Regression Based on Binary Classification Using Support Vector Machines

Marcin Orchel

AGH University of Science and Technology
in Poland

Support Vector Classification and Regression

- two the most popular problems in machine learning: classification and regression
- for both problems there were developed methods in the framework of Support Vector Machines (SVM): Support Vector Classification (SVC) and ε-insensitive Support Vector Regression (ε-SVR)
- similarities between SVC and ε-SVR:
  - quadratic optimization problems
  - they lead to sparse solutions
  - they originate from statistical learning theory
- for ϵ-SVR we have the additional parameter ε

ε-Support Vector Regression Idea

- find a function for which all examples fall between ε bounds

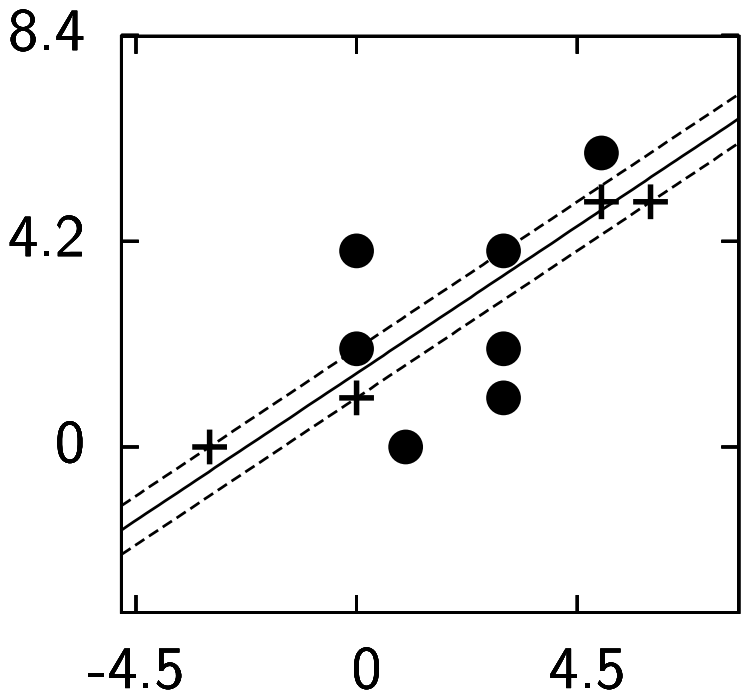

Figure: The idea of ε-SVR. In the figure, there are examples, support vectors (circles), a solution (solid line), and ε boundaries (dashed lines)

The idea of Regression Based on Binary Classification

- transformation of the data: regression examples are duplicated and transformed in the way that original examples are translated up and duplicated examples are translated down by a value of $\delta \geq 0$

