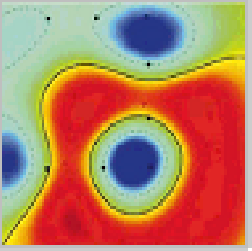
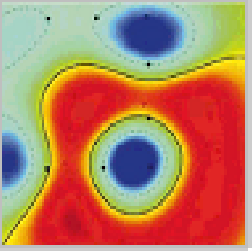


Fast Cross-Validation via Sequential Analysis

Tammo Krueger
Danny Panknin
Mikio Braun

Machine Learning Group
Technische Universitaet Berlin

16.12.2011 Big Learning Workshop

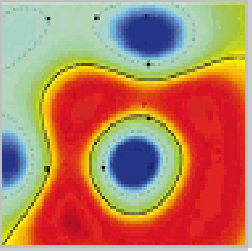


Motivation

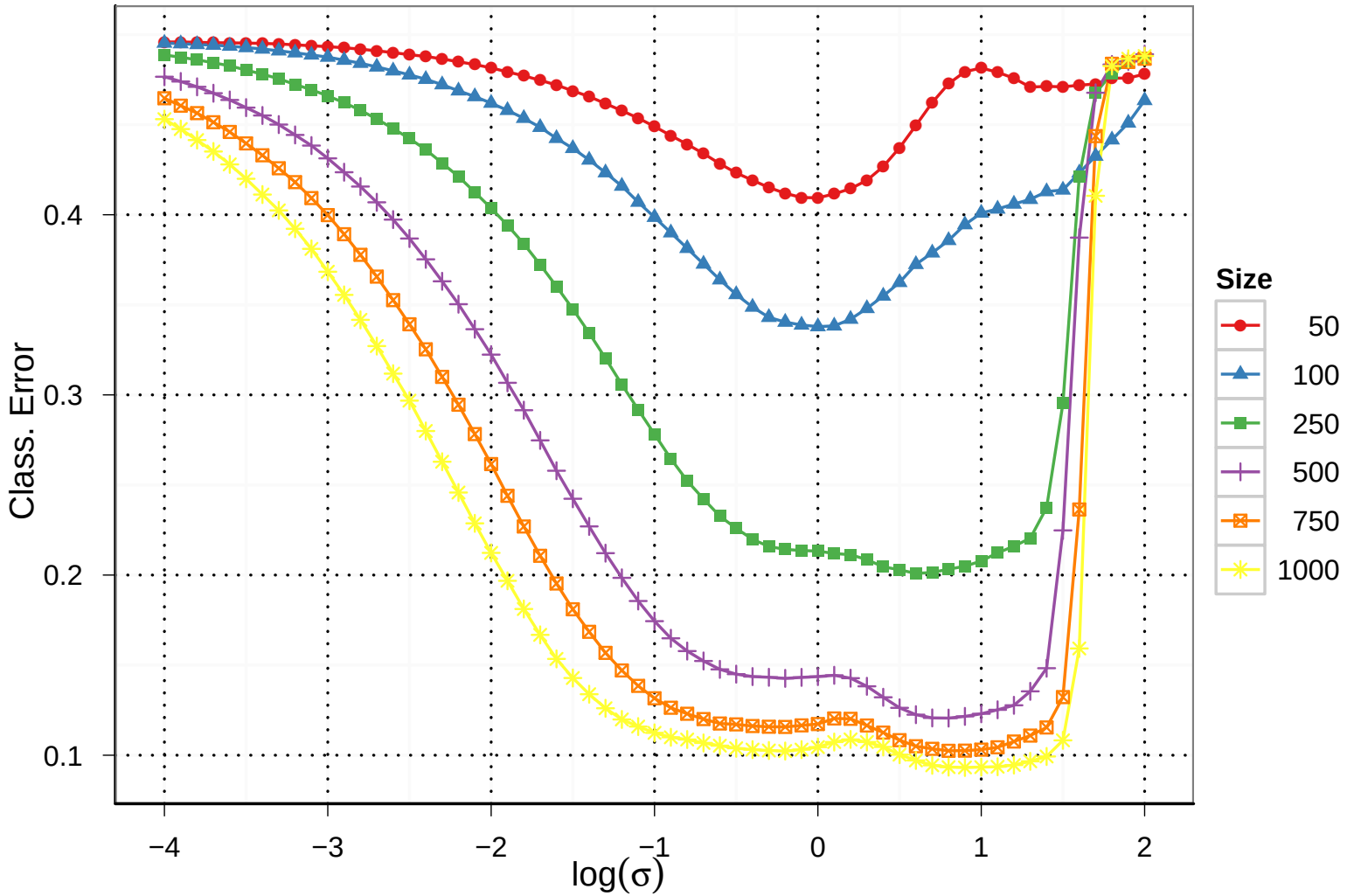
- Cross-validation is an indispensable tool for applied ML but unfortunately very time consuming
- Example *fitted Q iteration*:

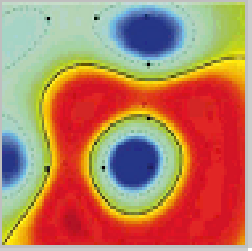
$$\overbrace{\underbrace{10}_{\text{parameter}} \times \underbrace{10}_{\text{fold}}}^{\text{Cross-Validation}} \times \underbrace{50}_{\text{max. iter.}} \times \underbrace{10}_{\text{reps.}} = 500,000 \text{ reg. problems}$$

