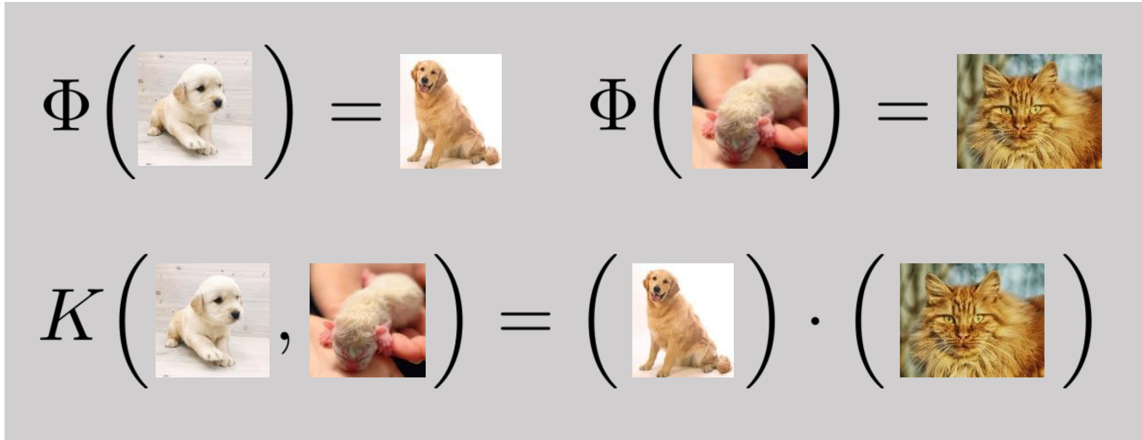
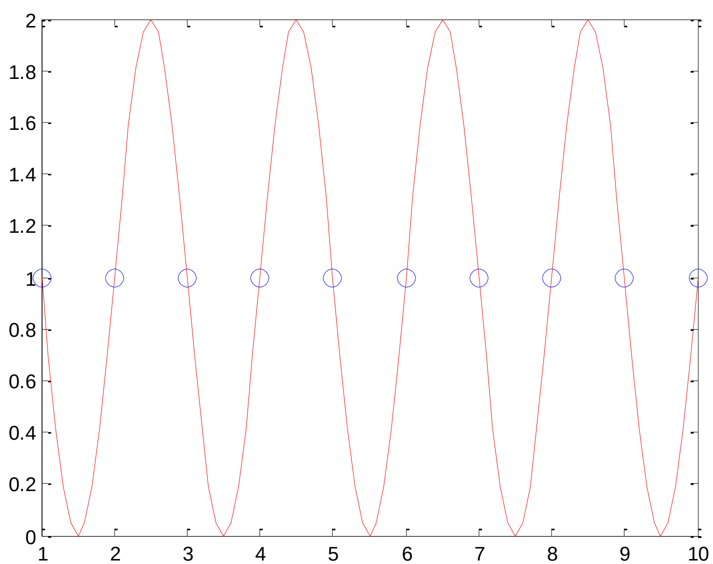


Kernel-based Learning:
Support Vector Regression

Pilsung Kang
School of Industrial Management Engineering
Korea University

Fitting Function

- Two objectives of function fitting
  - ✓ To fit a function, we minimize an error measure, called also loss function
  - ✓ We also like the function to be simple
    - Fewest basis functions
    - Simplest basis functions
    - Flatness is desirable

