meaning there is the same variance within the error terms.

Residuals are normally distributed. Using the Anderson-Darling test for normal distribution, given a p-value of 0.414969, the conclusion is that residuals are normally distributed.

Finally, performing a Durbin-Watson test to determine if either positive or negative correlation is present. Obtaining a value of 2.1429, the conclusion is that there is little to no auto-correlation in the residuals. Meaning that all patterns are explained by the model.

All metrics still indicate that the extended Lagrangian SVR model provides a decent fit to the data.

C. Diabetes Dataset

The Diabetes dataset is used in this example, which contains features about different patients. This dataset can be accessed from the *scikit-learn* library [24]. The objective is to predict a quantitative measure of disease progression one year after baseline.

There are 442 samples and 10 feature variables in this dataset, which the respective description can be consulted in [24].

D. Model Evaluation

The models that were implemented are the multiple linear regression, the random forest, XGBoost, Support Vector Regression (SVR) and the SVR based on Extended Lagrangian proposed by the authors [22]. The partitioning of the data into a training an test datasets was performed using a ratio of 70% and 30%, respectively. Additionally, a model is trained using the symbolic transformer method to perform a feature engineering process before to training [23]. This last model has the purpose of testing if there is an improvement if the performance of the Extended Lagrangian SVR model.

The comparison between the models was made using the performance metrics R^2, MSE and MAE, where the summary of the results obtained is presented in Table II. As can be seen in the table, the proposed model manages to improve MAE compared to the commonly applied models.

Model/Metric	C	γ	ε	λ	R^2	MSE	MAE
Linear Regression					0.4463	2963.368	43.6898
Random Forest					0.4423	2985.062	44.0939
XGBoost					0.4668	2853.892	43.2759
SVR	21.8337	0.0600	1.6738		0.4359	3019.506	43.5409
Extended-SVR	352.0067	0.0143	0.5195	0.8947	0.4544	2920.115	43.1549
Extended-SVR Symb. Transf.	289.1950	0.0054	0.3057	1.5645	0.4669	2853.587	41.8195

TABLE II: Metric Performance Comparison for Diabetes Dataset

The Extended Lagrangian SVR with the Symbolic Transformer [23] is the best-suited model for this example as is has the lowest MAE metric.

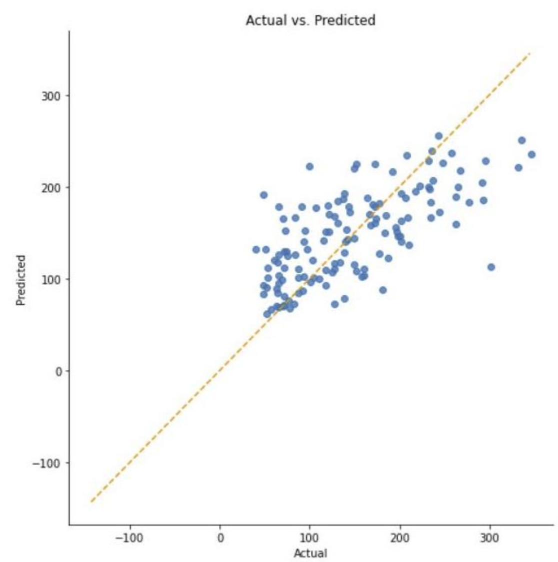

Fig. 3: Visualizing the differences between actual prices and predicted values.

In Fig. 3, there is a fairly strong correlation between the model's predictions and its actual results, it is observed a relatively even spread around the diagonal line. The R^2 metric indicates that the extended Lagrangian SVR using the Symbolic Transformer fits better than the other models compared in Table II.

In Fig. 4, it can be seen that the residual plot shows a fairly random pattern, meaning there is the same variance within the error terms.

Residuals are normally distributed. Using the Anderson-Darling test for normal distribution, given a p-value of

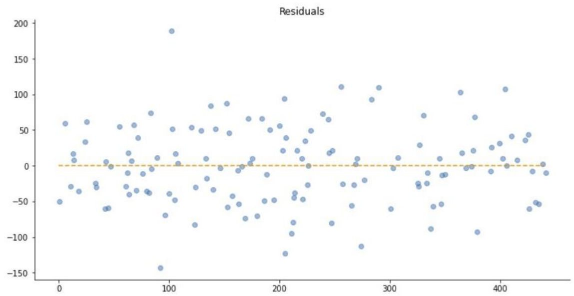

Fig. 4: Checking residuals

0.663172, the conclusion is that residuals are normally distributed.