- Directly optimizing the error landscape to avoid calculations difficult due to noise
- Our approach: use increasing subsets of the training data
  1. smaller subsets $\rightarrow$ less training time
  2. more training data $\rightarrow$ better error estimate
  3. relative behavior of parameter configurations converges

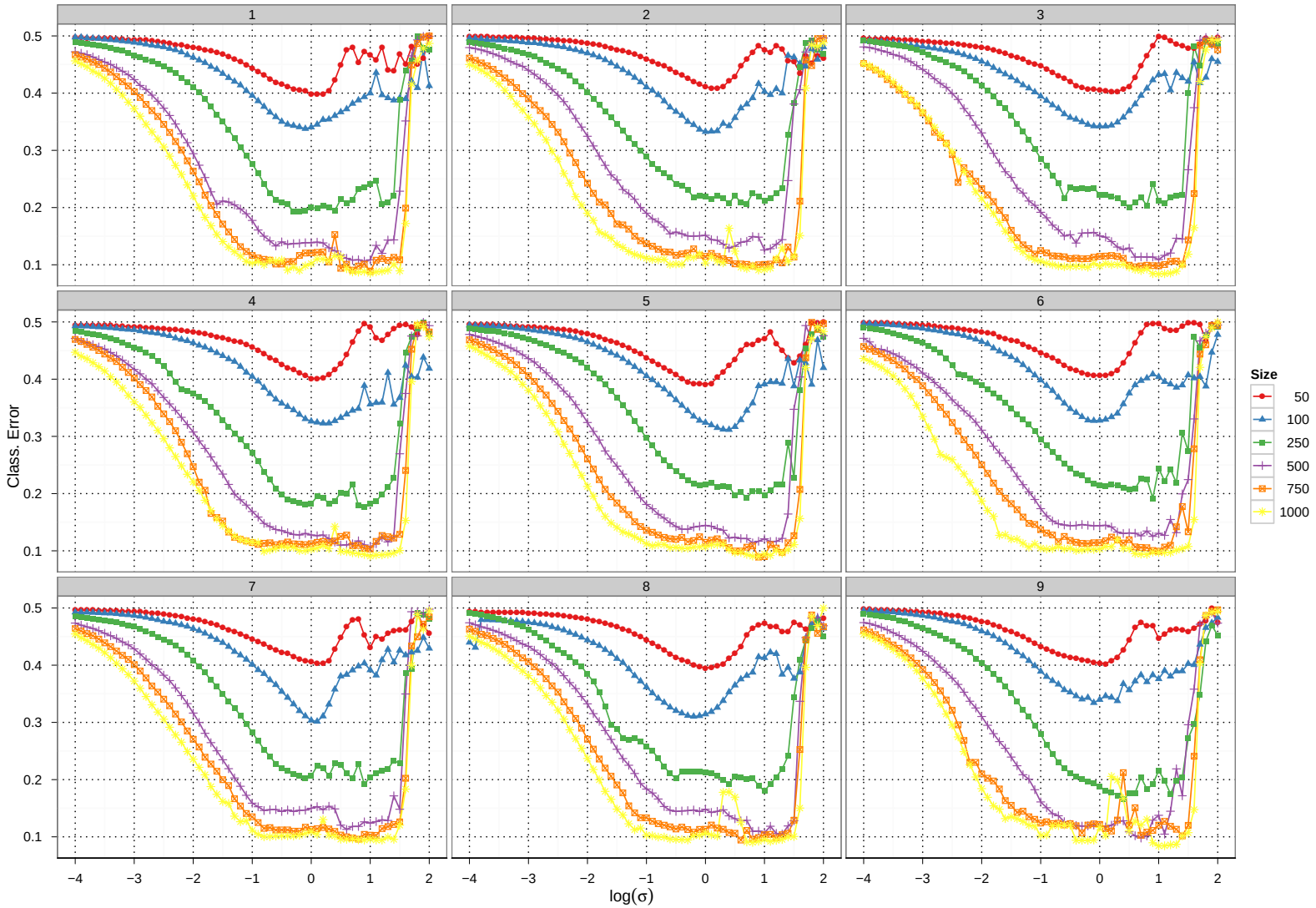


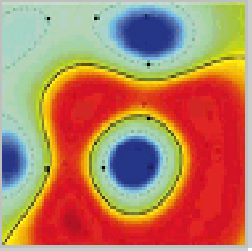
Motivation – Main Idea (Average over 500 Reps.)





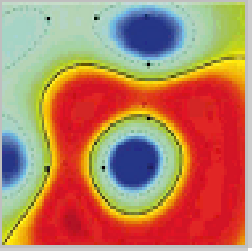
Motivation – Main Idea (Individual Reps.)





