Non-Functional Properties

Prof. Sven Apel

Universität des Saarlandes

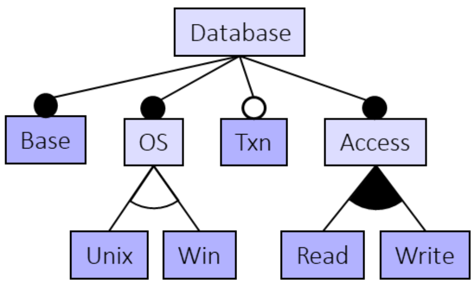
Part I

Motivation and Overview

Our view so far:

Domain Engineering

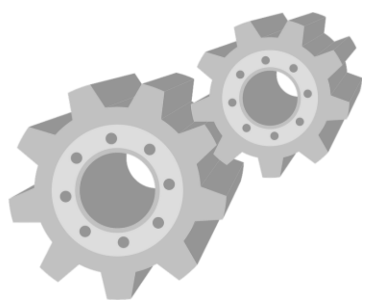
Feature model



Reusable / configurable code

Application Engineering

Unix ✓
Win ✗
Txn ✓
Read ✓
Write ✗

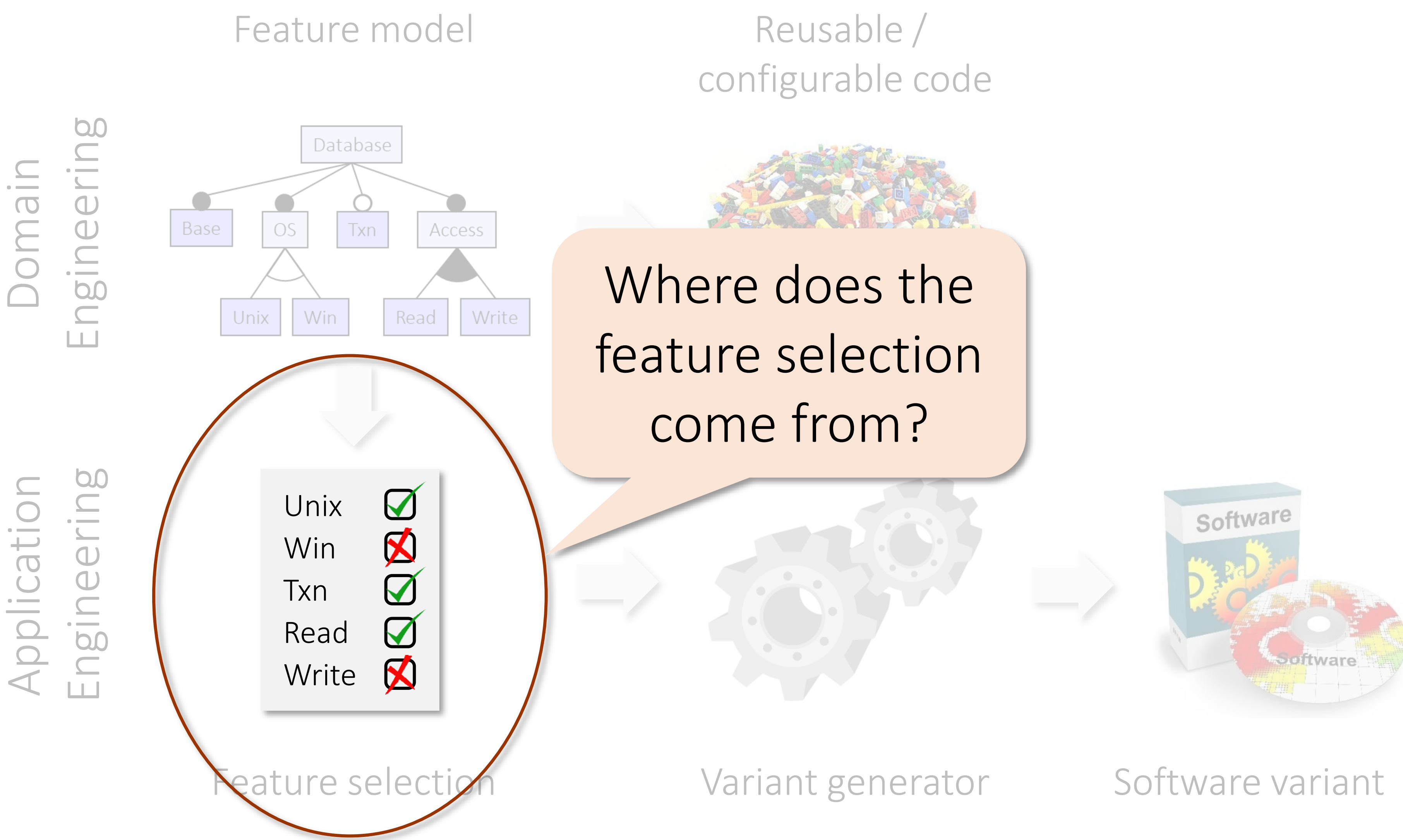


Feature selection

Variant generator

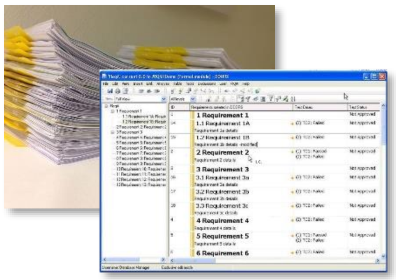
Software variant

Our view so far:

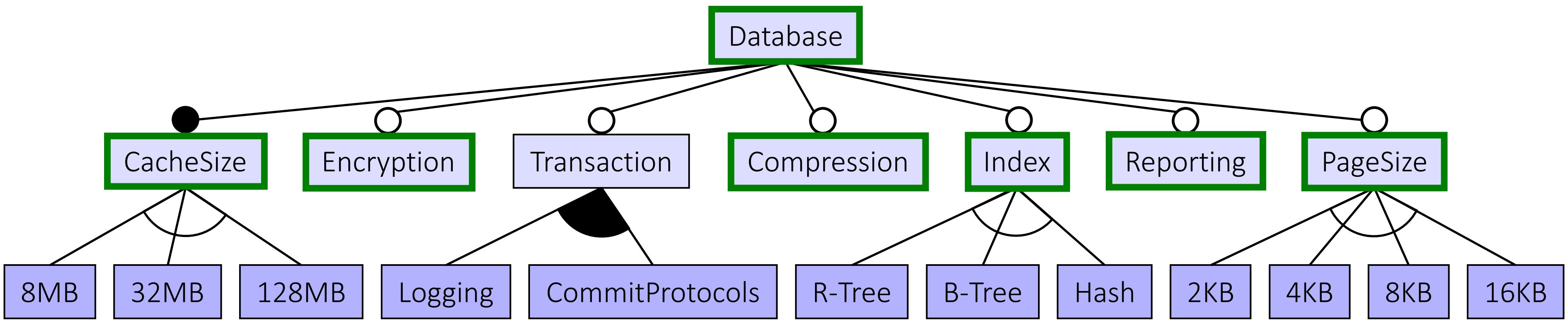


From Requirements to Configurations

Functional requirements



Features



Partial feature selection

Non-Functional Requirements

Performance

Memory consumption

Footprint

…

Definition(s)

Also known as *quality attributes*

Over 25 definitions

In general:

> *Any property of a product that is not related with functionality represents a non-functional property.*

Often there are *trade-offs* between properties!

Categorization

Quantitative

Response time (performance), throughput, etc.

