Kernel Methods

Support Vector Regression

Motivation

Regression Model with Noise

- The regression task is modeled as

$$y_n = g(\mathbf{x}_n) + \eta_n, \quad n = 1, 2, \ldots, N$$

  where η_n represents i.i.d. noise.
- Optimization in the presence of outliers (heavy-tailed distributions)

Optimal Loss Functions for Regression with Noise

Huber Loss Function

Assuming a symmetric pdf for the noise, the optimal minmax strategy for regression:

$$L(y, f(\mathbf{x})) = \begin{cases} \epsilon|y - f(\mathbf{x})| - \frac{\epsilon^2}{2}, & \text{if } |y - f(\mathbf{x})| > \epsilon, \\ \frac{1}{2}|y - f(\mathbf{x})|^2, & \text{if } |y - f(\mathbf{x})| \leq \epsilon. \end{cases}$$

ϵ-Insensitive Loss Function

An alternative that enhances computational efficiency:

$$L(y, f(x)) = \begin{cases} |y - f(x)| - \epsilon, & \text{if } |y - f(x)| > \epsilon, \\ 0, & \text{if } |y - f(x)| \leq \epsilon. \end{cases}$$

ϵ-Insensitive loss promotes sparsity in the solution, making it particularly useful in support vector machines for regression.

Support Vector Regression I

Slack Variables

$$
\begin{aligned}
y_n - f(\mathbf{x}_n) &\leq \epsilon + \xi_n \\
-(y_n - f(\mathbf{x}_n)) &\leq \epsilon + \tilde{\xi}_n
\end{aligned}
$$

Support Vector Regression II

Regularized Minimization of Slack Variables

The optimization problem is formulated as a regularized minimization task:

$$\min_{\boldsymbol{\theta},\xi,\tilde{\xi}} \left(\frac{1}{2}\|\boldsymbol{\theta}\|^2 + C\left(\sum_{n=1}^{N}(\xi_n) + \sum_{n=1}^{N}(\tilde{\xi}_n) \right) \right)$$

subject to the constraints:

$$\begin{aligned} y_n - f(x_n) &\leq \epsilon + \xi_n, \\ -(y_n - f(x_n)) &\leq \epsilon + \tilde{\xi}_n, \\ \xi_n, \tilde{\xi}_n &\geq 0. \end{aligned}$$

Support Vector Regression III

Lagrangian Formulation

The optimization task is approached by introducing Lagrange multipliers. The optimal solution $\hat{\theta}$ is then expressed as:

$$\hat{\theta} = \sum_{n=1}^{N} (\tilde{\lambda}_n - \lambda_n) x_n$$

Support Vector Identification

Support vectors are identified as points where the error is at least ϵ. Points with errors less than ϵ have zero Lagrange multipliers and thus do not influence the solution.

Support Vector Regression IV

Solution in RKHS

Assume $f(\boldsymbol{x}) = \boldsymbol{\theta}^{\boldsymbol{T}}\mathbf{x} + \boldsymbol{\theta_0}$. For tasks solved in a Reproducing Kernel Hilbert Space (RKHS), replace inner product with the kernel function mappings $\kappa(\cdot, x_n)$:

$$\hat{\theta}(\cdot) = \sum_{n=1}^{N} (\tilde{\lambda}_n - \lambda_n)\kappa(\cdot, x_n)$$

Prediction

Given a new input x, the prediction $\hat{y}(x)$ is calculated as:

$$\hat{y}(x) = \sum_{n=1}^{N_s} (\tilde{\lambda}_n - \lambda_n)\kappa(x, x_n) + \hat{\theta}_0$$

where N_s is the number of support vectors, showing how ϵ-insensitive loss leads to sparsity in the model.

Support Vector Regression V

Objective Function

Dual representation form:

$$\arg\max_{\tilde{\lambda},\lambda} \left(\sum_{n=1}^{N} (\tilde{\lambda}_n - \lambda_n) y_n - \epsilon \sum_{n=1}^{N} (\tilde{\lambda}_n + \lambda_n) - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} (\tilde{\lambda}_n - \lambda_n)(\tilde{\lambda}_m - \lambda_m) \kappa(x_n, x_m) \right)$$

Constraints for each $n = 1, 2, \ldots, N$:

- $0 \leq \tilde{\lambda}_n \leq C$
- $0 \leq \lambda_n \leq C$
- $\sum_{n=1}^{N} \tilde{\lambda}_n = \sum_{n=1}^{N} \lambda_n$

13.2 Smoothing using kernel methods

```
gamma = 1/(np.square(kernel_params))
regressor = SVR(kernel='rbf', gamma=gamma, C=C, epsilon=epsilon)
regressor.fit(x_col,y_row)
y_pred = regressor.predict(t_col)
```

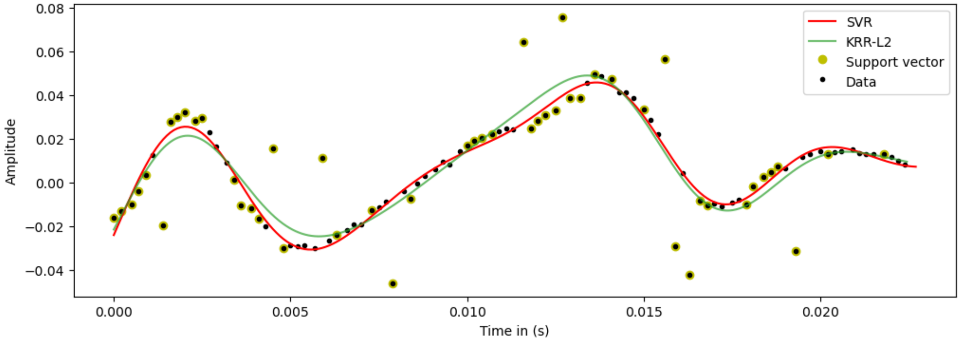

Figure: $\sigma = 0.004$, $C = 1e-2$