Motivation – Exploitation

- Observations:
  1. Individual runs are *noisy*, but at least we can see the tendency
  2. A lot of *underperforming* parameter configurations
  3. We can estimate the correct parameter on a *sufficiently large subset* of the data
- Exploitation:
  1. Transformation of the pointwise test errors of the configurations into a binary *top or flop* scheme
  2. Dropping of *significant loser configurations* along the way via tests from the sequential analysis framework
  3. *Early stopping* of the procedure, when we have seen enough data for a stable parameter estimation



Fast Cross-Validation Procedure – Algorithm

data points

conf.	d_1	d_2	d_3	$\cdots$	d_{n-1}	d_n	
c_1	-2.2	-1.9	-1.8		2.1	1.5	flop
c_2	-1.9	-2.4	-2.3	$\cdots$	1.9	2.4	flop
c_3	-1.4	-0.9	-0.7		0.5	0.5	flop
$\vdots$		$\vdots$			$\vdots$		
c_{k-2}	0.6	0.6	0.7		-0.8	-0.4	top
c_{k-1}	0.1	0.5	0.7	$\cdots$	-0.9	-0.1	top
c_k	0.5	0.4	0.6		-0.3	0.0	top

pointwisePerformance matrix

❶ →

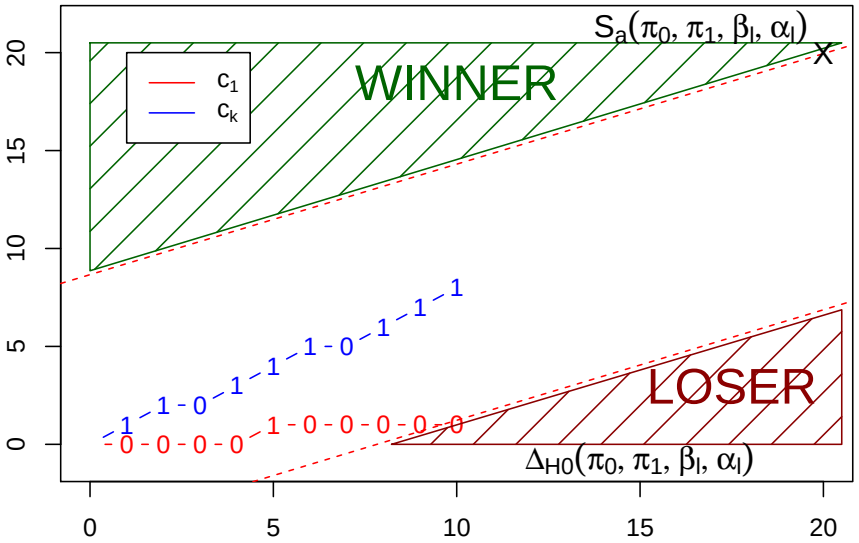
steps

1	2	3	4	5	6	7	8	9	10	
0	0	0	0	1	0	0	0	0	0	(†)
0	1	0	0	0	0	0	0	0	0	(†)
0	1	1	0	0	1	0	0	0	0	
					$\vdots$					
0	1	0	1	1	1	1	0	1	1	
0	1	1	1	1	1	0	1	1	1	
1	1	0	1	1	1	0	1	1	1	