Energy- and memory consumption

Measurable properties, metric scale

→ *Easy to evaluate*

Qualitative

Extensibility

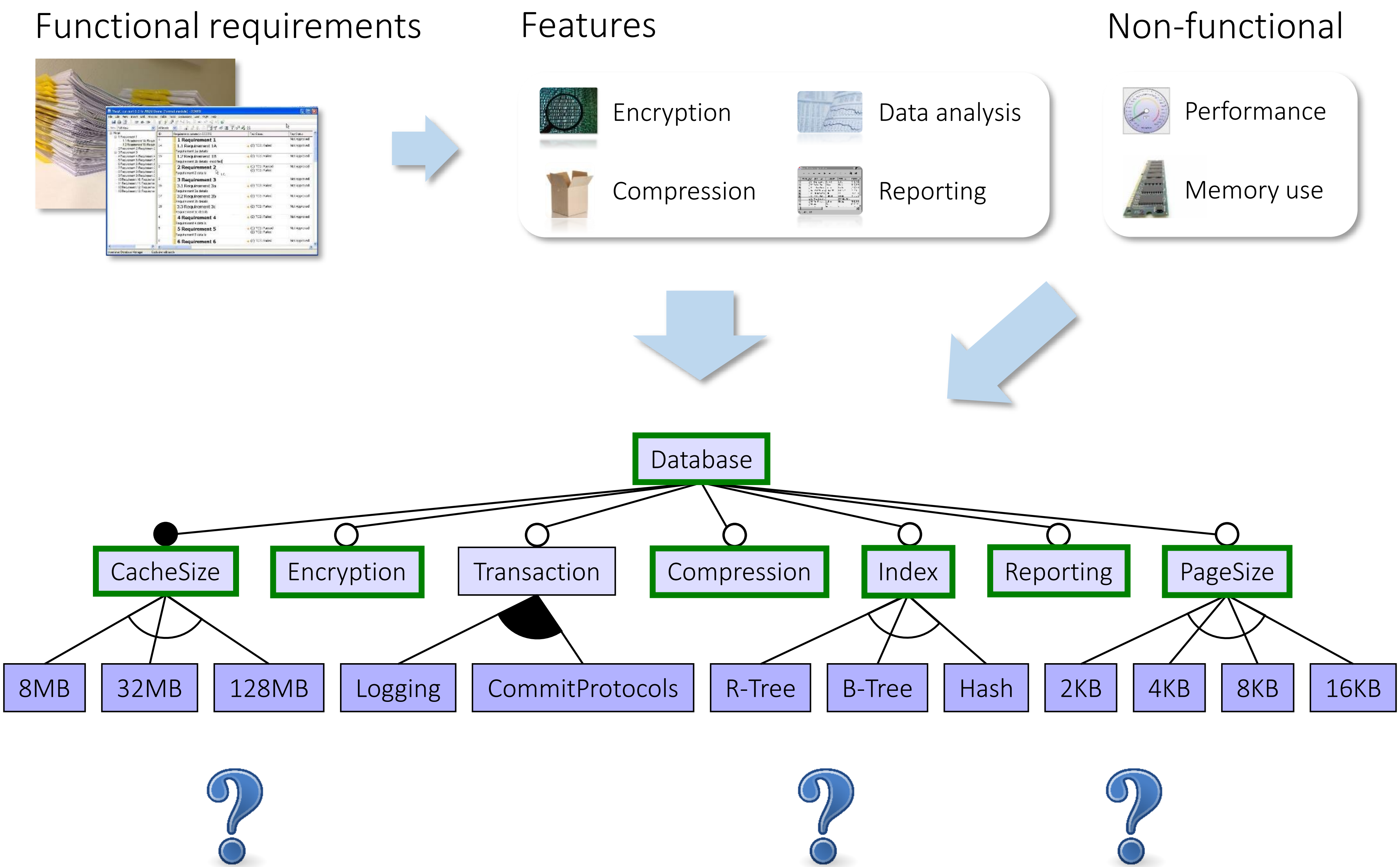
Error freeness

Robustness

Security

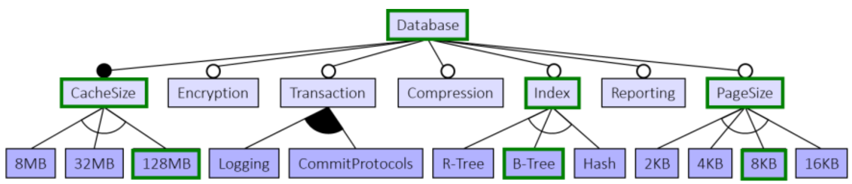
→ *No direct measurement (often, no suitable metric)*

Incorporating Non-Functional Properties



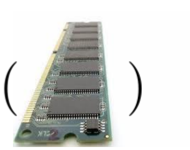
Practical Relevance

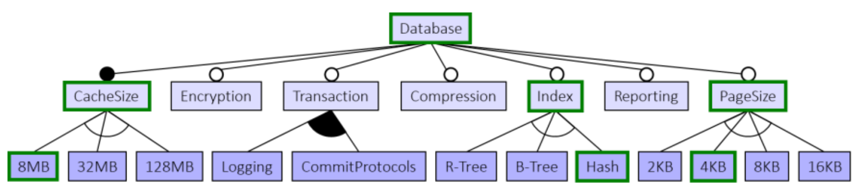
What is the binary footprint of a variant for a given feature selection?



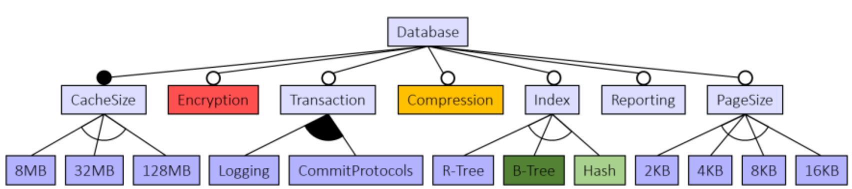
425 KB

What is the best feature selection to minimize memory consumption?

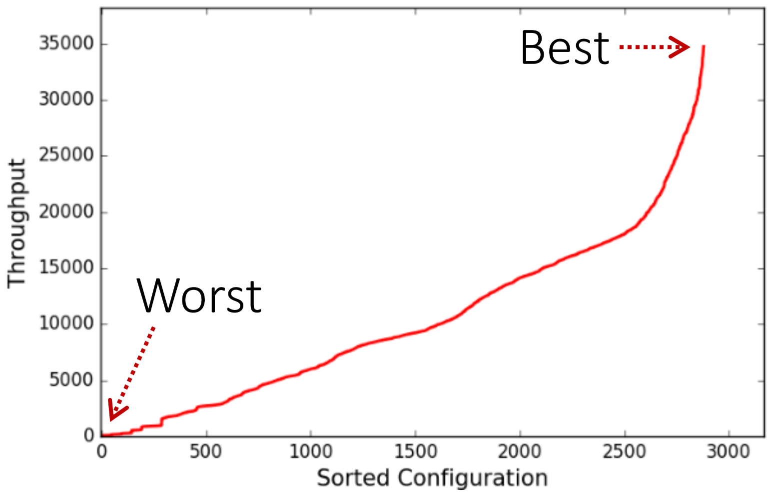
min()



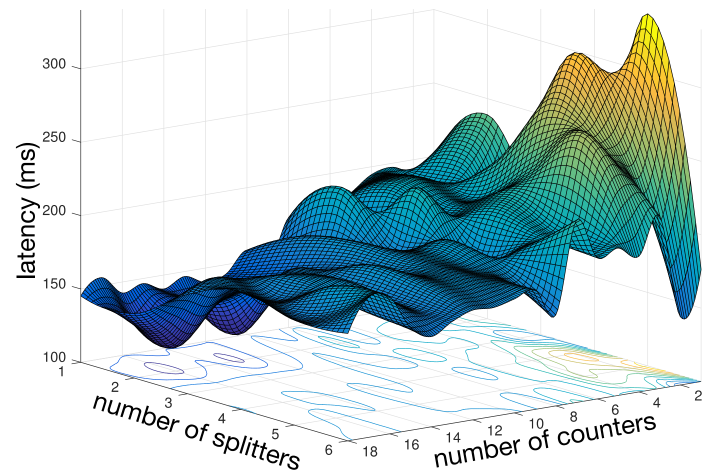
What are the performance critical features?



Practical Relevance



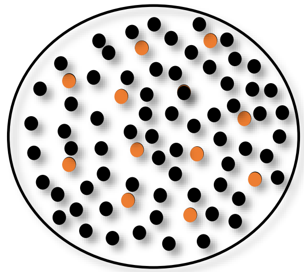
Best configuration is *480 times* better than worst configuration



Only by tweaking *2 of 200 options* of Apache Storm led to *100% change in latency*

Overview

Configuration space
$|C| = \mathcal{O}(2^{|F|})$

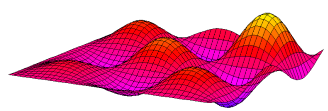


1. Measurement

Goal: Measure non-functional properties, e.g., via sampling (select a minimal, representative set of configurations)
Challenges: Size of configuration space, constraints, interactions
Key domains: Combinatorial testing, artificial intelligence, search-based software engineering, design of experiments

Performance model

$$\Pi : C \rightarrow \mathbb{R}$$



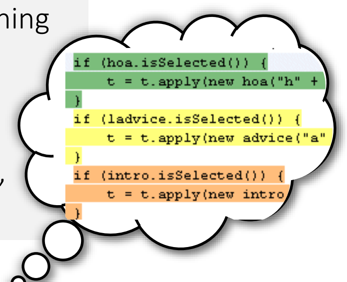
2. Learning a model

Goal: Learn a model accurately describing performance of all configurations
Challenges: Dimensionality of the learning problem, interactions
Key domains: machine learning, statistics

Goal

4. Comprehension

Goal: Explainable machine learning with white-box models
Challenges: Trade-off between explainability and accuracy
Key domains: machine learning, software analysis, testing



$$\begin{array}{ccc} c_i & = & \{\ldots\} \\ \vdots & \vdots & \vdots \\ c_k & = & \{\ldots\} \end{array}$$

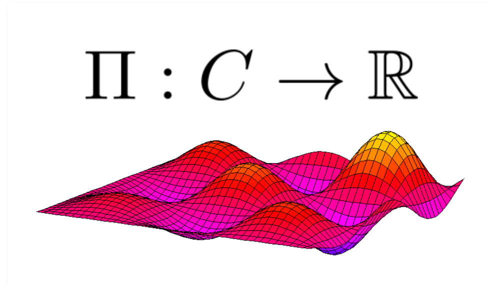
3. Optimization

Goal: Finding optimal configurations in a single or multi-objective scenario
Challenges: Size of configuration space, constraints, interactions, lack of ground truth
Key domains: search-based software engineering, meta-heuristics, machine learning, artificial intelligence, mathematical optimization

System understanding

Optimal configuration(s)

Predictive Modeling / Machine Learning



?

Linear regression (Least-square, Lasso, Ridge, ...)

Regression trees (CART, Random Forests, ...)