Finally, performing a Durbin-Watson test to determine if either positive or negative correlation is present. Obtaining a value of 2.0868, the conclusion is that there is from little to no auto-correlation in the residuals. Meaning that all patterns are explained by the model.

All metrics still indicate that the extended Lagrangian SVR model provides a suitable fit to the data.

IV. Conclusions

This paper has proposed a generalized Lagrange multiplier method and derived generalized KKT conditions for support vector regression, which includes a weighted elastic net regularization structure. It was shown that the extended Lagrange SVR models outperform the classic SVR models in predicting the prices of houses in Boston and predicting quantitative measure of disease (Diabetes) progression one year after baseline.

A disadvantage of this new model proposal would be the increasing time of the optimization for the new hyper-parameter λ; if the hyper-parameter search is exhaustive (e.g., Grid Search), the process will suffer the curse of dimensionality; the number of times required to evaluate the model during hyper-parameter optimization grows exponentially in the number of parameters. This is the reason for using intelligent methods (Bayesian Optimization) in this paper, where better performance models can be obtained since this method requires fewer model evaluations and has lower variance because it intelligently searches the parameter space.

On the other side, the advantage of this new model proposal is that this new elastic net structure leads to good results. It gives the possibility to reduce the number of support vectors used to create the model, through the L_1 norm. Also, the L_2 helps the eigenvalues to be positive, making the problem strictly convex, and giving some stability in the case that some support vectors are correlated.

References

[1] V. N. Vapnik, *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.

[2] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998. [Online]. Available: https://doi.org/10.1023/A:1009715923555

[3] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. World Scientific, 2002. [Online]. Available: https://www.worldscientific.com/worldscibooks/10.1142/5089

[4] S. Abe, *Support Vector Machines for Pattern Classification*, 2nd ed. Springer, 2004.

[5] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004. [Online]. Available: https://doi.org/10.1023/B:STCO.0000035301.49549.88

[6] S. Boyd and L.Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[7] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2001.

[8] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[9] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996. [Online]. Available: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02080.x

[10] X. Han and L. Clemmensen, "On weighted support vector regression," *Quality and Reliability Engineering International*, vol. 30, no. 6, pp. 891–903, 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.1654

[11] M. Jaggi, "An equivalence between the Lasso and support vector machines," *Arxiv*, vol. abs/1303.1152, 2013. [Online]. Available: http://arxiv.org/abs/1303.1152

[12] A. Y. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proceedings of the Twenty-First International Conference on Machine Learning*, ser. ICML '04. Association for Computing Machinery, 2004.

[13] L. Wang, J. Zhu, and H. Zou, "The doubly regularized support vector machine," *Statistica Sinica*, vol. 16, no. 2, pp. 589–615, 2006.

[14] J. López, S. Maldonado, and M. Carrasco, "Double regularization methods for robust feature selection and svm classification via dc programming," *Information Sciences*, vol. 429, pp. 377–389, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025517310976

[15] A. N. Tikhonov, "On the solution of ill-posed problems and the method of regularization," *Dokl. Akad. Nauk SSSR*, vol. 151, no. 3, pp. 501–504, 1963.

[16] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005. [Online]. Available: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2005.00503.x

[17] Q. Zhou, W. Chen, S. Song, J. Gardner, K. Weinberger, and Y. Chen, "A reduction of the elastic net to support vector machines with an application to GPU computing," 2015. [Online]. Available: https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9856

[18] M. Li, "Generalized Lagrange multiplier method and KKT conditions with an application to distributed optimization," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 2, pp. 252–256, 2019.

[19] F. Nogueira, "Bayesian Optimization: Open source constrained global optimization tool for Python," 2014. [Online]. Available: https://github.com/fmfn/BayesianOptimization

[20] scikit-learn developers, "sklearn.datasets.load_boston," 2020. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_boston.html

[21] S. Chaudhary, "Boston house price prediction," March 2019. [Online]. Available: https://www.kaggle.com/shreayan98c/boston-house-price-prediction

[22] G. Álvarez, "SVR-Extended," https://github.com/Gegori1/SVR-Extended, 2021.

[23] T. Stephens, "Api reference gplearn." [Online]. Available: https://gplearn.readthedocs.io/en/stable/reference.html#symbolic-transformer

[24] scikit-learn developers, "sklearn.datasets.load_diabetes," 2020. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_diabetes.html