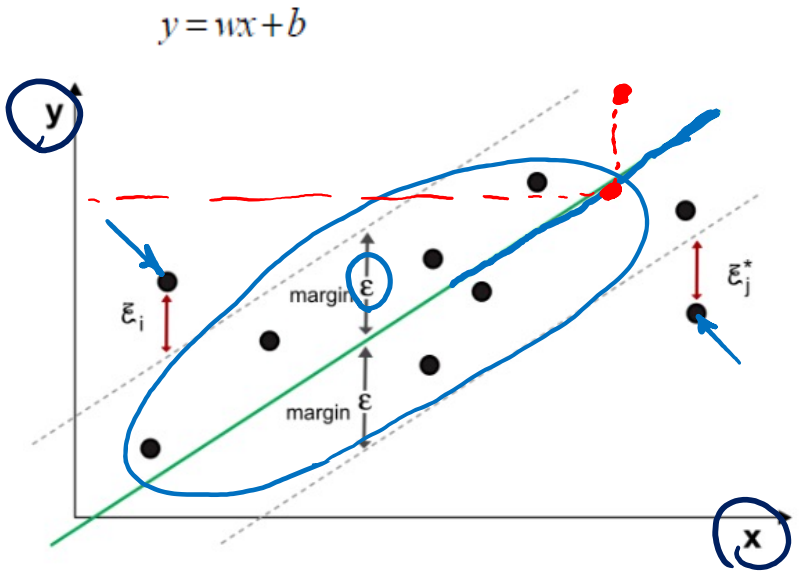
Part 26- Support Vector Machines Regression

Prof. Pedram Jahangiry



Topics

Part 23

- SVM Geometry
- SVM Motivation

Part 24

- Maximum Margin Classifier (MMC)
- Support Vector Classifiers (SVC)

Part 25

- Support Vector Machines (SVM)

Part 26

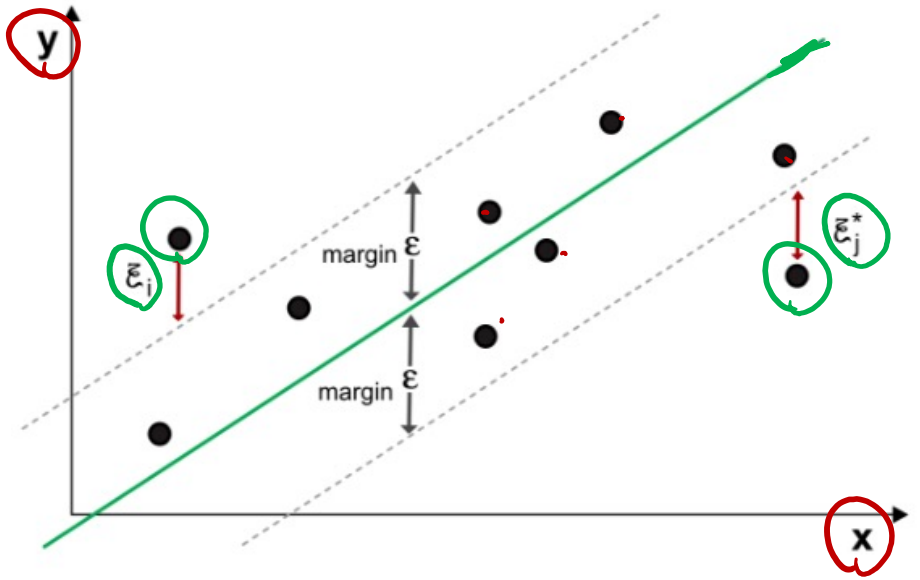
- Support Vector Regressors (SVR)

Part 27

- Multiple class classification
- SVM pros and cons
- SVM applications in Finance

SVM for regression (Support Vector Regressors)