Support vector machines

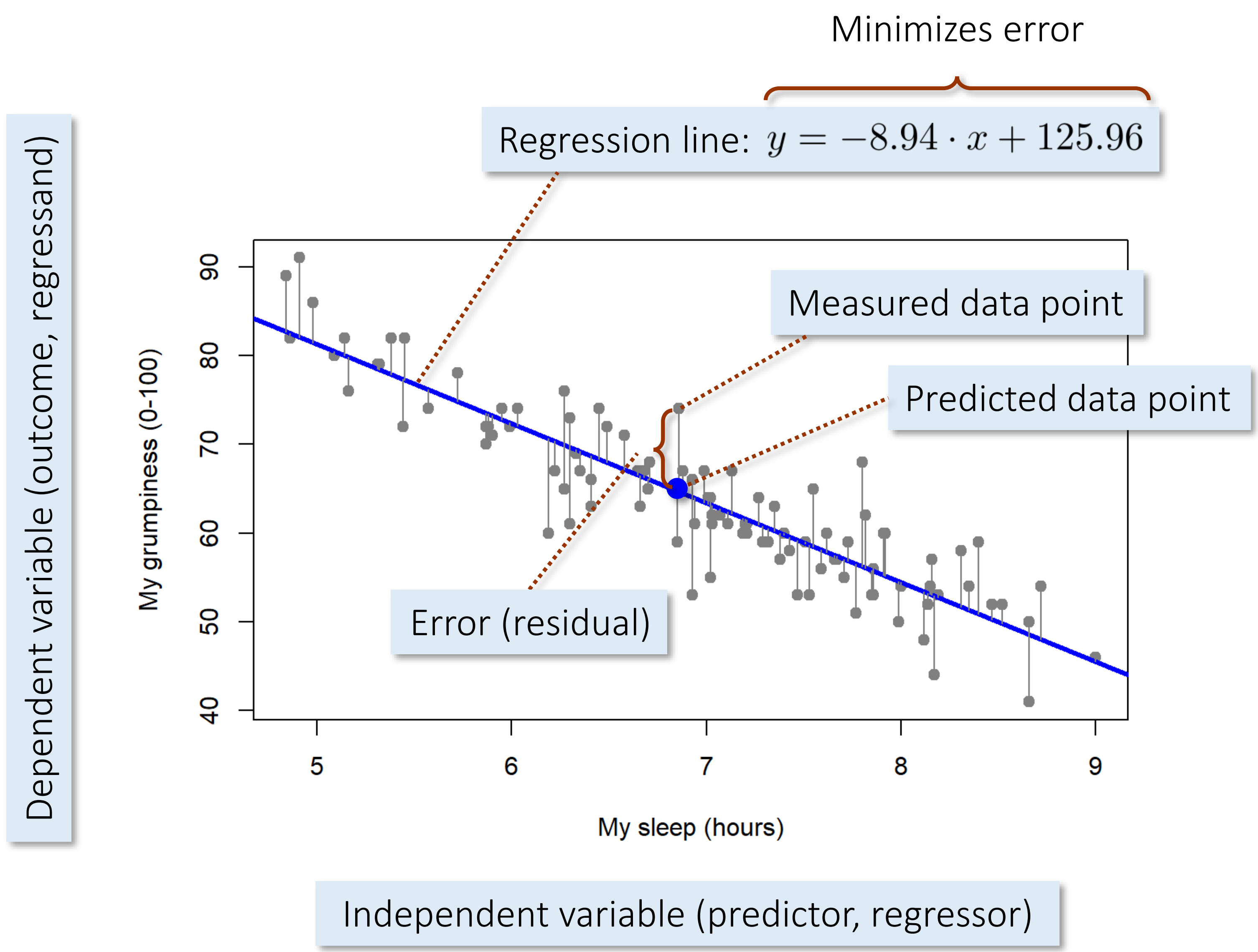
Neural networks

...

Part II

Learning Influence Models

Linear Regression



Linear Regression

Model structure:

Influences

Error

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$$

Property
(e.g., performance)

Features

Single observation: $(y, \mathbf{x})$
(one sample)

All observations: $(\mathbf{y}, X)$
(whole sample)

1. Sampling observations:
Influences and errors are unknown!

Linear Regression

Observations:

$$(\mathbf{y}, X)$$

Optimization:

$$\min \quad \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta})$$

Predictor (performance model):

$$\Pi(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

Example model:

$$\Pi\big((x_1, x_2)\big) = 15 - 8.9\, x_1 + 154\, x_2$$

2.Learning a predictor:

Minimizing the overall error when assigning influences

Performance is non-linear!

```
x264
     --no-progress

     --rc-lookahead 60
     --ref 9
     -o trailer_480p24.x264
     trailer_2k_480p24.y4m
```

661s

```
x264 --quiet
     --no-progress

     --rc-lookahead 60
     --ref 9
     -o trailer_480p24.x264
     trailer_2k_480p24.y4m
```

487s (-174s)

```
x264
     --no-progress
     --no-asm
     --rc-lookahead 60
     --ref 9
     -o trailer_480p24.x264
     trailer_2k_480p24.y4m
```

551s (-110s)

```
x264 --quiet
     --no-progress
     --no-asm
     --rc-lookahead 60
     --ref 9
     -o trailer_480p24.x264
     trailer_2k_480p24.y4m
```

324s (-337s)

Performance is non-linear!

```
x264
     --no-progress

     --rc-lookahead 60
     --ref 9
     -o trailer_480p24.x264
     trailer_2k_480p24.y4m
```

661s

```
x264 --quiet
     --no-progress

     --rc-lookahead 60
     --ref 9
     -o trailer_480p24.x264
     trailer_2k_480p24.y4m
```

487s (-174s)

Influences of individual features do not add up in combination!

```
x264
     --no-progress
     --no-asm
     --rc-lookahead 60
     --ref 9
     -o trailer_480p24.x264
     trailer_2k_480p24.y4m
```

551s (-110s)

```
     --no-progress
     --no-asm
     --rc-lookahead 60
     --ref 9
     -o trailer_480p24.x264
     trailer_2k_480p24.y4m
```

324s (-337s)

Feature Selection

We cannot include all possible combinations of features as distinct variables!

$$y_i = \beta_0 + \beta_1 \boxed{x_{i1}} + \ldots + \beta_p \boxed{x_{ip}} + \varepsilon_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$$

Features & interactions

Features only: $I^{(1)} = \mathcal{O}(|F|)$

All pairs of features: $I^{(2)} = \mathcal{O}(|F|^2)$

...

All feature combinations: $I^* = \mathcal{O}(2^{|F|})$

Which features to select and how?

Feature Selection: Example

Features

$$X_a = \begin{pmatrix} 1 & 0 & 0 & 20 & 0 \\ 0 & 1 & 1 & 50 & 16 \\ 1 & 0 & 1 & 100 & 32 \\ 1 & 1 & 0 & 50 & 32 \\ 1 & 1 & 1 & 20 & 32 \\ 0 & 0 & 0 & 100 & 0 \\ 1 & 1 & 1 & 100 & 16 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 833 \\ 411 \\ 290 \\ 799 \\ 753 \\ 514 \\ 416 \end{pmatrix}$$

$$\boldsymbol{\beta}_a = \begin{pmatrix} 102.4 & 84.3 & 54.1 & 5.4 & 1.3 \end{pmatrix} \qquad \bar{\varepsilon}_a = 41\%$$

Feature Selection: Example

Features Interactions

$$X_b = \begin{pmatrix} 1 & 0 & 0 & 20 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 50 & 16 & 0 & 0 & 800 \\ 1 & 0 & 1 & 100 & 32 & 0 & 0 & 0 \\ 1 & 1 & 0 & 50 & 32 & 1 & 0 & 1600 \\ 1 & 1 & 1 & 20 & 32 & 1 & 1 & 640 \\ 0 & 0 & 0 & 100 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 100 & 16 & 1 & 1 & 1600 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 833 \\ 411 \\ 290 \\ 799 \\ 753 \\ 514 \\ 416 \end{pmatrix}$$

$$\boldsymbol{\beta}_a = \begin{pmatrix} 102.4 & 84.3 & 54.1 & 5.4 & 1.3 \end{pmatrix} \qquad \bar{\varepsilon}_a = 41\%$$

$$\boldsymbol{\beta}_b = \begin{pmatrix} 132.3 & 81.3 & 56.6 & 3.5 & 1.9 & 14.1 & 5.4 & 2.4 \end{pmatrix} \qquad \bar{\varepsilon}_b = 25\%$$

Feature Selection: Example

Features Interactions Functions

$$X_c = \begin{pmatrix} 1 & 0 & 0 & 20 & 0 & 0 & 0 & 0 & 400 & - \\ 0 & 1 & 1 & 50 & 16 & 0 & 0 & 800 & 2500 & 1.2 \\ 1 & 0 & 1 & 100 & 32 & 0 & 0 & 0 & 10^4 & 0.9 \\ 1 & 1 & 0 & 50 & 32 & 1 & 0 & 1600 & 2500 & 1.5 \\ 1 & 1 & 1 & 20 & 32 & 1 & 1 & 640 & 400 & 1.5 \\ 0 & 0 & 0 & 100 & 0 & 0 & 0 & 0 & 10^4 & - \\ 1 & 1 & 1 & 100 & 16 & 1 & 1 & 1600 & 10^4 & 1.2 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 833 \\ 411 \\ 290 \\ 799 \\ 753 \\ 514 \\ 416 \end{pmatrix}$$

$$\boldsymbol{\beta}_a = \begin{pmatrix} 102.4 & 84.3 & 54.1 & 5.4 & 1.3 \end{pmatrix} \qquad \bar{\varepsilon}_a = 41\%$$

$$\boldsymbol{\beta}_b = \begin{pmatrix} 132.3 & 81.3 & 56.6 & 3.5 & 1.9 & 14.1 & 5.4 & 2.4 \end{pmatrix} \qquad \bar{\varepsilon}_b = 25\%$$

$$\boldsymbol{\beta}_c = \begin{pmatrix} 130.3 & 83.5 & 54.2 & 0.01 & 0 & 14.1 & 5.4 & 1.4 & 2.1 & 8.8 \end{pmatrix} \qquad \bar{\varepsilon}_c = 7.4\%$$

Feature Selection: Example

Features | Interactions | Functions

$$
X_c = \begin{pmatrix}
1 & 0 & 0 & 20 & 0 & 0 & 0 & 0 & 400 & - \\
0 & 1 & 1 & 50 & 16 & 0 & 0 & 800 & 2500 & 1.2 \\
1 & 0 & 1 & 100 & 32 & 0 & 0 & 0 & 10^4 & 0.9 \\
1 & 1 & 0 & 50 & 32 & 1 & 0 & 1600 & 2500 & 1.5 \\
1 & 1 & 1 & 20 & 32 & 1 & 1 & 640 & 400 & 1.5 \\
0 & 0 & 0 & 100 & 0 & 0 & 0 & 0 & 10^4 & - \\
1 & 1 & 1 & 100 & 16 & 1 & 1 & 1600 & 10^4 & 1.2
\end{pmatrix}
\qquad
\mathbf{y} = \begin{pmatrix} 833 \\ 411 \\ 290 \\ 799 \\ 753 \\ 514 \\ 416 \end{pmatrix}
$$

$$\boldsymbol{\beta}_a = \begin{pmatrix} 102.4 & 84.3 & 54.1 & 5.4 & 1.3 \end{pmatrix} \qquad \bar{\varepsilon}_a = 41\%$$

$$\boldsymbol{\beta}_b = (132.3 \;\; 81.3 \qquad \bar{\varepsilon}_b = 25\%$$

$$\boldsymbol{\beta}_c = (130.3 \;\; 83.5 \qquad \bar{\varepsilon}_c = 7.4\%$$

Search for the best model in *multiple iterations* adding *more and more variables* of *increasing complexity*!