trace matrix

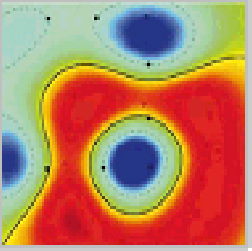
❷ ↙

❸ ↓

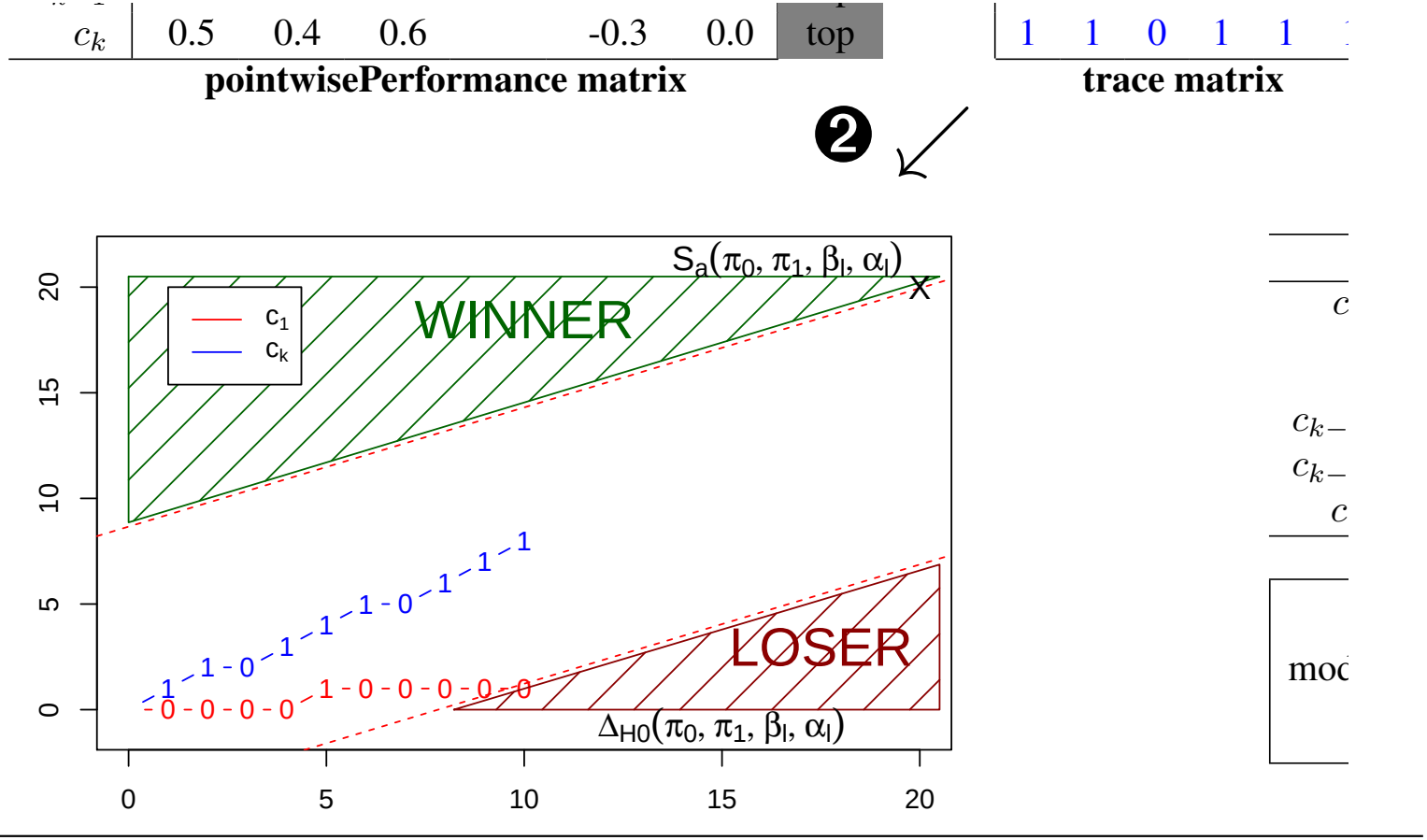


	7	8	9	10	
c_3	0	0	0	0	
$\vdots$		$\vdots$			$\stackrel{?}{=}$
c_{k-2}	1	0	1	1	
c_{k-1}	0	1	1	1	
c_k	0	1	1	1	

