CUbiC

CENTER FOR COGNITIVE UBIQUITOUS COMPUTING

Multiple Kernel Learning for Efficient Conformal Predictions

NIPS Workshop on New Directions in Multiple Kernel Learning

Vineeth Balasubramanian
Shayok Chakraborty
Sethuraman Panchanathan

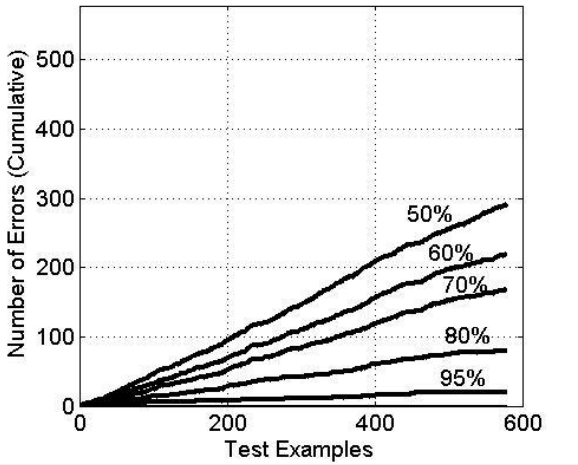
{vineeth.nb,schakr10, panch}@asu.edu

Conformal Predictions

A recent framework for reliable confidence estimation

Desirable Properties

- Calibration of obtained confidence values in online setting
- Can be generalized to any classification/regression algorithm



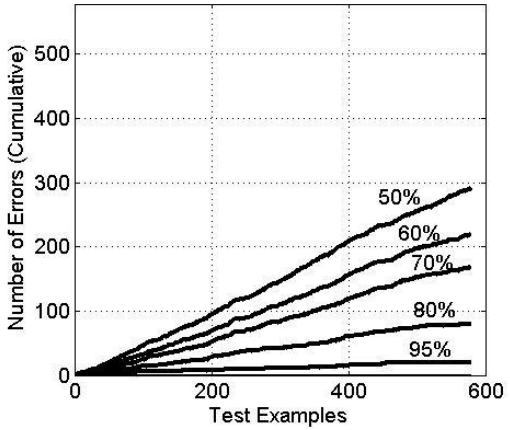
Developed By

- Vovk, Shafer, Gammerman

Theoretical Basis

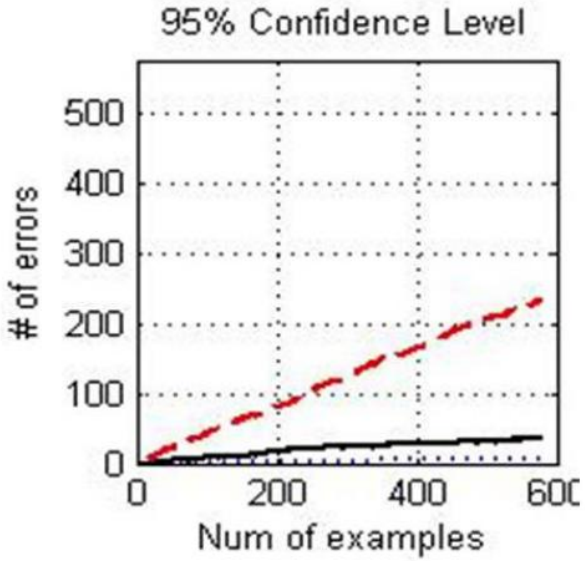
- Algorithmic randomness, Transductive inference, Hypothesis testing