Empirical Results

System	Domain	# Binary Opt.	# Numeric Opt.	# Constraints	# Configs
Dune MGS	Multi-Grid Solver	8	3	20	2 304
HIPA^{cc}	Image Processing	31	2	416	13 485
HSMGP	Stencil-Grid Solver	11	3	45	3 456
JavaGC	Runtime Env.	12	23	4	10^{31}
SaC	Compiler	53	7	10	10^{23}
x264	Video Encoder	8	13	0	10^{27}

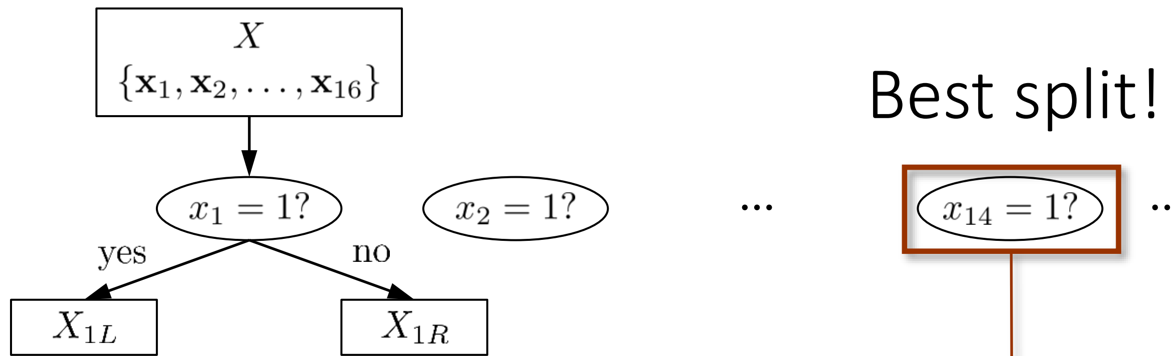
	Plackett-Burman Design		Random Design	
System	Feature-wise	Pair-wise	Feature-wise	Pair-wise
Dune MGS	8.8 %	**8.3 %**	20.1 %	22.1 %
HIPA^{cc}	13.8 %	**10.7 %**	14.2 %	13.9 %
HSMGP	1.7 %	**1.5 %**	4.5 %	2.8 %
JavaGC	21.9 %	**18.8 %**	31.3 %	24.6 %
SaC	**16.0 %**	25.0 %	21.1 %	30.7 %
x264	21.2 %	15.0 %	14.2 %	**13.5 %**

Regression Trees (CART)

Running example:

Configuration	Features																Performance (s)
$\mathbf{x}_i$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	y_i
$\mathbf{x}_1$	1	1	0	1	1	1	1	0	1	0	0	1	1	0	0	1	651
$\mathbf{x}_2$	1	1	1	1	1	1	0	1	1	1	0	0	1	0	1	0	536
$\mathbf{x}_3$	1	1	1	1	0	0	0	0	1	1	0	0	1	0	0	1	581
$\mathbf{x}_4$	1	0	0	0	0	0	1	0	1	1	0	0	1	0	1	0	381
$\mathbf{x}_5$	1	1	0	1	0	0	0	1	1	1	0	0	1	0	1	0	424
$\mathbf{x}_6$	1	1	0	0	1	0	1	1	1	1	0	0	1	0	0	1	615
$\mathbf{x}_7$	1	0	1	0	1	1	1	0	1	1	0	0	1	0	1	0	477
$\mathbf{x}_8$	1	0	1	0	0	0	0	1	1	0	0	1	1	1	0	0	263
$\mathbf{x}_9$	1	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	272
$\mathbf{x}_{10}$	1	1	1	1	0	0	0	1	1	0	0	1	1	1	0	0	247
$\mathbf{x}_{11}$	1	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	612
$\mathbf{x}_{12}$	1	0	1	1	1	0	0	0	1	0	0	1	1	0	1	0	510
$\mathbf{x}_{13}$	1	1	1	1	0	1	1	0	1	0	1	0	1	0	0	1	555
$\mathbf{x}_{14}$	1	1	0	0	1	0	1	1	1	0	0	1	1	1	0	0	264
$\mathbf{x}_{15}$	1	0	1	0	0	1	1	1	1	0	0	1	1	0	0	1	576
$\mathbf{x}_{16}$	1	0	1	0	1	0	1	1	1	0	1	0	1	1	0	0	268

Regression Trees (CART)



Sample mean

$$\ell_{X_{iL}} = \frac{1}{|X_{iL}|} \sum_{(\mathbf{x}_j, y_j) \in X_{iL}} y_j \qquad \ell_{X_{iR}} = \frac{1}{|X_{iR}|} \sum_{(\mathbf{x}_j, y_j) \in X_{iR}} y_j$$

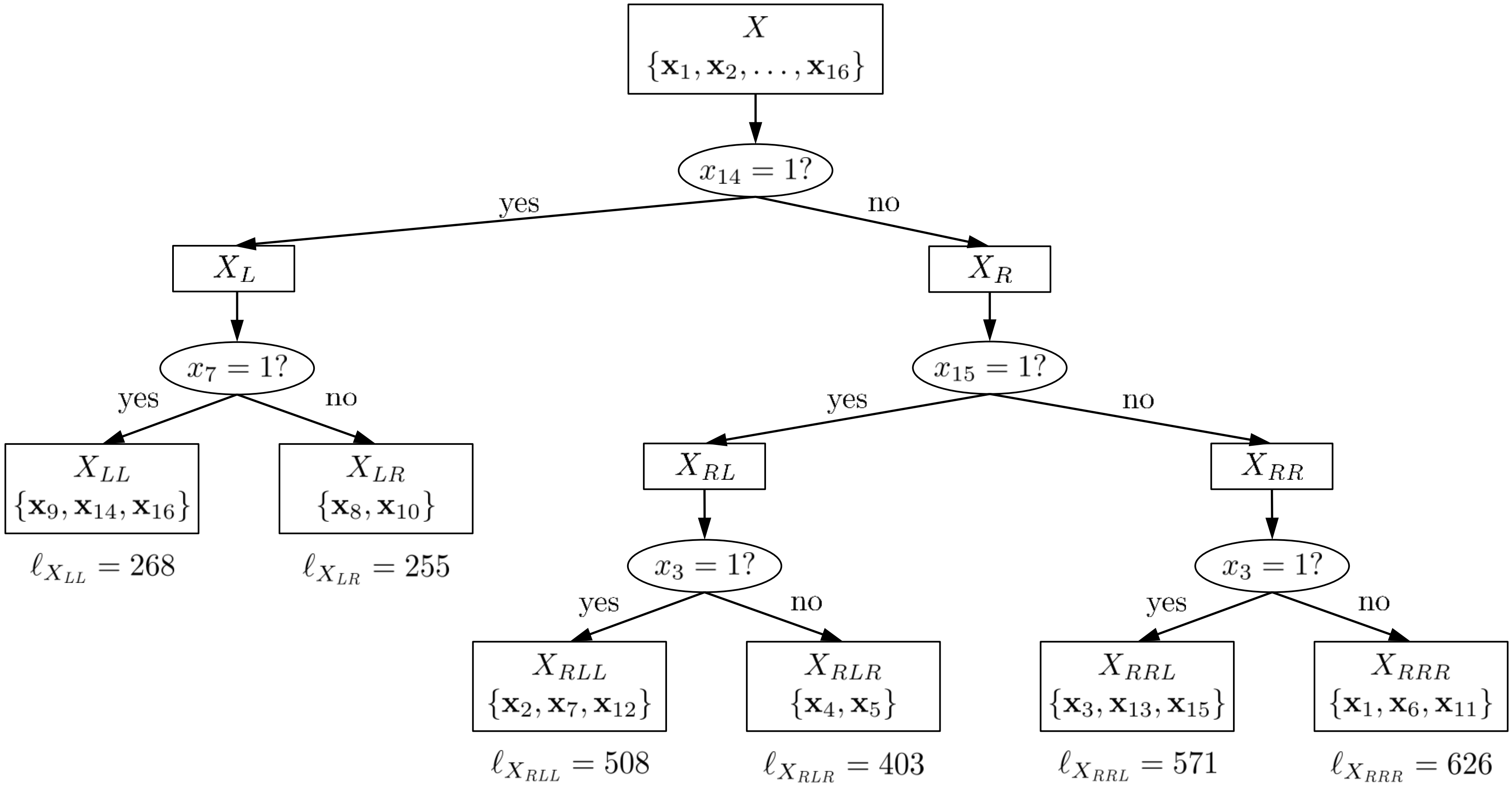
Squared error loss

$$\sum_{(\mathbf{x}_j, y_j) \in X_{iL}} (y_j - \ell_{X_{iL}})^2 \qquad \sum_{(\mathbf{x}_j, y_j) \in X_{iR}} (y_j - \ell_{X_{iR}})^2$$

Sum of squared error loss

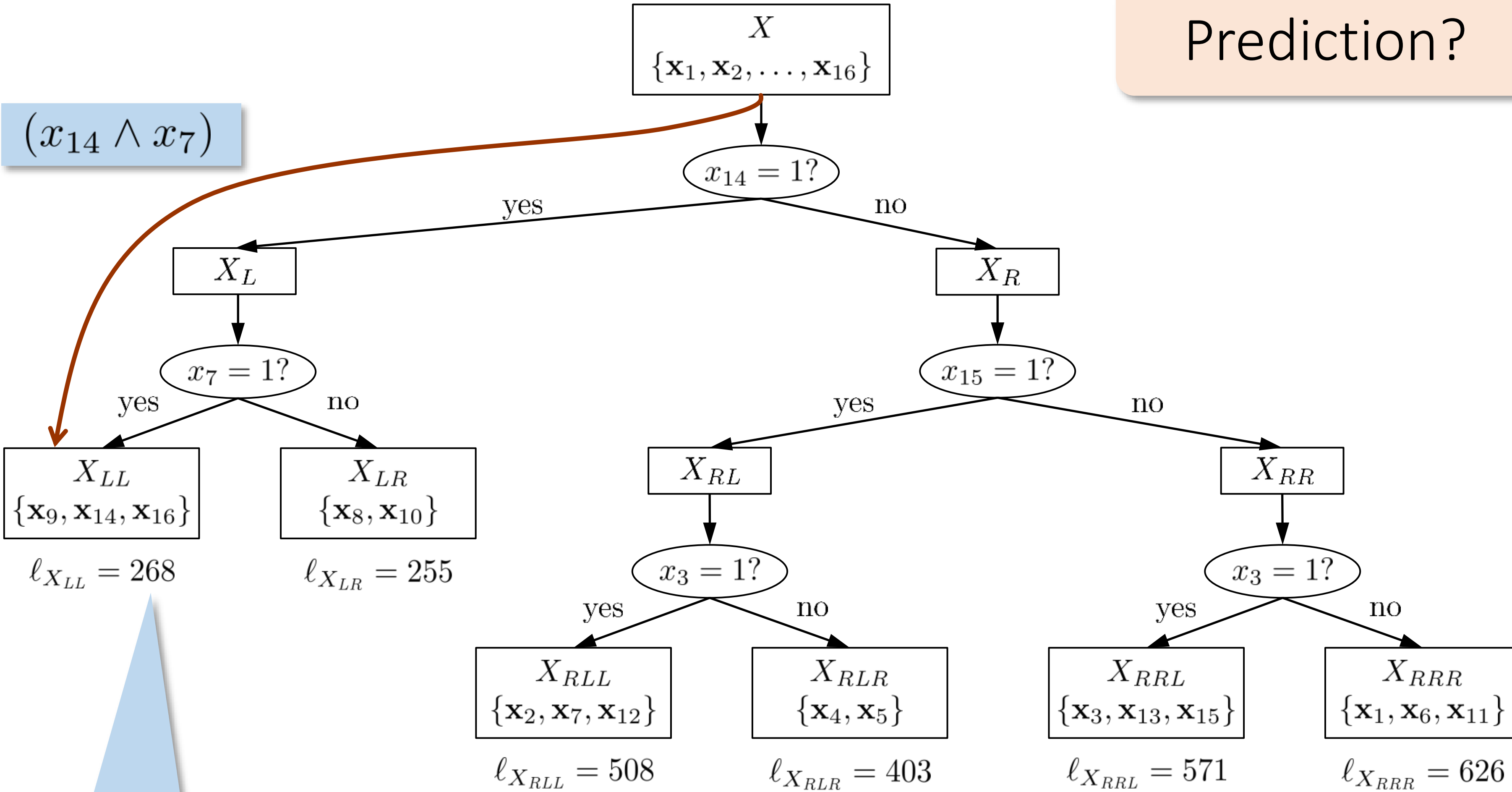
$$\sum_{(\mathbf{x}_j, y_j) \in X_{iL}} (y_j - \ell_{X_{iL}})^2 + \sum_{(\mathbf{x}_j, y_j) \in X_{iR}} (y_j - \ell_{X_{iR}})^2$$