$$\Delta = N/20$$
$$\text{modelSize} = 10\Delta$$
$$\text{n} = N - 10\Delta$$

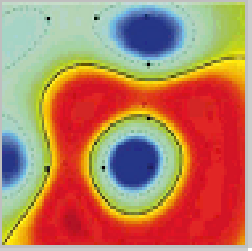


Meta-parameters – Selection of Test Parameters

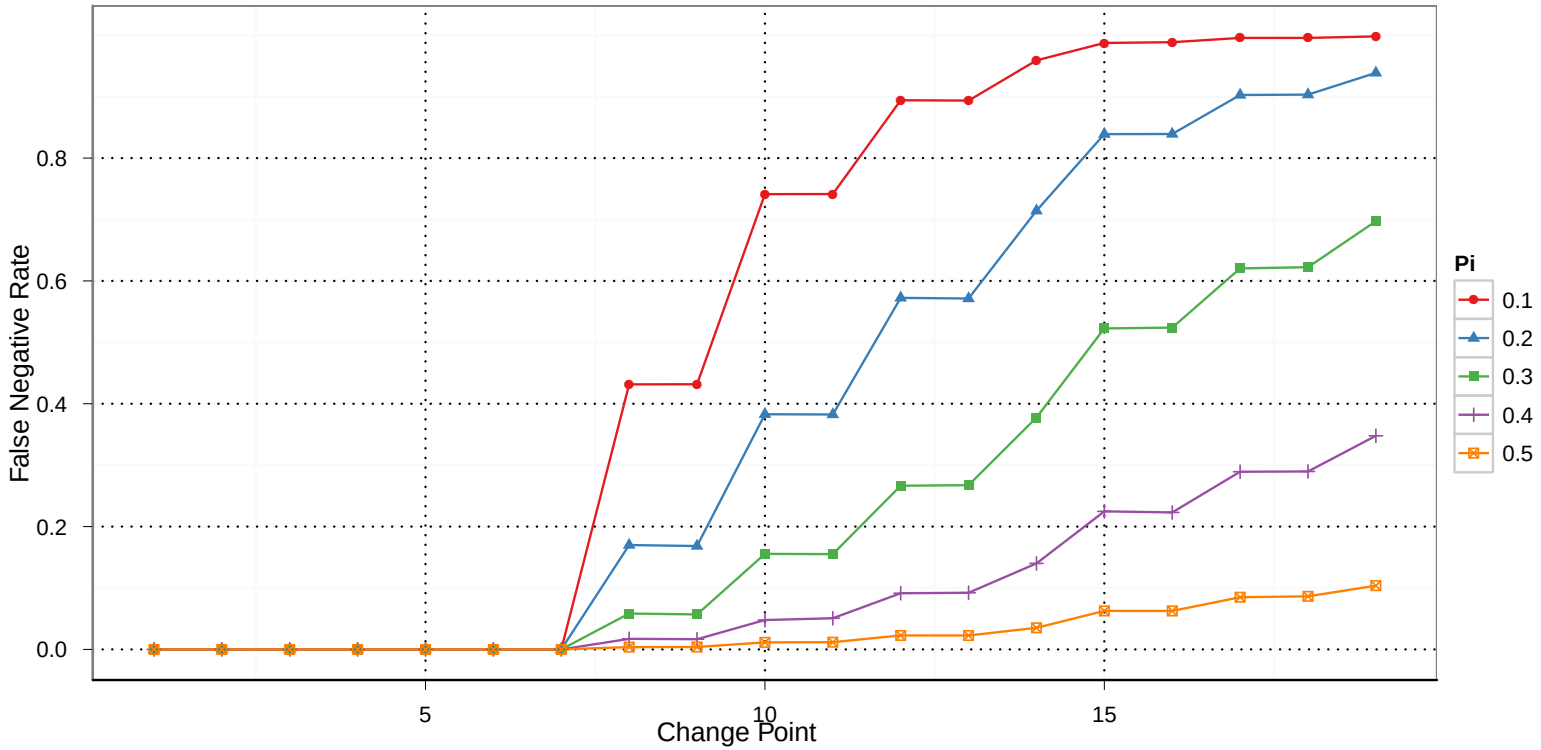


$$(\pi_0, \pi_1) = \underset{\pi_0', \pi_1'}{\text{argmax}} \triangle_{H_0}(\pi_0', \pi_1', \beta_I, \alpha_I)$$

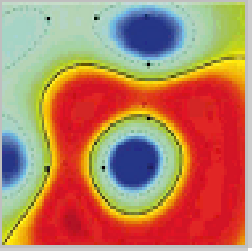
$$\text{s.t. } S_a(\pi_0', \pi_1', \beta_I, \alpha_I) \in (\text{steps} - 1, \text{steps}]$$