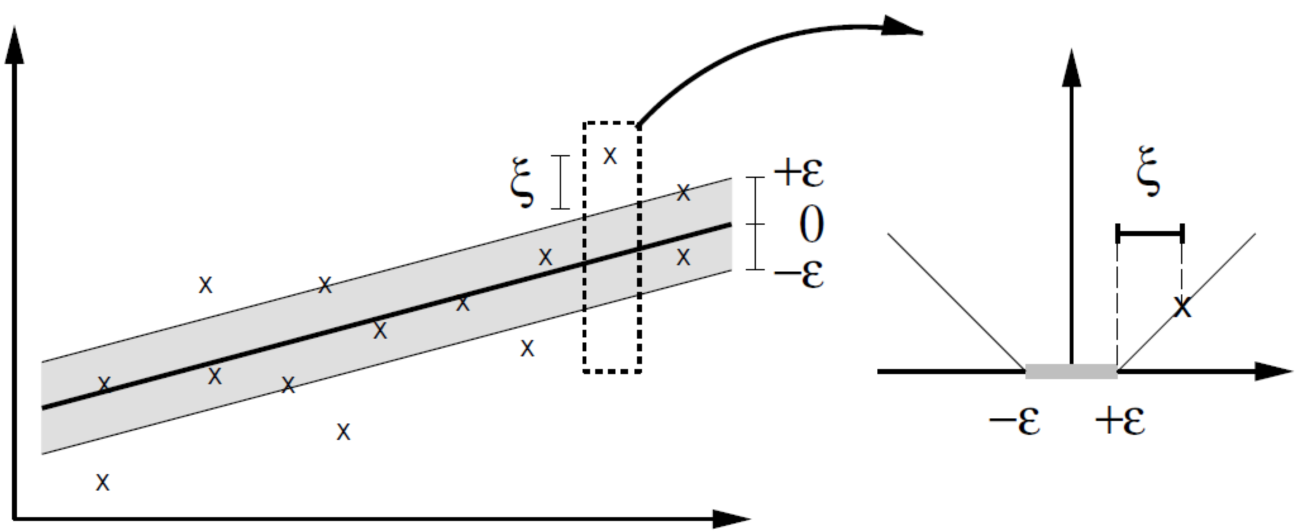


Why not draw a flat line rather than a sine function?

Support Vector Regression (SVR)

Smola and Scholkopf (2004)

- Combine loss function and flatness as a single objective
  - ✓ SVM was developed in the 1960s
  - ✓ Its extension to regression, i.e. SVR, is developed in 1997
- ε-SVR
  - ✓ Loss function



Support Vector Regression (SVR)

- SVR Formulation
  - ✓ Estimating a linear regression

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$$

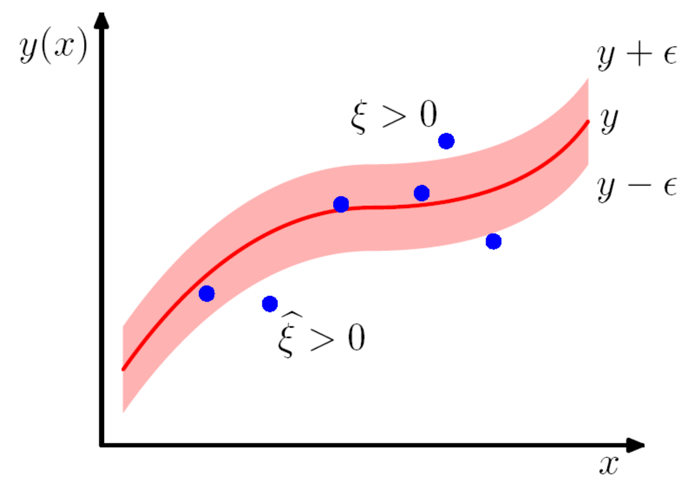
  - ✓ with precision ε by minimizing

$$\min \quad \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

$$s.t. \quad (\mathbf{w}^T\mathbf{x}_i + b) - y_i \leq \epsilon + \xi_i$$

$$y_i - (\mathbf{w}^T\mathbf{x}_i + b) \leq \epsilon + \xi_i^*$$

$$\xi_i,\ \xi_i^* \geq 0$$



Support Vector Regression (SVR)

- Primal Lagrangian

$$L_P = \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}(\eta_i\xi_i + \eta_i^*\xi_i^*)$$

$$- \sum_{i=1}^{n}\alpha_i(\epsilon + \xi_i + y_i - \mathbf{w}^T\mathbf{x}_i - b) - \sum_{i=1}^{n}\alpha_i^*(\epsilon + \xi_i^* - y_i + \mathbf{w}^T\mathbf{x}_i + b)$$

$$\alpha_i^{(*)}, \eta_i^{(*)} \geq 0$$

- Take the derivative w.r.t. b, w, ξ

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \qquad \frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)\mathbf{x}_i = 0$$

$$\frac{\partial L}{\partial \xi_i^{(*)}} = C - \alpha_i^{(*)} - \eta_i^{(*)}$$

Support Vector Regression (SVR)

- Dual Lagrangian Problem

$$L_D = -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\mathbf{x}_i^T\mathbf{x}_j - \epsilon\sum_{i,j=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i,j=1}^{n} y_i(\alpha_i^* - \alpha_i)$$

$$s.t.\ \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0, \quad \alpha_i, \alpha_i^* \in [0, C]$$

- Decision Function

$$\mathbf{w} = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)\mathbf{x}_i \quad \Rightarrow \quad f(\mathbf{x}) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)\mathbf{x}_i^T\mathbf{x} + b$$