is minimal!


X
$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{16}\}$
$x_{14} = 1?$
yes
no
X_L
X_R
$x_7 = 1?$
yes
no
$x_{15} = 1?$
yes
no
X_{LL}
$\{\mathbf{x}_9, \mathbf{x}_{14}, \mathbf{x}_{16}\}$
$\ell_{X_{LL}} = 268$
X_{LR}
$\{\mathbf{x}_8, \mathbf{x}_{10}\}$
$\ell_{X_{LR}} = 255$
X_{RL}
X_{RR}
$x_3 = 1?$
yes
no
$x_3 = 1?$
yes
no
X_{RLL}
$\{\mathbf{x}_2, \mathbf{x}_7, \mathbf{x}_{12}\}$
$\ell_{X_{RLL}} = 508$
X_{RLR}
$\{\mathbf{x}_4, \mathbf{x}_5\}$
$\ell_{X_{RLR}} = 403$
X_{RRL}
$\{\mathbf{x}_3, \mathbf{x}_{13}, \mathbf{x}_{15}\}$
$\ell_{X_{RRL}} = 571$
X_{RRR}
$\{\mathbf{x}_1, \mathbf{x}_6, \mathbf{x}_{11}\}$
$\ell_{X_{RRR}} = 626$

Prediction?

$(x_{14} \wedge x_7)$

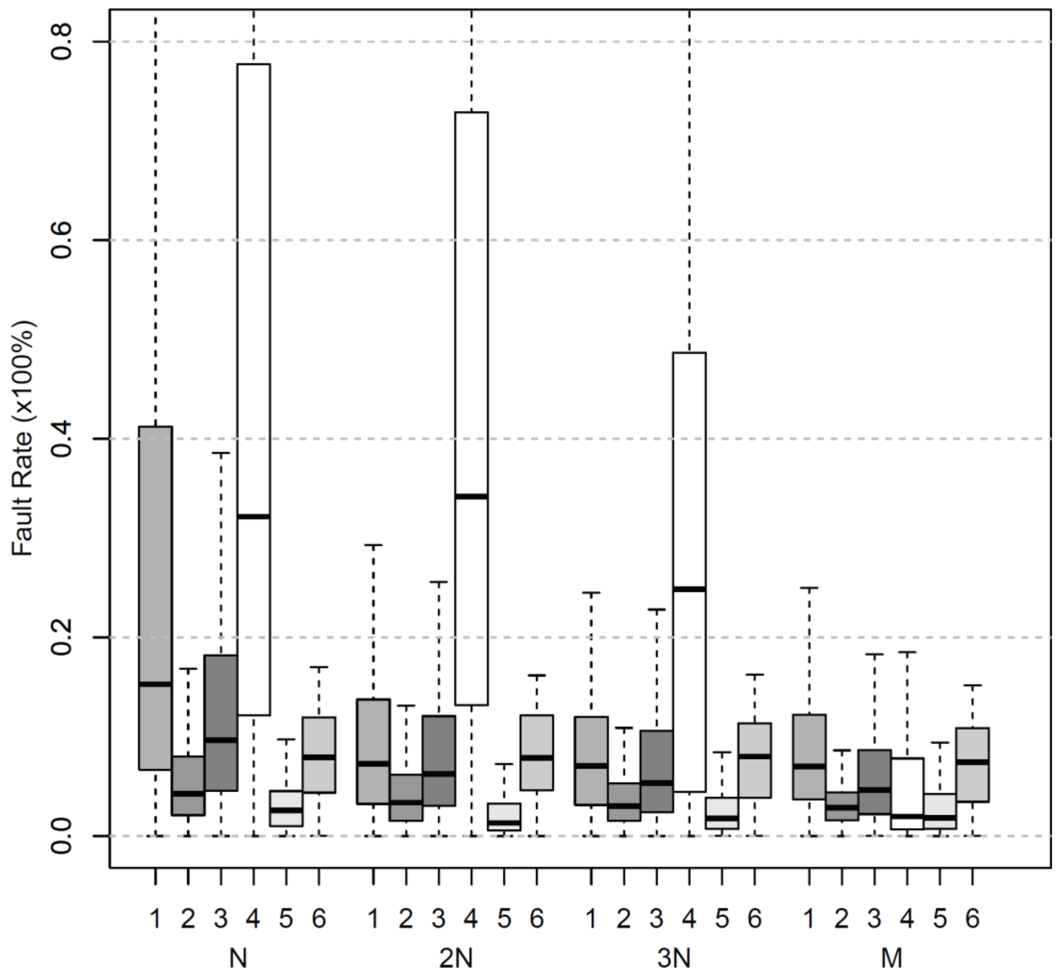


Predicted value
for any configuration
selecting feature x_{14}
and feature x_7

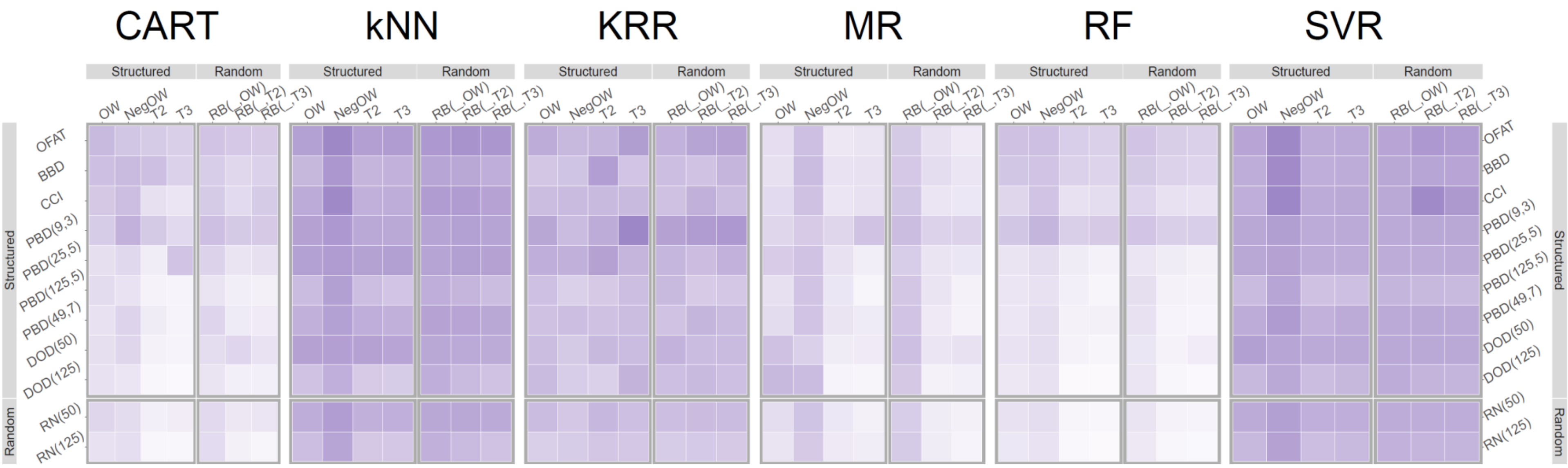
Empirical Results

	System	Domain	Lang.	LOC	$\|P\|$	N	M
1	APACHE	Web Server	C	230,277	192	9	29
2	LLVM	Compiler	C++	47,549	1,024	11	62
3	x264	Encoder	C	45,743	1,152	16	81
4	BERKELEY DB	Database	C	219,811	2,560	18	139
5	BERKELEY DB	Database	JAVA	42,596	400	26	48
6	SQLITE	Database	C	312,625	3,932,160	39	566

$N = |F|$ $M =$ Number of configurations required by pair-wise sampling



More Sampling & Learning Strategies



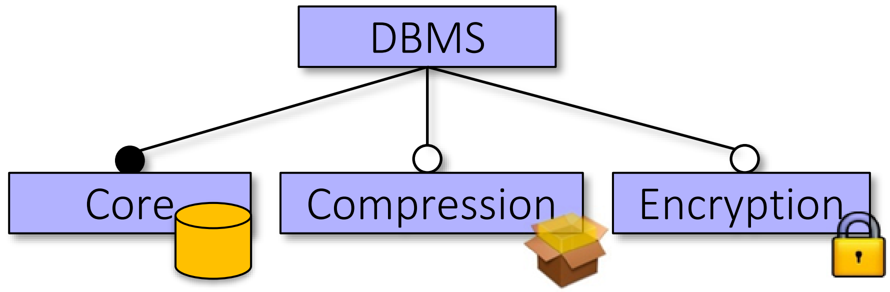
Learning strategies	
CART	Classification and Regression Tree
kNN	k-Nearest-Neighbors Regression
KRR	Kernel Ridge Regression
MR	Multiple Linear Regression
RF	Random Forest
SVR	Support Vector Regression

