Relevance Vector Machine for regression and comparaison with the SVR

Hadrien Hendrikx
School of Engineering (STI)
École Polytechnique
Fédérale de Lausanne
Lausanne, Switzerland
hadrien.hendrix@epfl.ch

Gregoire Gallois-Montbrun
School of Engineering (STI)
École Polytechnique
Fédérale de Lausanne
Lausanne, Switzerland
gregoire.gallois-montbrun@epfl.ch

Louis Faury
School of Engineering (STI)
École Polytechnique
Fédérale de Lausanne
Lausanne, Switzerland
louis.faury@epfl.ch

***Abstract*—Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.**

I. Introduction

II. Theoretical background

A. Support Vector machine for Regression

The Support Vector machine for Regression (SVR) extends the SVM method for regression tasks. Let us first consider the simple linear regression case for a dataset $\{X, \mathbf{t}\} = \{(x_1, \dots, x_N)^T, (t_1, \dots, t_n)^T\}$, where we minimize a regularized error function given by :

$$E(w) = \frac{1}{2}\sum_{n=1}^{N}\left\{w^T\phi(x_n) - t_n\right\}^2 + \frac{\lambda}{2}\|w\|^2 \tag{1}$$

To obtain sparse solution, the quadratic term is replaced by a *ε-insensitive error function* (see [1]) denoted $E_\varepsilon(\cdot)$ with :

$$E_\varepsilon(y(x) - t) = \begin{cases} 0, & \text{if } |y(x) - t| < \varepsilon \\ |y(x) - t| < \varepsilon & \text{otherwise} \end{cases} \tag{2}$$

Therefore, the quantity to be minimized can be expressed as :

$$C\sum_{n=1}^{N} E_\varepsilon(w^T\phi(x_n) - t) + \frac{1}{2}\|w\|^2 \tag{3}$$

where C is a regularization parameter.
As for the SVM, one can introduce *slack variables* in order to transform this optimization program into a quadratic programming problem (quadratic objective, linear constraints). Once the problem solved, predictions are made using :

$$y(x) = \sum_{n=1}^{N}(a_n - \hat{a}_n)k(x, x_n) + b \tag{4}$$

where we introduced the kernel $k(x, x') = \phi(x)^T\phi(x')$. The coefficients $\{a_n\}$ and $\{\hat{a}_n\}$ actually are Lagrange multipliers for the QP problem, and provide a *sparse* solutions in the datapoints. The *support vectors* (data point used for predictions) are those for which $a_n \neq 0$ or $\hat{a}_n \neq 0$, in other words those that lie on the boundary or outside of the ε-tube defined by the loss function in equation (2).
As for the SVM, onecan adopt a $\nu - SVR$ formulation to have an lower-bound control on the number of retained support vectors.

B. Relevance Vector machine for Regression

III. Results

A. Datasets presentation

B. Results

IV. Discussion

V. Conclusion

References

[1] Vladimir N Vapnik. The nature of statistical learning theory. 1995.