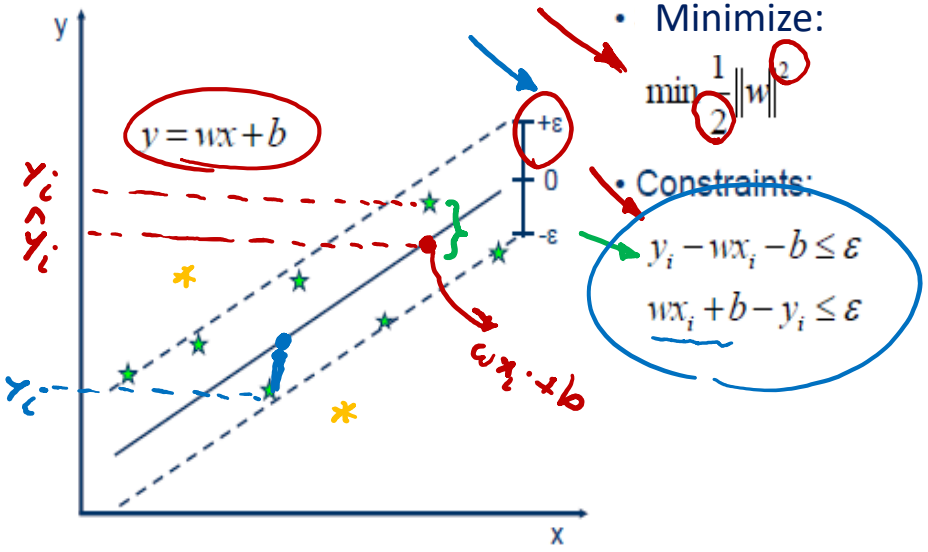
- The idea of SVM classification can be transposed to regression problems.
- However, the role of the margin is different. Our objective, is to basically find the hyperplane that holds maximum training observations within the margin ϵ (tolerance level).



SVR optimization

C ↑ More flexible var ↑ bias ↓

C ↓ bias ↑ var ↓



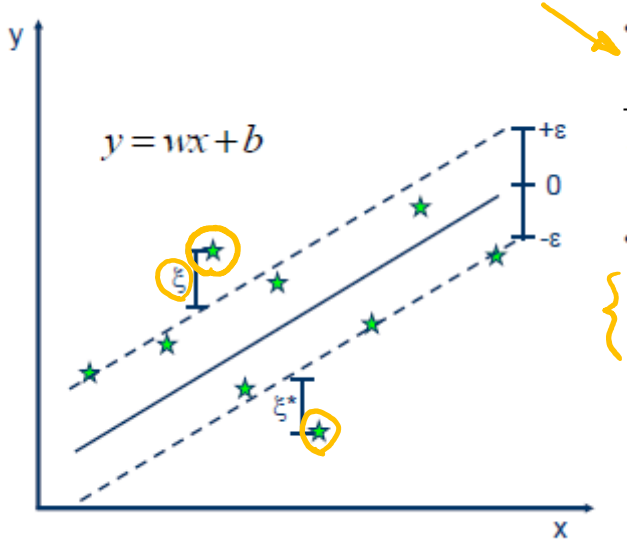
- Minimize:

$$\min \frac{1}{2}\|w\|^2$$

- Constraints:

$$y_i - wx_i - b \le \varepsilon$$

$$wx_i + b - y_i \le \varepsilon$$



- Minimize:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$

- Constraints:

$$y_i - wx_i - b \le \varepsilon + \xi_i$$

$$wx_i + b - y_i \le \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \ge 0$$

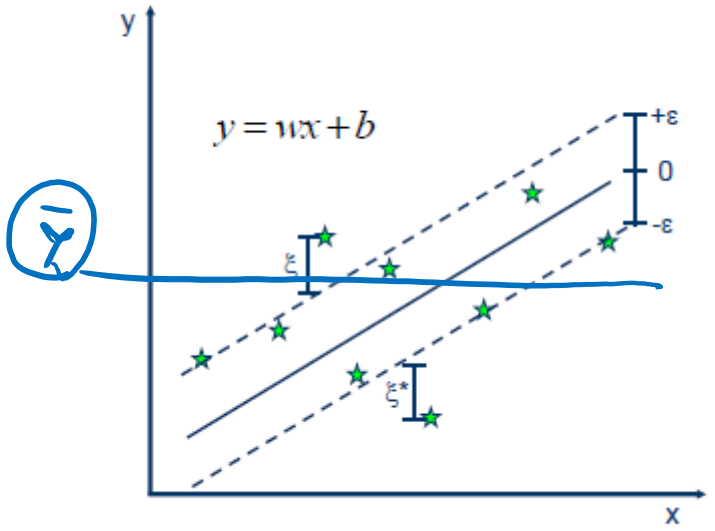
Source: https://www.saedsayad.com/support_vector_machine_reg.htm