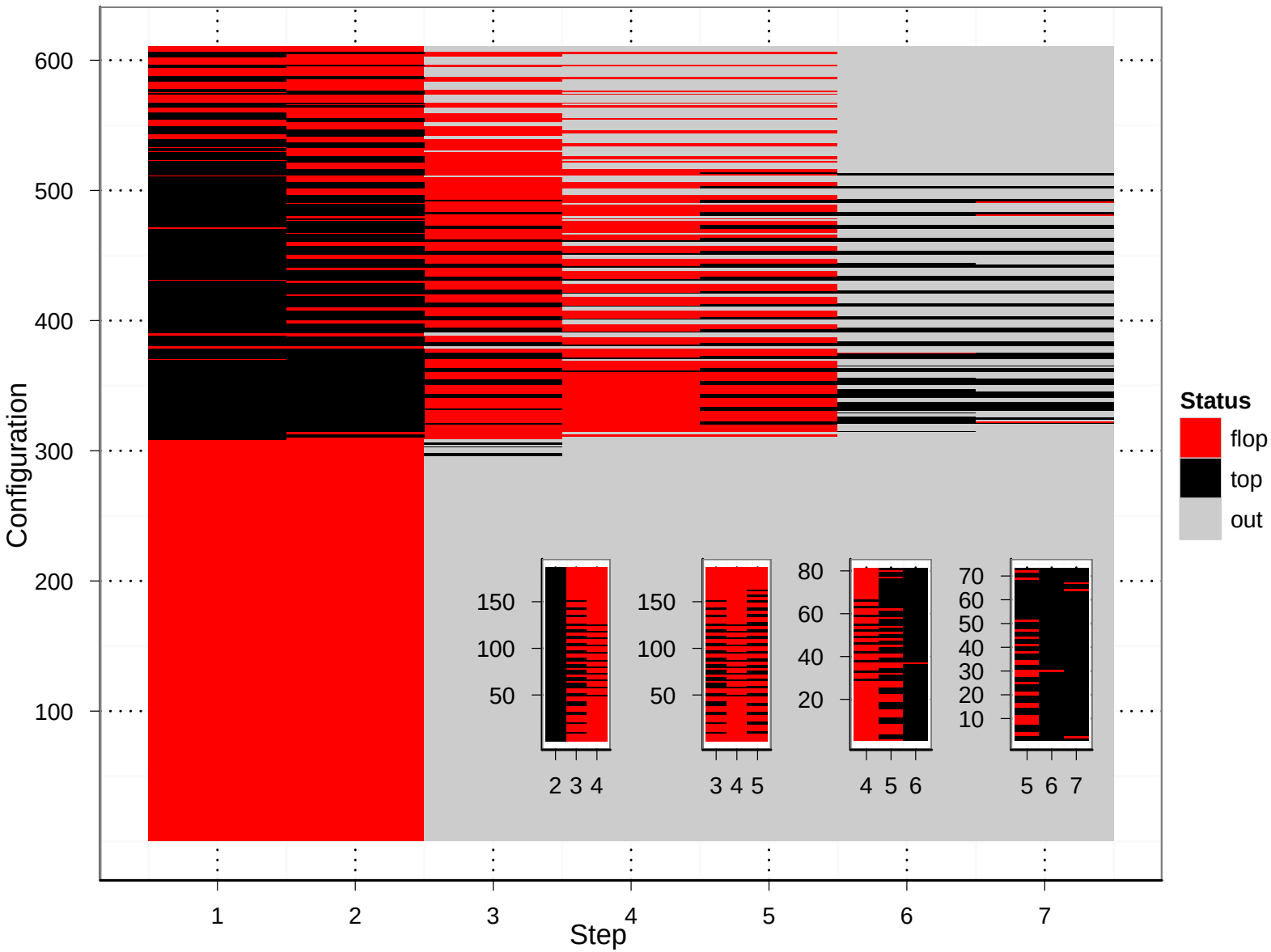
Meta-parameters – False Negative Rate

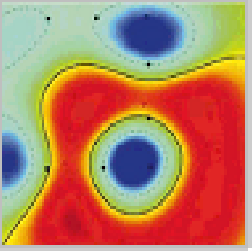


$$\underbrace{0 \leq \frac{cp}{\text{steps}} \leq \frac{\log \frac{\beta_I}{1-\alpha_I} \log \frac{\pi_1}{\pi_0}}{\log \frac{1-\beta_I}{\alpha_I} \log \frac{1-\pi_1}{1-\pi_0}}}_{\textit{security zone}\ \text{(false negative rate of 0)}} \quad \text{with steps} \geq \left\lceil \log \frac{1-\beta_I}{\alpha_I} / \log 2 \right\rceil$$



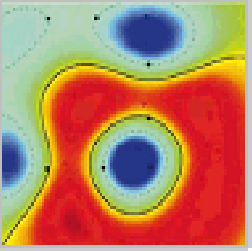
Fast Cross-Validation Procedure – Example Run



