Support Vector Regression

SV classification uses $y \in \{1, -1\}$, while regression tries to predict $y \in \Re$.

SV regression uses ϵ-insensitive loss:

$$E(y, z) = \begin{cases} 0 & \text{if } |y - z| \leq \epsilon \\ |y - z| - \epsilon & \text{otherwise} \end{cases}$$

or equivalently:

$$E(y, z) = \max(0, |y - z| - \epsilon)$$

One can think of SV regression as fitting a tube of radius ϵ to the data.

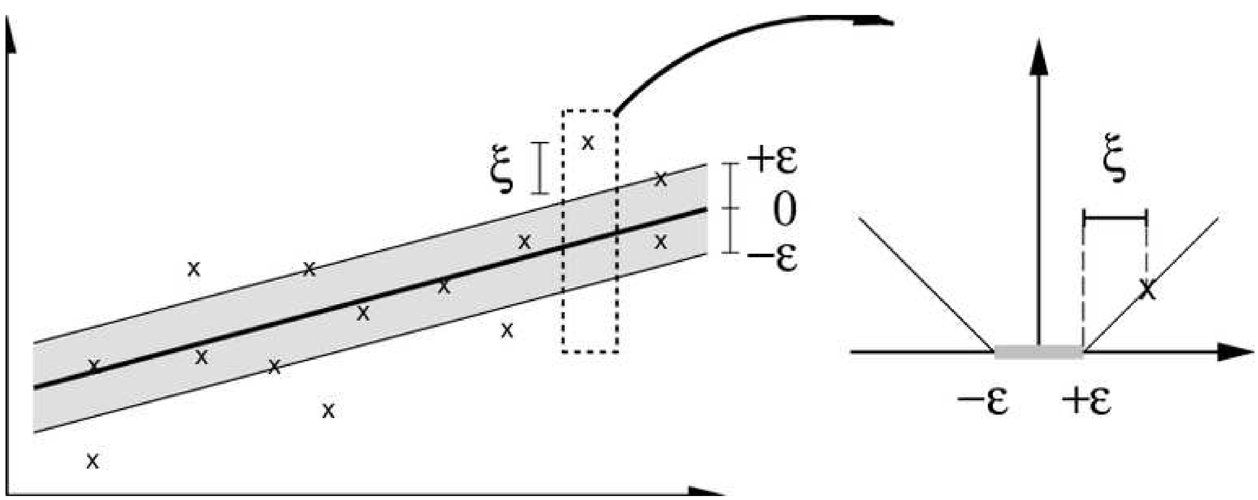

Figure is from B. Scholkopf and A. Smola, *Learning with Kernels*, MIT Press, 2002.

The optimal fit for $f_a(\mathbf{x})$ is found by solving:

$$\begin{aligned} \text{minimize}\ & \frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^{m} \max(0, |y - z| - \epsilon) \\ \text{subject to}\ & f_a(\mathbf{x}_i) - y_i \leq \epsilon + \xi_i \\ & y_i - f_a(\mathbf{x}_i) \leq \epsilon + \xi_i^* \\ & \xi_i \geq 0,\ \ \xi_i^* \geq 0 \end{aligned}$$

where ξ_i is the error if $f_a(\mathbf{x}_i)$ is too high and ξ_i^* is the error if $f_a(\mathbf{x}_i)$ is too low.

Applying the usual math tricks results in:

$$f_a(\mathbf{x}) = b + \sum_{i=1}^{m} \alpha_i k(\mathbf{x}, \mathbf{x}_i)$$

where the α_i weights are found by solving:

$$\begin{aligned} \text{maximize}\ & -\epsilon \sum_{i=1}^{m} |\alpha_i| + \sum_{i=1}^{m} \alpha_i y_i \\ & - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \end{aligned}$$

subject to $-C \leq \alpha_i \leq C$ and $\sum_{i=1}^{m} \alpha_i = 0$