Support Vector Regression (SVR)

- Dual Lagrangian Problem with Kernel Trick

$$L_D = -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(\mathbf{x}_i, \mathbf{x}_j) - \epsilon\sum_{i,j=1}^{n}(\alpha_i^* + \alpha_i) + \sum_{i,j=1}^{n} y_i(\alpha_i^* - \alpha_i)$$

$$s.t. \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0, \quad \alpha_i, \alpha_i^* \in [0, C]$$

- Decision Function

$$\mathbf{w} = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)\Phi(\mathbf{x}_i) \quad \Rightarrow \quad f(\mathbf{x}) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)K(\mathbf{x}_i, \mathbf{x}) + b$$

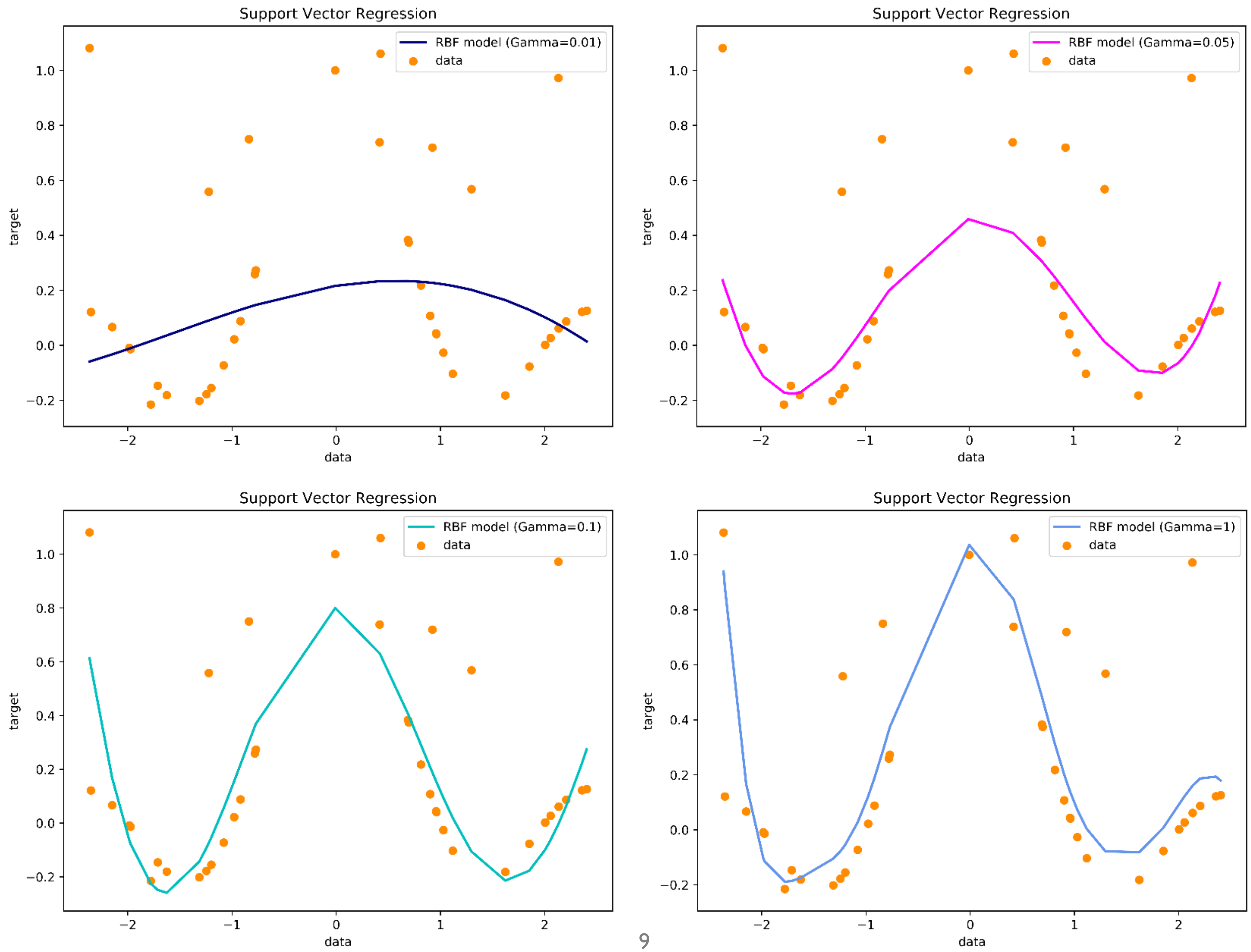
Support Vector Regression (SVR)