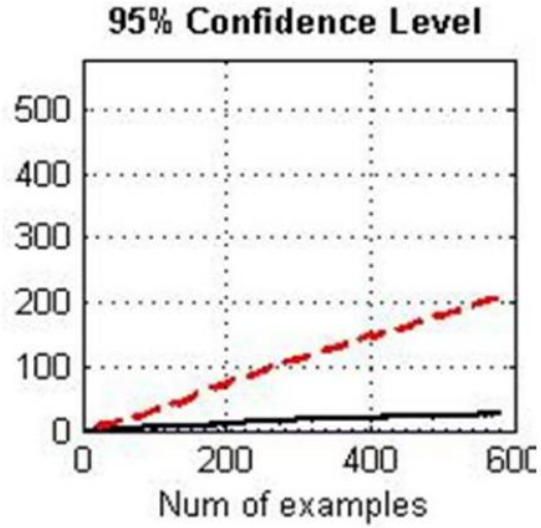
Motivation

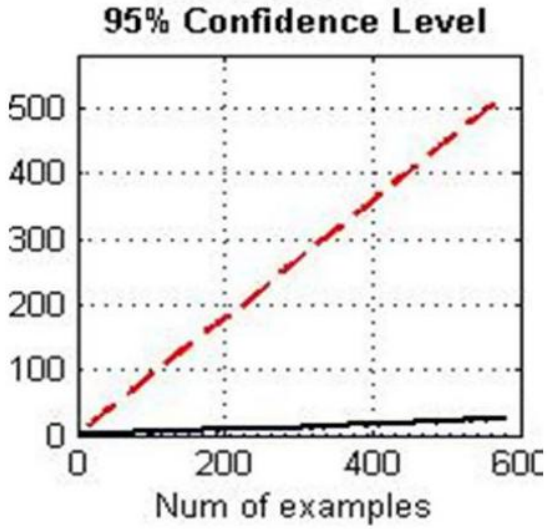


Validity

Efficiency







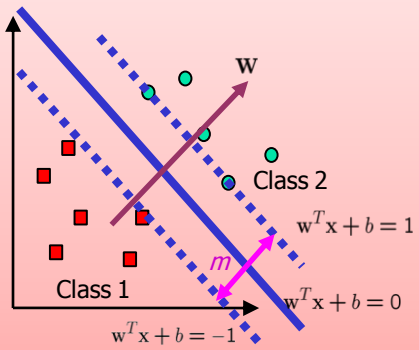
SVM

K-NN

BPNN

Kernel Learning for Efficiency Maximization

Maximize the margin between the class



Minimize $\frac{1}{2}||\mathbf{w}||^2$

subject to

$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1$

Minimize the variance within each class

$$\arg\max_w \frac{w^T S_b w}{w^T S_w w}$$

Within-scatter Matrix

Minimize $w^T S_w w$

$$\min \frac{1}{2} w^T (\lambda S_w + I) w$$

subject to $y_i(w^T x_i + b) \geq 1 \ \forall i = 1, 2, \ldots, n$

Kernel Learning for Efficiency Maximization

$$\min \frac{1}{2} w^T (\lambda S_w + I) w$$

$$\text{subject to } y_i (w^T x_i + b) \geq 1 \ \forall i = 1, 2, \ldots, n$$

$$\Lambda = \lambda S_w + I$$

$$\hat{w} = \Lambda^{1/2} w$$

$$\hat{x}_i = \Lambda^{-1/2} x_i$$

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

$$\text{such that } y_i (\hat{w}^T \hat{x}_i + b) \geq 1 \ \forall i = 1, 2, \ldots, n$$

Kernel Learning for Efficiency Maximization

$$\min \frac{1}{2} w^T (\lambda S_w + I) w$$

$$\text{subject to } y_i (w^T x_i + b) \geq 1 \; \forall i = 1, 2, \ldots, n$$

Existing kernel learning method (Extended Level Method, Xu 2009) used to learn the kernel

$$\Lambda = \lambda S_w + I$$

$$\hat{w} = \Lambda^{1/2} w$$

$$\hat{x}_i = \Lambda^{1/2} x_i$$

Standard SVM Problem!

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

$$\text{such that } y_i (\hat{w}^T \hat{x}_i + b) \geq 1 \; \forall i = 1, 2, \ldots, n$$

Results

Cardiac Patient Dataset