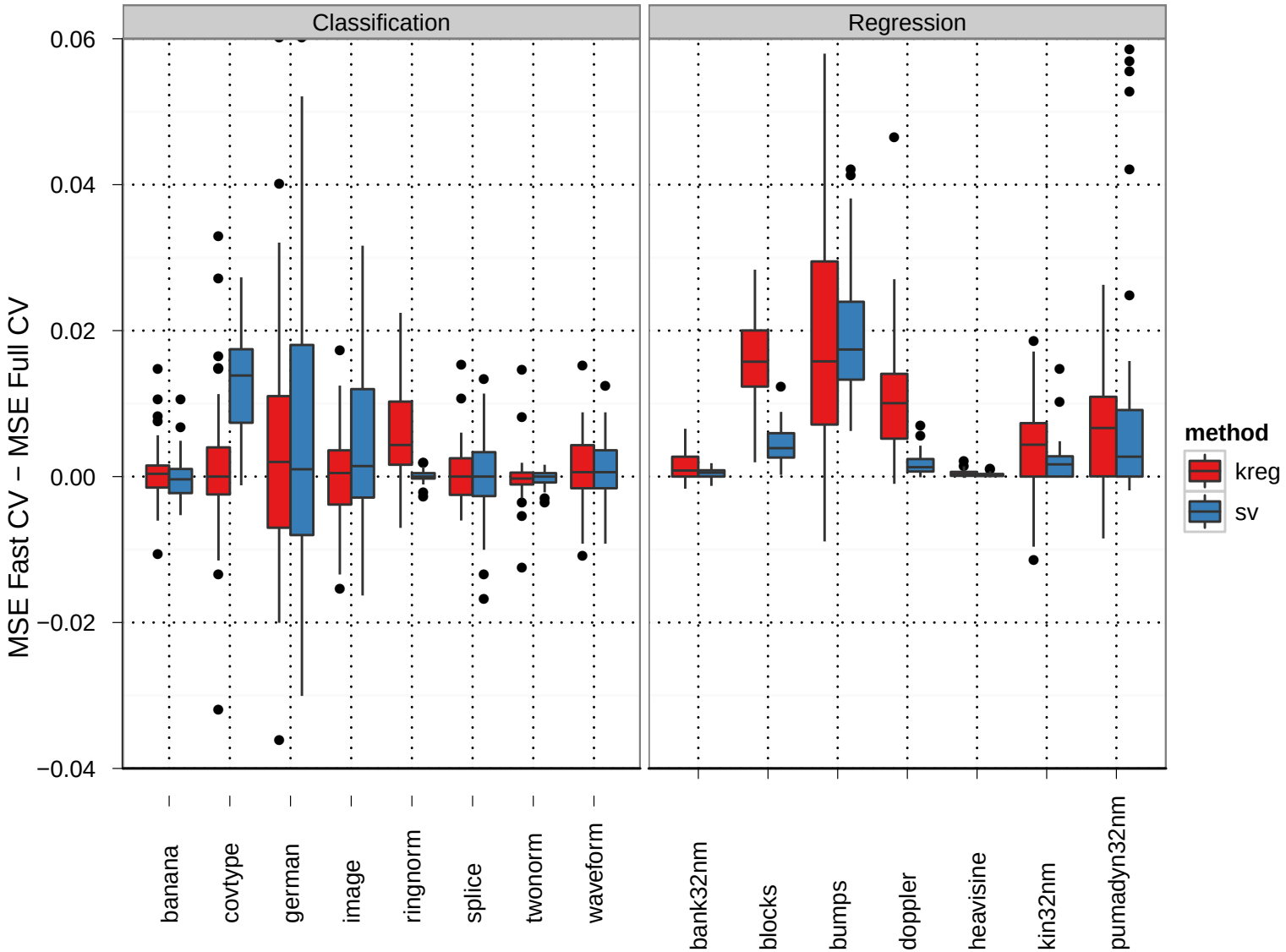


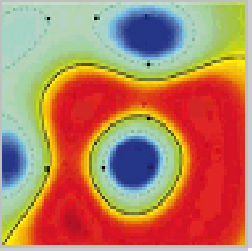
Experimental Setup

- 8 classification and 7 regression data sets
- For each dataset:
  - $\frac{1}{2}$ for parameter estimation, $\frac{1}{2}$ for test error estimation
  - SVM/SVR and Kernel Ridge Regression/Kernel Logistic Regression with Gaussian kernel using 610 parameter configurations
  - Parameter estimation with:
    - Full 10-fold cross-validation
    - Fast cross-validation procedure with 10 steps
- Repeated 50 times with different splits for each dataset
- Compare:
  - Test error difference of fast versus full cross-validation
  - Relative speed factor, i.e. $\frac{\text{time full cross-validation}}{\text{time fast cross-validation}}$