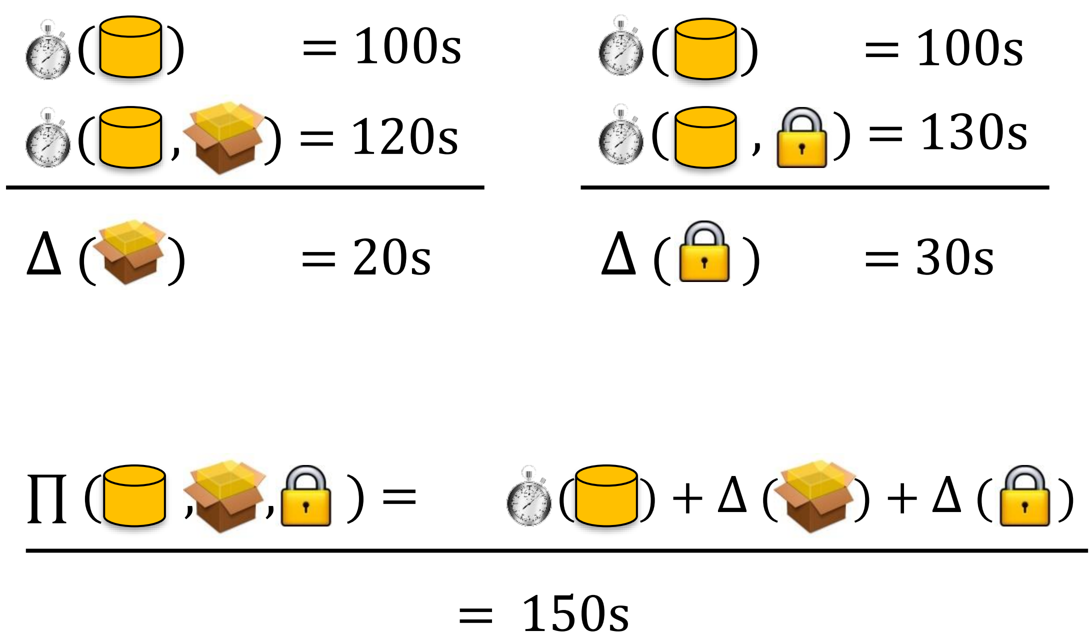
Sampling strategies	
OW	Option-wise
T*	T-wise
NewOW	Negative Option-wise
RB	Random (Binary)
OFAT	One Factor At A Time
BBD	Box-Behnken Design
CCI	Central Composite Inscribed Design
PBD	Plackett-Burman Design
DOD	D-Optimal Design
RN	Random (Numeric)

Part III

Non-Functional
Feature Interactions

Example


DBMS
Core
Compression
Encryption


= 100s
= 120s
Δ
= 20s
= 100s
= 130s
Δ
= 30s
∏
=
+ Δ
+ Δ
= 150s

Example

Π

Predicted performance

Measured performance

$\Pi(\ ,\ ,\ ) = (\ ) + \Delta(\ ) + \Delta(\ )$
$= 100s \quad + 20s \quad + 30s$
$= \mathbf{150s}$

$(\ ,\ ,\ ) = \mathbf{140s}$

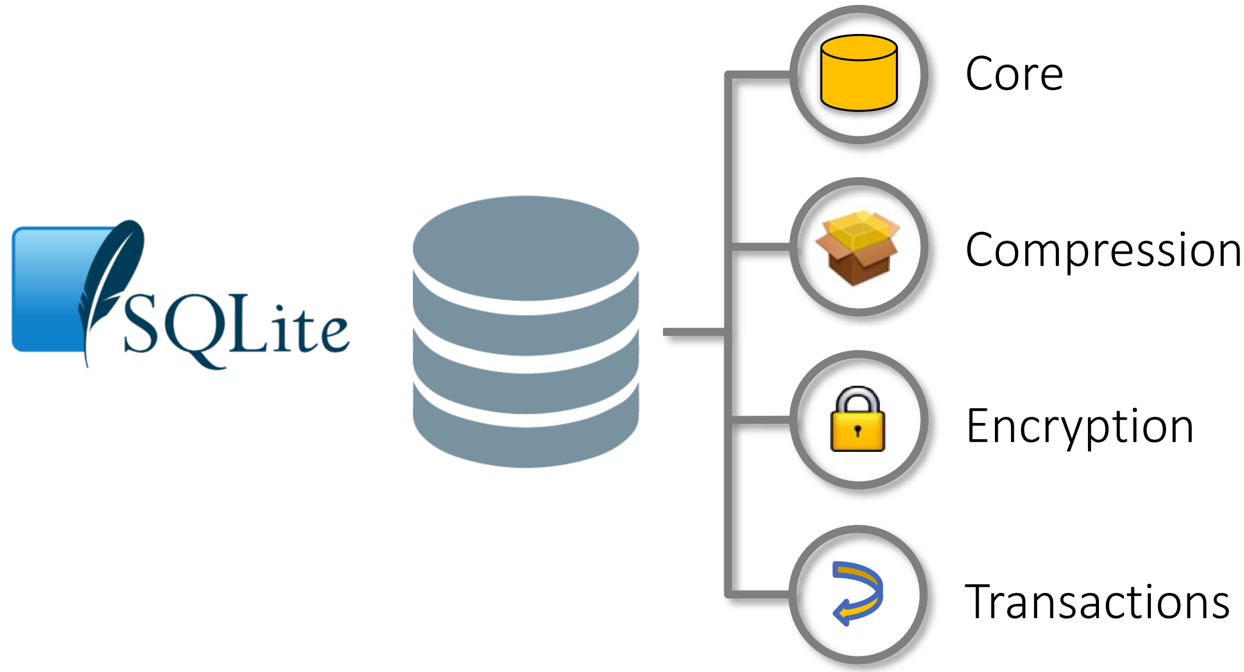
Feature interaction: due to encrypting compressed data

Definition

$$\frac{\begin{array}{c} f \notin c_{\min} \qquad f' \notin c_{\min} \\ \text{selectable}(\Psi, f) \qquad \text{selectable}(\Psi, f') \qquad \text{selectable}(\Psi, f \wedge f') \\ \Big| \big(\text{⏱}(c_{\min} \cup \{f\}) + \text{⏱}(c_{\min} \cup \{f'\})\big) - \text{⏱}\big(c_{\min} \cup \{f, f'\}\big) \Big| > \varepsilon \end{array}}{\text{interact}(f, f', \Psi)}$$

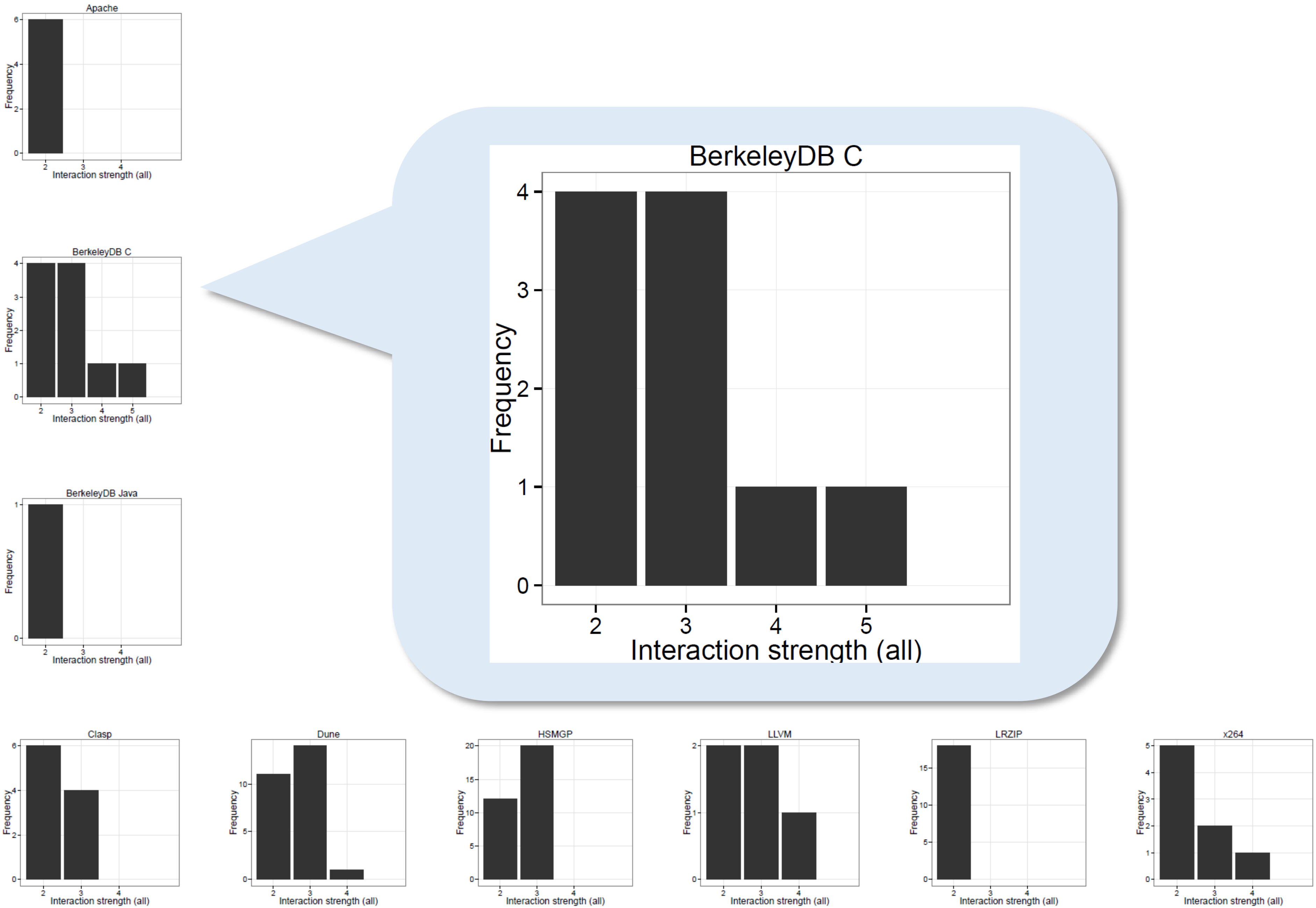
$c_{\min}$: Minimal configuration (minimal set of features)

Influence Models

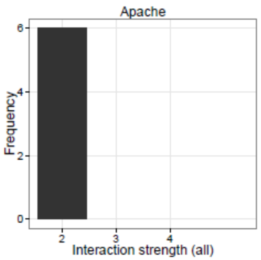


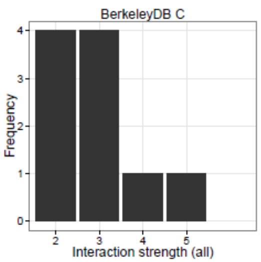
Feature Interaction!