The idea of Regression Based on Binary Classification

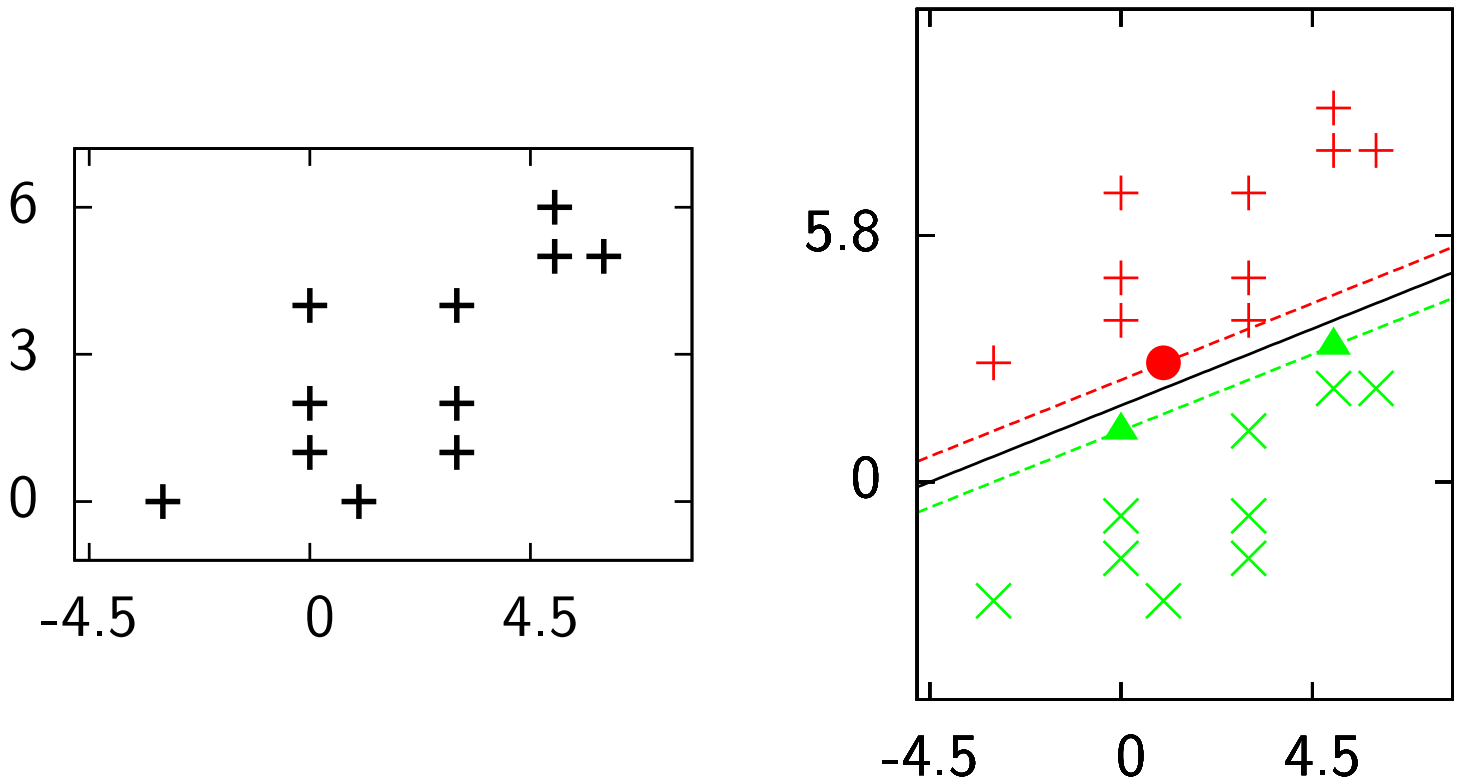

Figure: The idea of the transformation of the problem in δ-SVR for 2d. In the left figure, there are regression example points. In the right figure, there are classification example points after transformation, support vectors (triangles and circles), solution (a solid line), margin lines (dashed lines)

The idea of Regression Based on Binary Classification

1. transformation of the problem: the output y of the regression problem is incorporated to the feature set as the additional feature
2. a classification problem is solved with the new data setting
3. the solution of the classification problem is transformed into the regression function $y = \ldots$

Transforming the Curve into the Function

- SVC can return nonlinear solutions using the kernel trick
- nonlinear classifier can lead to the solution that cannot be transformed into the function (more than one y value for some values of the remaining features)
- popular kernels: polynomial, sigmoid, radial basis functions (RBF)
- we propose a slightly modification of these kernels by skipping the last variable, which is summed separately, e.g.

$$\left(\sum_{i=1}^{m+1} x_i y_i\right)^d \quad \rightarrow \quad \left(\sum_{i=1}^{m} x_i y_i\right)^d + x_{m+1} y_{m+1}$$

Details of Regression Based on Binary Classification

- How to choose the best value of φ?
- we use a simple search method
- we have found that it is better to compare regression performance on training data than classification performance while choosing φ
- for symmetrical, unimodal distributions of regression data, the transformed classification problem has the same optimal solution as the original regression problem for any φ

Advantages

- we need only the SVC solver for classification and regression problems
- modifications of SVC original problem such as adding a priori knowledge could be directly used for regression problems
- for example we can create improved reduced models for regression problems by using SVC with margin knowledge per example

Results

- test performed on synthetic data with added Gaussian noise and on real world data sets
- with φ we can control the number of support vectors
- results show that regression based on binary classification is able to achieve comparable to the ε-SVR or better generalization performance on unseen data

Conclusions

- the new regression is an alternative for ϵ-SVR
- it is based on SVC, therefore it is easier to use some modifications of SVC like incorporating a priori knowledge directly for regression problems