SVR optimization

$C \propto \frac{1}{\text{Regularization}}$

$C = 0$ Max Reg

$C \uparrow$ Reg $\downarrow$ More flexible



- Minimize:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right)$$

- Constraints:

$$y_i - wx_i - b \le \varepsilon + \xi_i$$

$$wx_i + b - y_i \le \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \ge 0$$

Source: https://www.saedsayad.com/support_vector_machine_reg.htm

Kernel SVR optimization

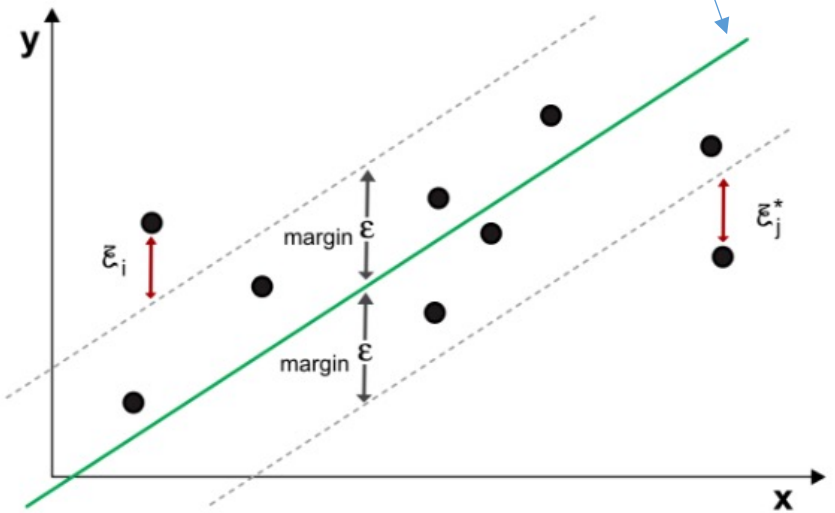
$$\underset{w,b}{\text{Min}} \ \frac{1}{2}||w||^2 + C\sum_{i=1}^{I}(\xi_i+\xi_i^*)$$

$$\left(\sum_{k=1}^{K} w_k\phi(x_{i,k}) + b\right) - y_i \le \epsilon + \xi_i^*$$

$$y_i - \left(\sum_{k=1}^{K} w_k\phi(x_{i,k}) + b\right) \le \epsilon + \xi_i$$

$$\xi_i, \xi_i^* \ge 0 \ \forall_i$$

Predictions: $\sum_{k=1}^{K} w_k\phi(x_{i,k}) + b$



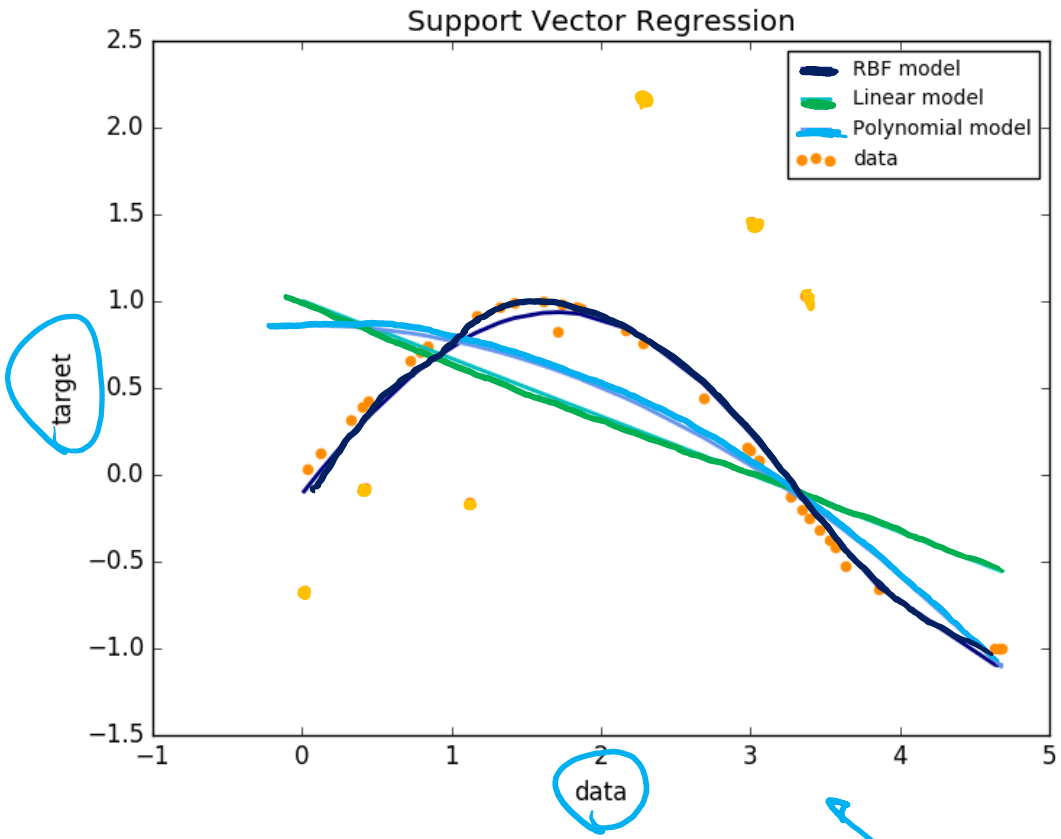
- Goal: minimize the sum of squared weights subject to the error being small enough
- This is somewhat the opposite of the penalized linear regressions which seek to minimize the error, subject to the weights being small enough