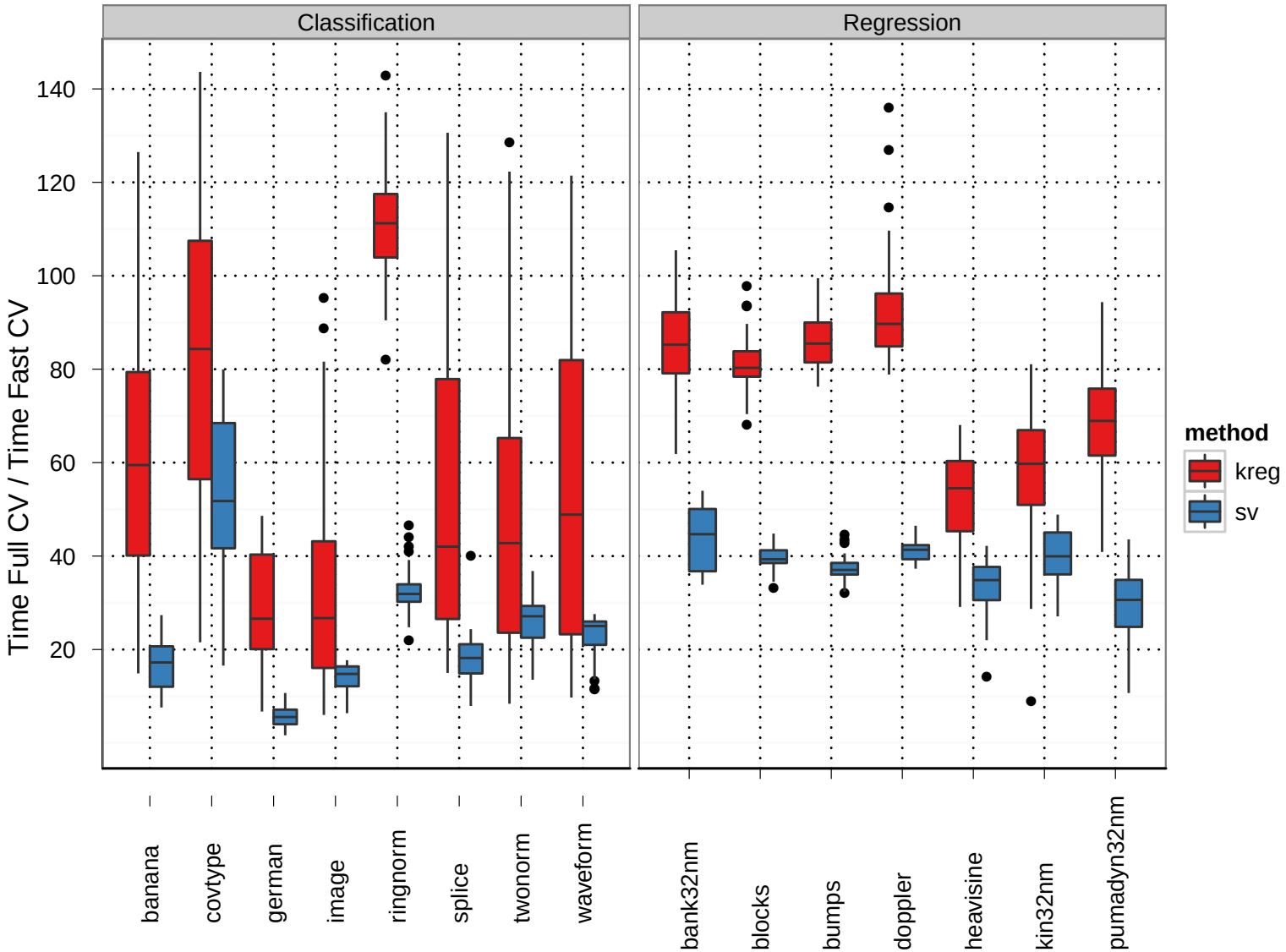


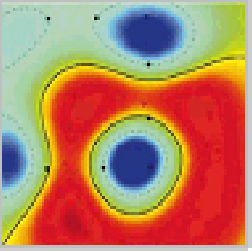
Experiments – Test Error





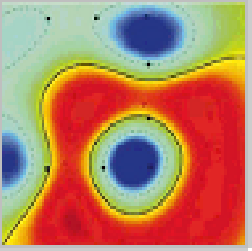
Experiments – Speed Increase



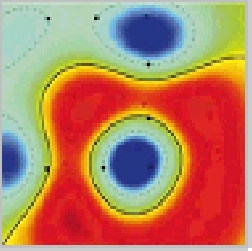


Fast Cross-Validation Procedure – Summary

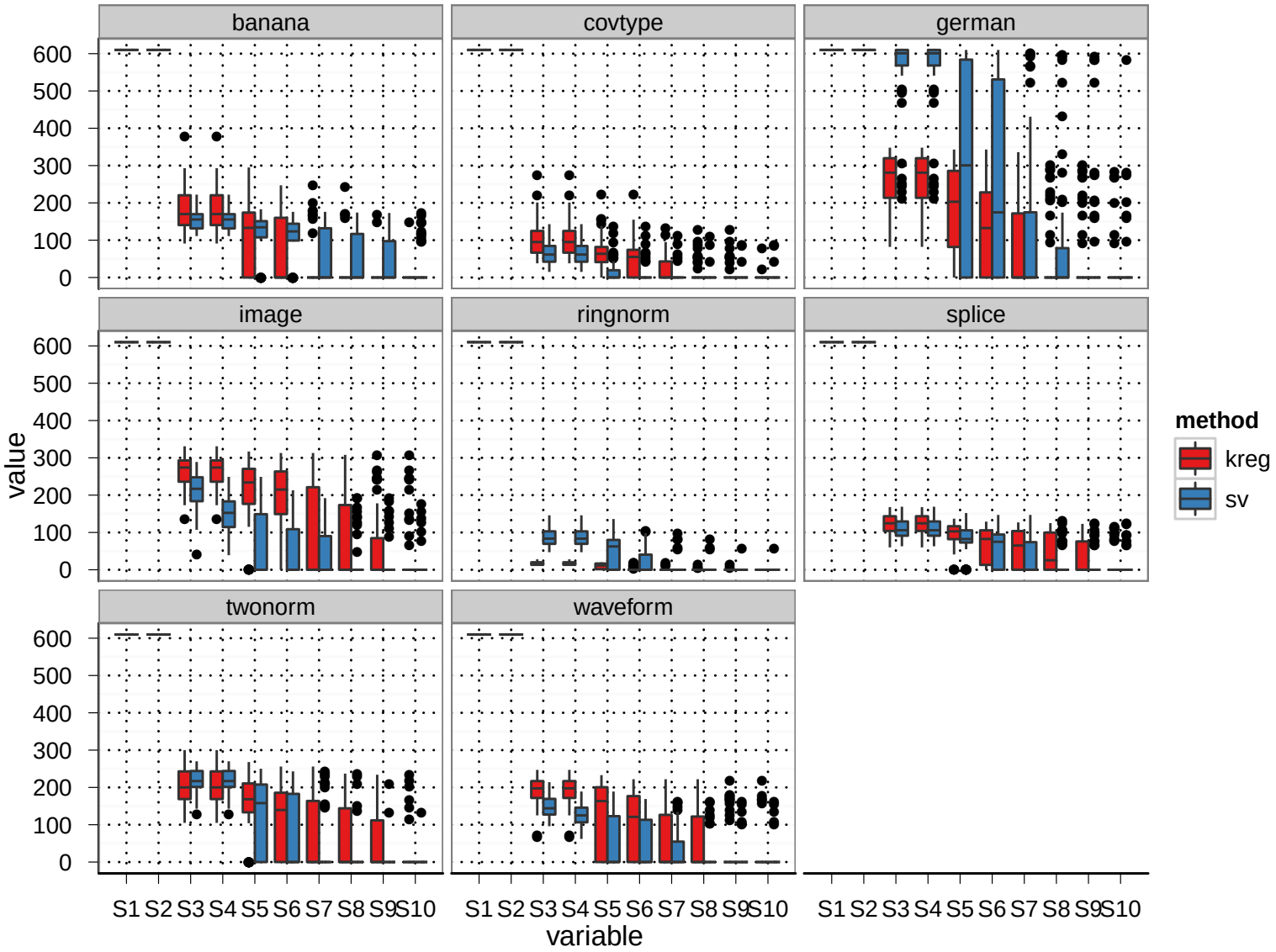
- Motivation: we can estimate the correct parameter on a *sufficiently large subset* of the data
- Transformation: *Race of configurations* evaluated on linearly increasing subsets of the data
- At each step of this race:
  1. Transform the test errors on individual data points of the remaining configurations into a binary *top or flop* scheme
  2. Drop significant loser configurations along the way using tests from the sequential analysis framework
  3. Apply distribution free testing techniques to decide, whether we have gathered enough evidence for a stable parameter estimation

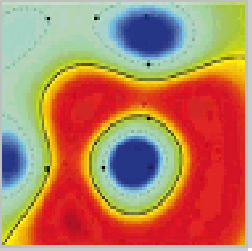


Questions? Remarks?
Thanks for your attention!



Experiments – Traces Classification





Experiments – Traces Regression