- Various Loss Functions for SVR

	loss function	density model
ε–insensitive	$c(\xi) = \lvert\xi\rvert_\varepsilon$	$p(\xi) = \frac{1}{2(1+\varepsilon)} \exp(-\lvert\xi\rvert_\varepsilon)$
Laplacian	$c(\xi) = \lvert\xi\rvert$	$p(\xi) = \frac{1}{2} \exp(-\lvert\xi\rvert)$
Gaussian	$c(\xi) = \frac{1}{2}\xi^2$	$p(\xi) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{\xi^2}{2})$
Huber's robust loss	$c(\xi) = \begin{cases} \frac{1}{2\sigma}(\xi)^2 & \text{if } \lvert\xi\rvert \le \sigma \\ \lvert\xi\rvert - \frac{\sigma}{2} & \text{otherwise} \end{cases}$	$p(\xi) \propto \begin{cases} \exp(-\frac{\xi^2}{2\sigma}) & \text{if } \lvert\xi\rvert \le \sigma \\ \exp(\frac{\sigma}{2} - \lvert\xi\rvert) & \text{otherwise} \end{cases}$
Polynomial	$c(\xi) = \frac{1}{p}\lvert\xi\rvert^p$	$p(\xi) = \frac{p}{2\Gamma(1/p)} \exp(-\lvert\xi\rvert^p)$
Piecewise polynomial	$c(\xi) = \begin{cases} \frac{1}{p\sigma^{p-1}}(\xi)^p & \text{if } \lvert\xi\rvert \le \sigma \\ \lvert\xi\rvert - \sigma\frac{p-1}{p} & \text{otherwise} \end{cases}$	$p(\xi) \propto \begin{cases} \exp(-\frac{\xi^p}{p\sigma^{p-1}}) & \text{if } \lvert\xi\rvert \le \sigma \\ \exp(\sigma\frac{p-1}{p} - \lvert\xi\rvert) & \text{otherwise} \end{cases}$

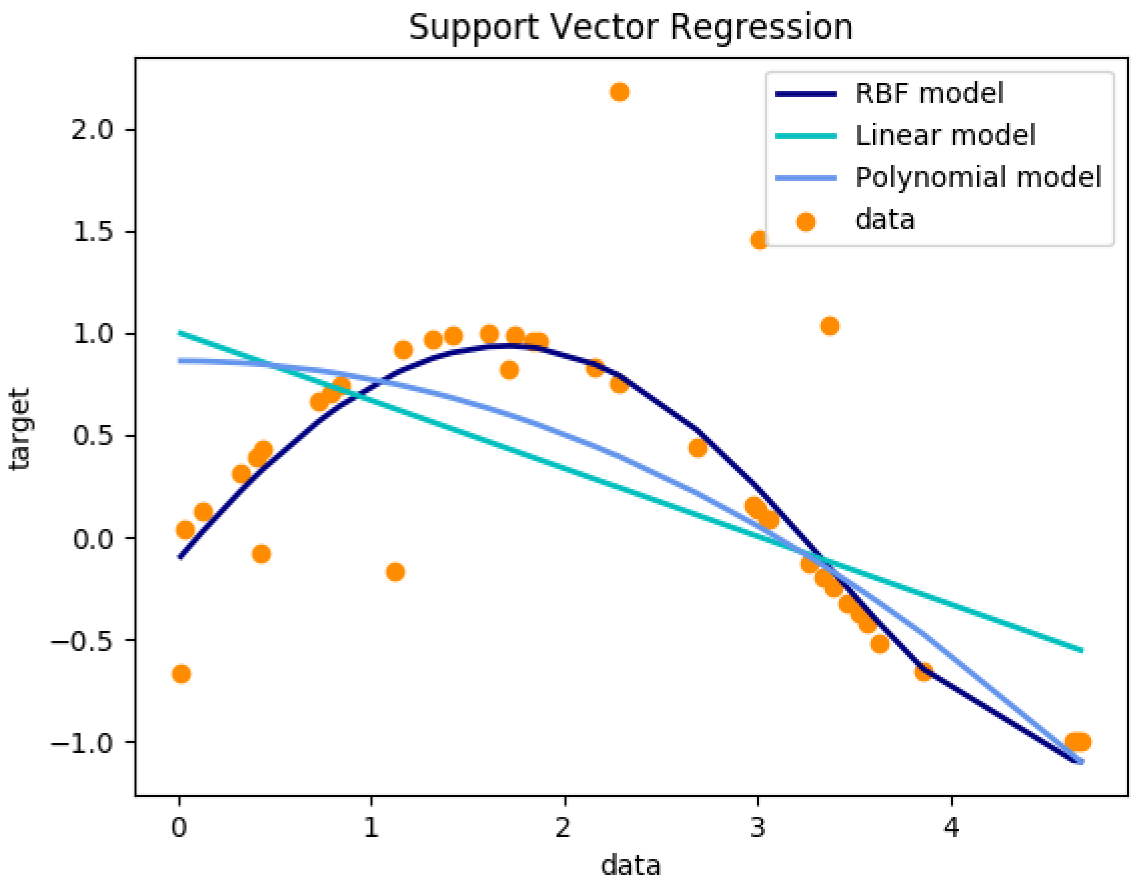
Support Vector Regression (SVR)