Min MSE

LASSO Min MSE s.t $|\sum|w_i|| < \alpha$

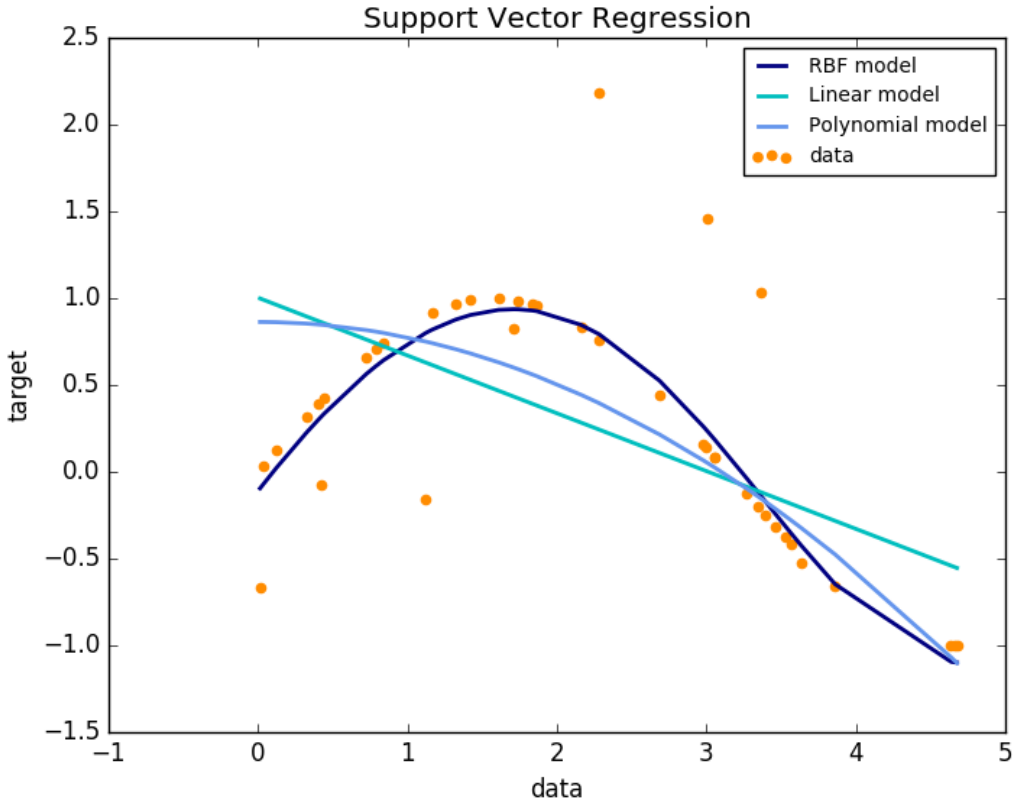
Ridge Min MSE s.t $\sum(w_i^2) < \beta$

SVR using Linear and Non-linear Kernels



Source: Scikit learn documentation

SVR using Linear and Non-linear Kernels



Source: Scikit learn documentation