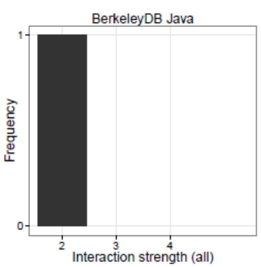
Performance Interactions



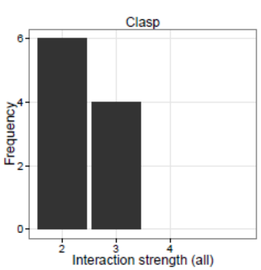
Performance Interactions

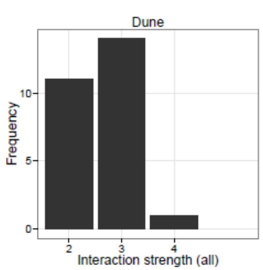


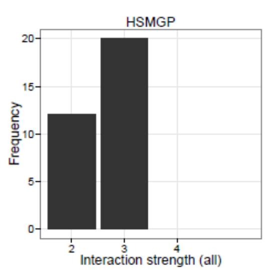


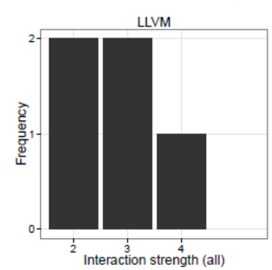


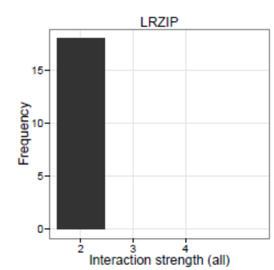
Two confirmed performance bugs in the LLVM extension Polly

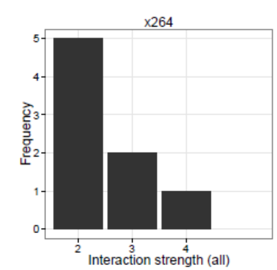




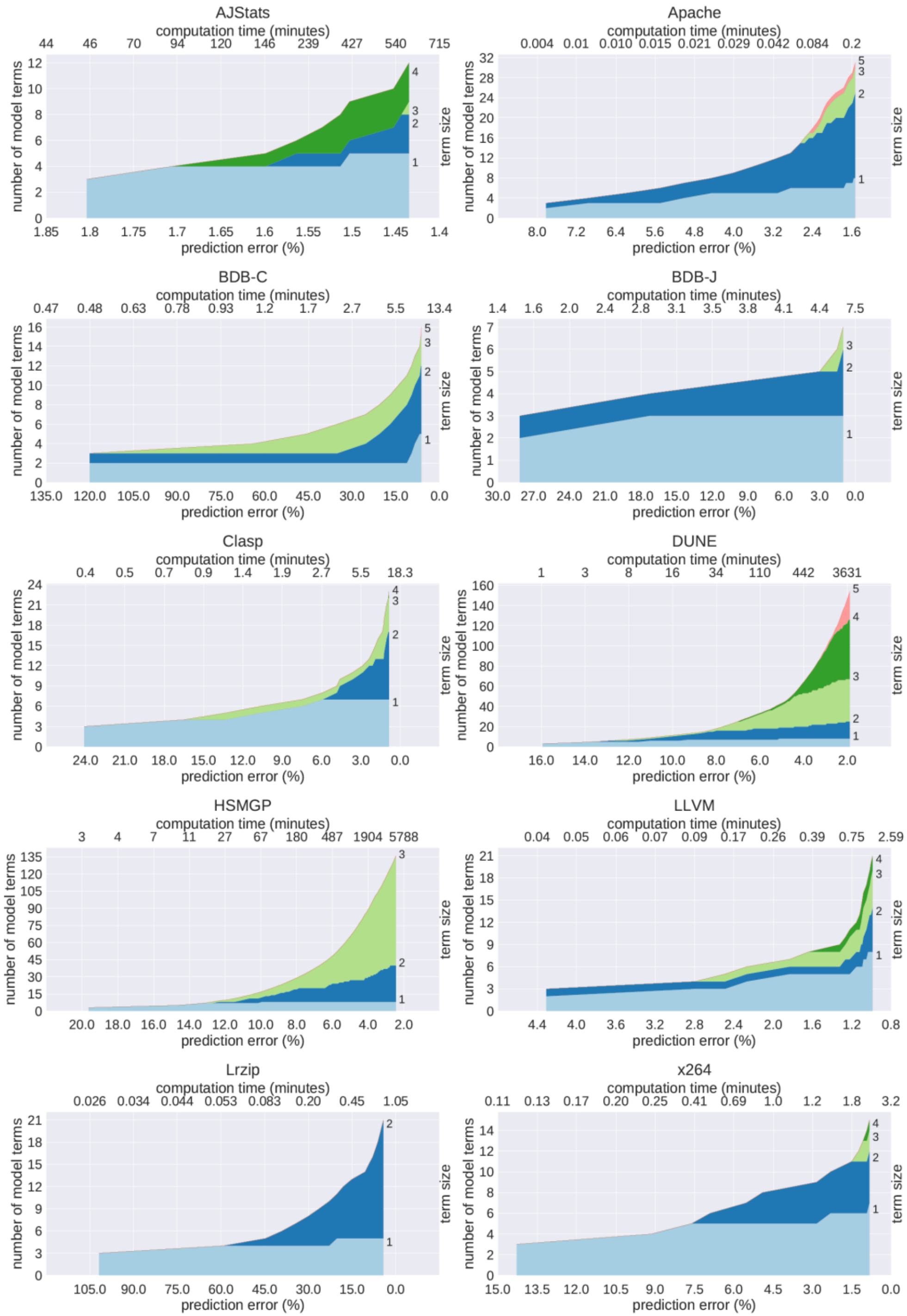








Feature Interactions vs. Prediction Error

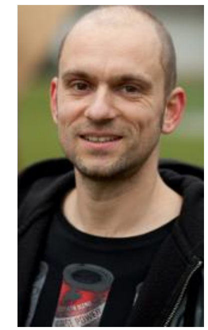


Comprehension

JUQUEEN – Jülich Blue Gene/Q

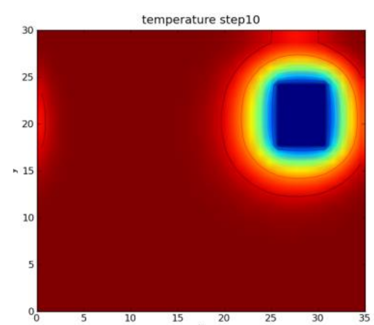
(458.752 Cores, 5.9 Petaflops)

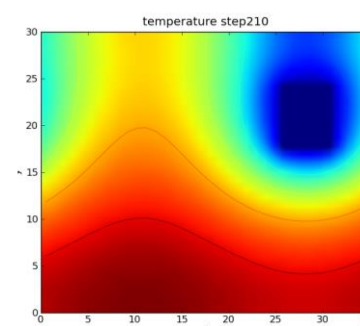
Matthias Bolten



Christian Engwer

$$\frac{\partial u}{\partial t} - \alpha \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) = 0$$