- Fitted functions with different ε



Support Vector Regression (SVR)

- Fitted function with different Kernel functions



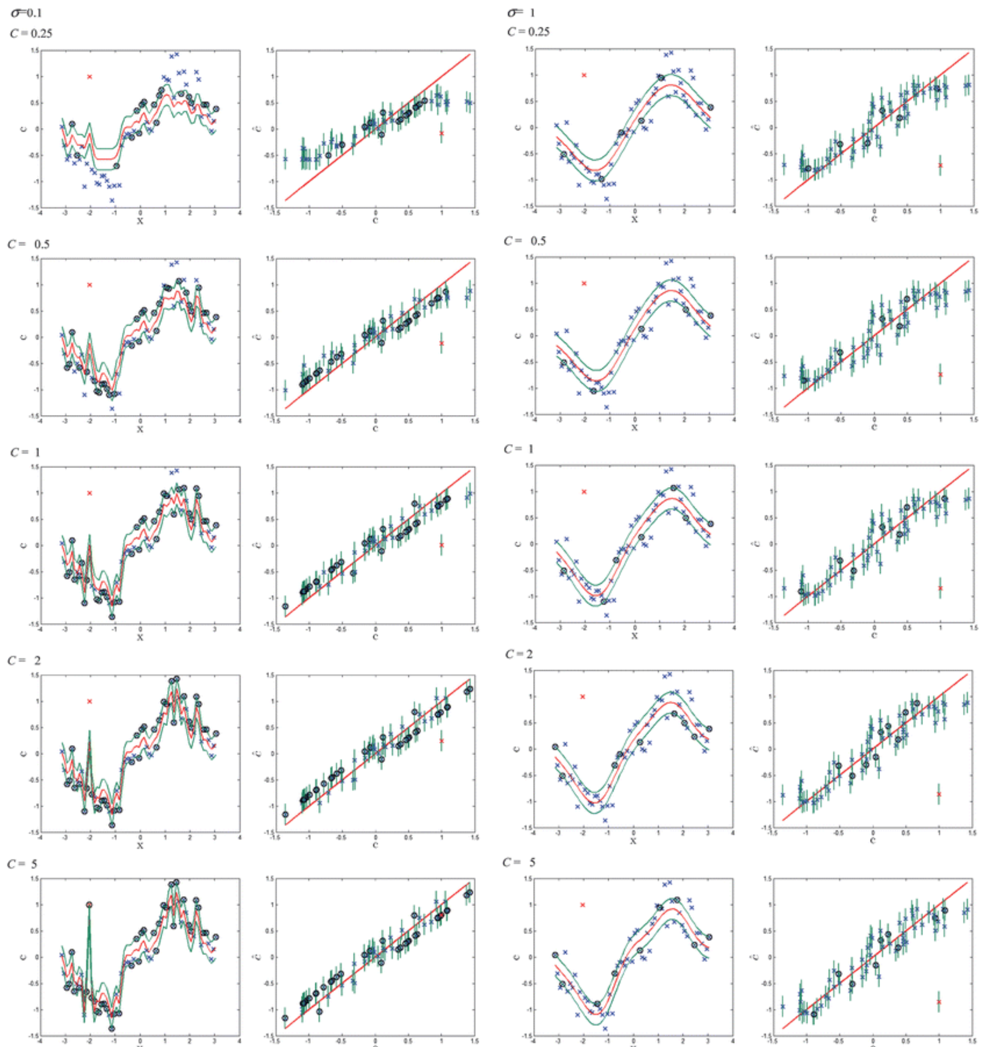
http://scikit-learn.org/stable/auto_examples/svm/plot_svm_regression.html#sphx-glr-auto-examples-svm-plot-svm-regression-py