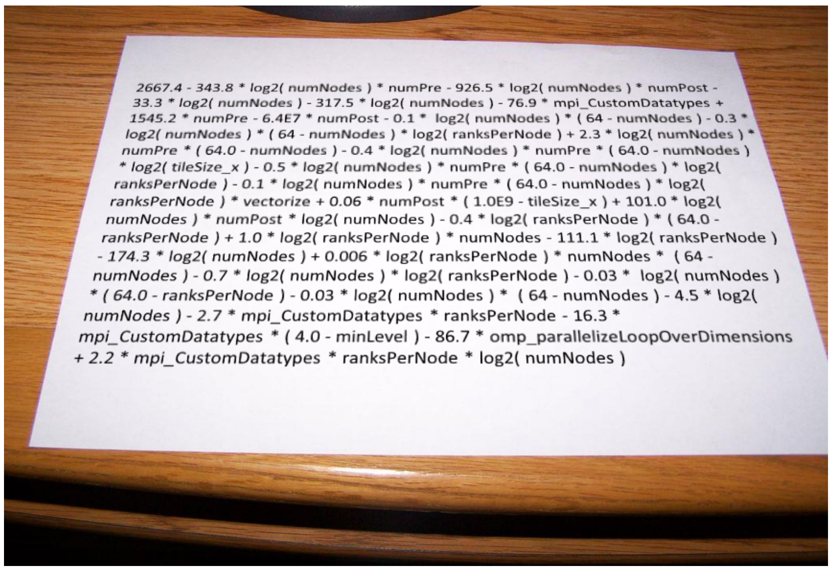




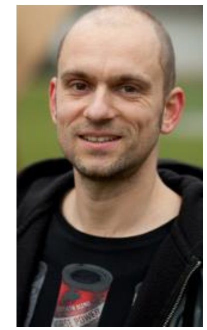
Comprehension

SPPEXA

ExaStencils

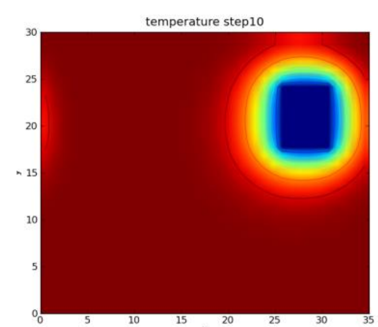


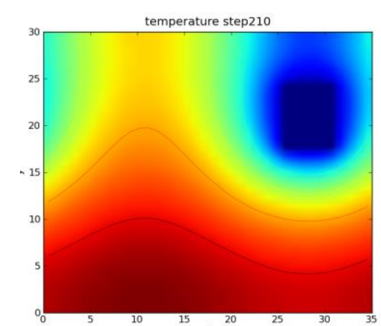
Matthias Bolten

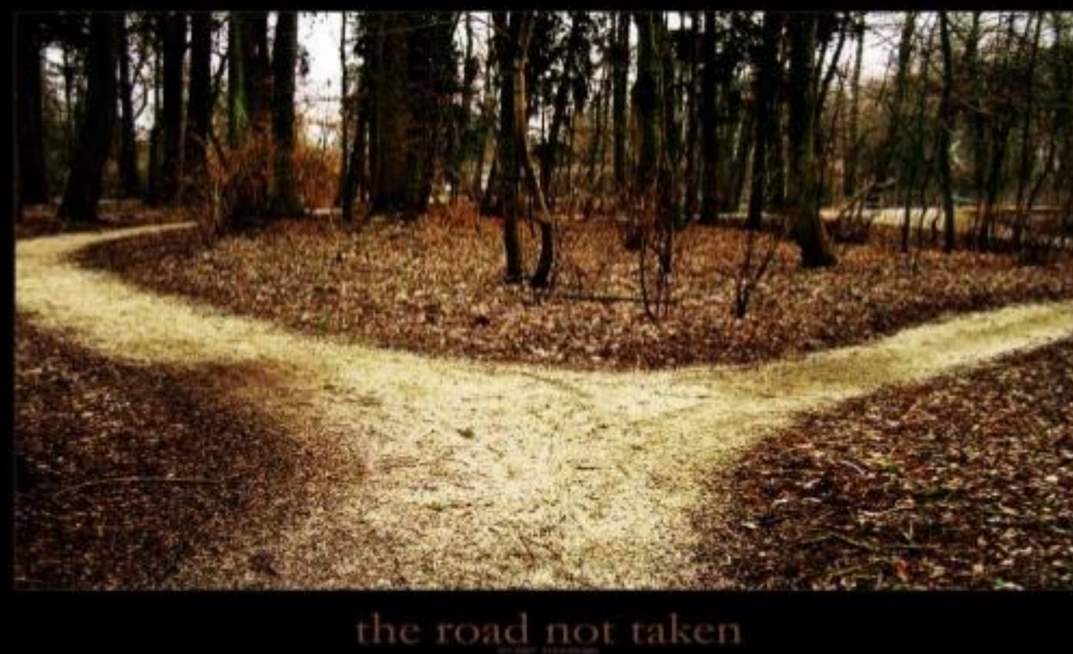


Christian Engwer

$$\frac{\partial u}{\partial t} - \alpha \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) = 0$$



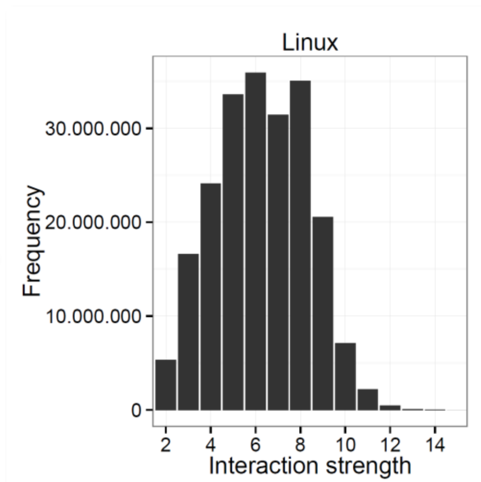




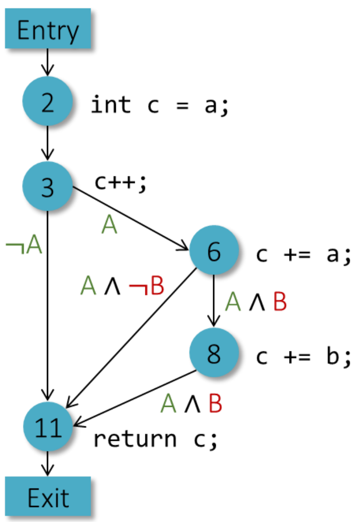
Two Roads Diverged In A Yellow Wood,
And Sorry I Could Not Travel Both.

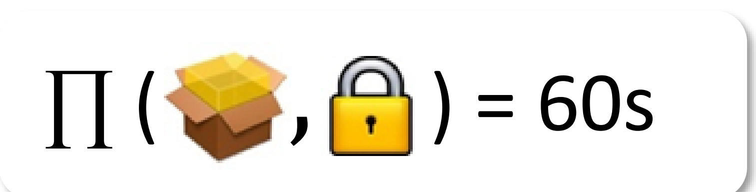
Perspectives

Properties & Distribution?



Correlation?





Prediction?

Literature

T. Xu, et al.: *Hey, you have given me too many knobs!: understanding and dealing with over-designed configuration in system software*. Proceedings ESEC/FSE. 307-319, ACM, 2015

N. Siegmund, et al.: *Performance-influence models for highly configurable systems*. Proceedings ESEC/FSE, 284-294, ACM, 2015

J. Guo, et al.: *Data-efficient performance learning for configurable systems*. EMSE 23(3): 1826-1867, 2018

C. Kaltenecker, et al. *The interplay of sampling and machine learning for software performance prediction*. IEEE Software, 2020

S. Kolesnikov, et al.: *Tradeoffs in modeling performance of highly configurable software systems*. SoSyM 18(3): 2265-2283, 2019

S. Apel, et al.: *Exploring feature interactions in the wild: The new feature-interaction challenge*. Proceedings FOSD, 1-8, ACM, 2013

Quiz

Model A:

$$1000 - 250 \cdot \textsf{AccessLog} - 150 \cdot \textsf{HostnameLookups}$$

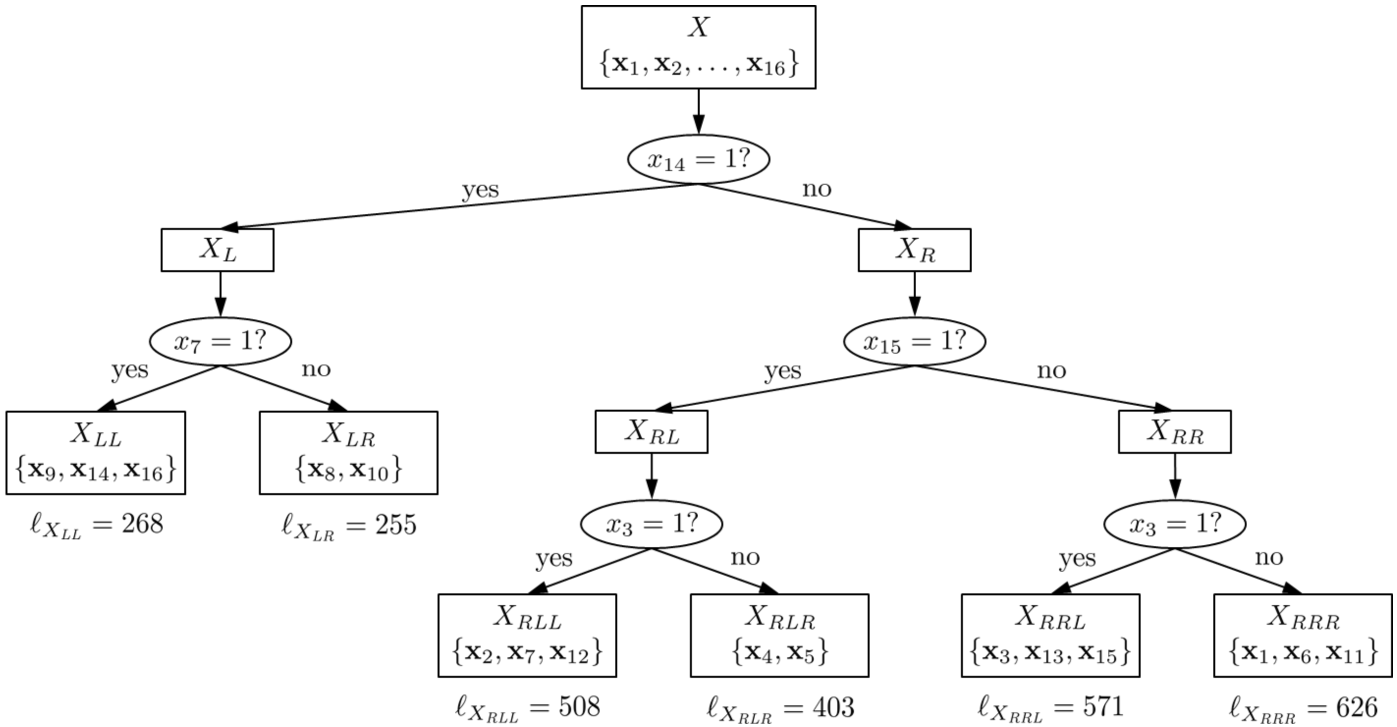
Model B:

$$\begin{aligned}&1000 - 250 \cdot \textsf{AccessLog} - 150 \cdot \textsf{HostnameLookups}\\&+100 \cdot \textsf{AccessLog} \cdot \textsf{HostnameLookups}\\&+2 \cdot \textsf{AccessLog} \cdot \textsf{EnableSendfile} \cdot \textsf{KeepAlive}\\&+1 \cdot \textsf{EnableSendfile} \cdot \textsf{FollowSymLinks} \cdot \textsf{Handle}\end{aligned}$$

Calculate the predicted performance:

Configuration	Measured	Prediction (Model A)	Prediction (Model B)
AccessLog	750		
HostnameLookups	850		
AccessLog, HostnameLookups	600		
AccessLog, EnableSendfile, KeepAlive	750		
AccessLog, EnableSendfile, Handle	750		
AccessLog, EnableSendfile, FollowSymLinks, Handle	750		

Quiz



Calculate the predicted performance:

Configuration	Prediction
$\neg x_3 \wedge \neg x_{14} \wedge x_{15}$	
$x_3 \wedge \neg x_7 \wedge \neg x_{14} \wedge \neg x_{15}$	
$x_7 \wedge x_{14} \wedge \neg x_{16}$	
$\mathbf{x}_4$	
$\mathbf{x}_8$	
$\mathbf{x}_{10}$	