고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics

Support Vector Regression (SVR)

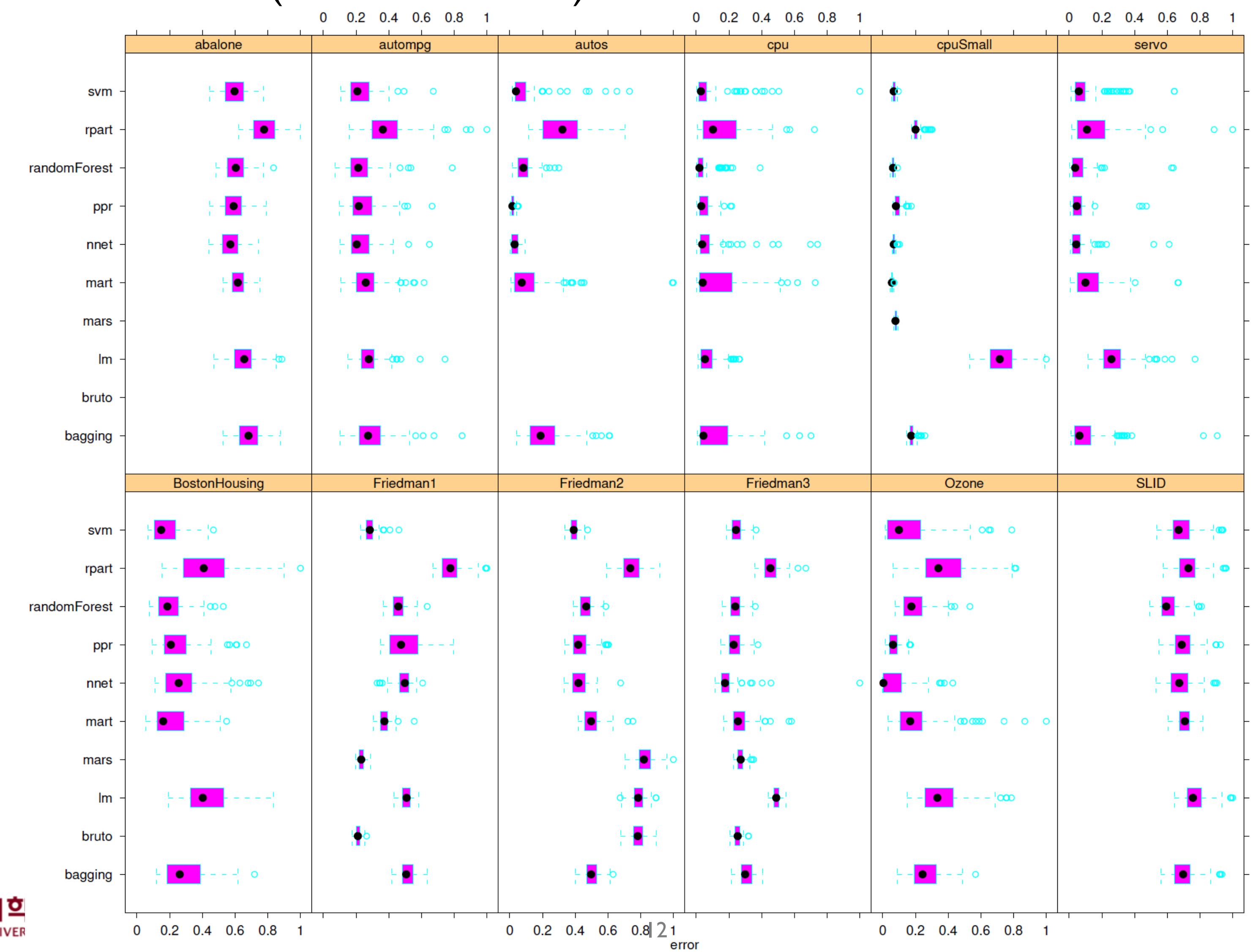
- SVR with different sigma and cost combinations



http://pubs.rsc.org/-/content/articlehtml/2010/an/b918972f

Support Vector Regression (SVR)

- SVR Performance (in terms of MSE)




Q & A

References

Research Papers

- Müller, K., Mika, S., Rätsch, G., Tsuda, K., and Schölkopf, B. (2001). An introduction to kernel-based learning algorithms. IEEE Transactions on Neural Networks 12(2): 181-201.
- Smola A.J. and Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing 14(3): 199-222.