What can be done with ML in Software Engineering research?

Paul Temple

ML and SE 2022-2023

Contact: paul.temple@irisa.fr

DiverSE
Diversity-Centric Software Engineering

About me

About me


MANCHE
Lannion
Saint-Malo
Morlaix
Guingamp
Brest
SAINT-BRIEUC
Dinan
Fougères
FINISTERE
29
CÔTES-D'ARMOR
22
ILLE-ET-VILAINE
35
RENNES
Pontivy
QUIMPER
MORBIHAN
56
Lorient
VANNES
Redon
OCEAN ATLANTIQUE

About me: Background

- 2008 – 2010: 1^{st} and 2^{nd} year study $\rightarrow$ computer science with an emphasis on data and image processing
- 2010 – 2013: Engineering school $\rightarrow$ computer science with an emphasis on data and image processing
- 2012 – 2013: Master's Degree on data and image processing

About me: Background

- 2008 – 2010: 1^{st} and 2^{nd} year study $\rightarrow$ computer science with an emphasis on data and image processing
- 2010 – 2013: Engineering school $\rightarrow$ computer science with an emphasis on data and image processing
- 2012 – 2013: Master's Degree on data and image processing
- 2014 – 2015: Lab engineer in biometrics & epayment team (Caen)

About me: Background

- 2008 – 2010: 1^{st} and 2^{nd} year study $\rightarrow$ computer science with an emphasis on data and image processing
- 2010 – 2013: Engineering school $\rightarrow$ computer science with an emphasis on data and image processing
- 2012 – 2013: Master's Degree on data and image processing
- 2014 – 2015: Lab engineer in biometrics & epayment team (Caen)
- 2015 – 2018: PhD in Software Engineering $\rightarrow$ ML with software variability

About me: Background

- 2008 – 2010: 1^{st} and 2^{nd} year study $\rightarrow$ computer science with an emphasis on data and image processing
- 2010 – 2013: Engineering school $\rightarrow$ computer science with an emphasis on data and image processing
- 2012 – 2013: Master's Degree on data and image processing
- 2014 – 2015: Lab engineer in biometrics & epayment team (Caen)
- 2015 – 2018: PhD in Software Engineering $\rightarrow$ ML with software variability
- 2019 – 2022: Post-doc in Namur $\rightarrow$ EoS VeriLearn project: SE for ML and ML for SE
- 2022 – ...: Associate Professor in DiverSE at UnivRennes $\rightarrow$ SE for ML and ML for SE

About me: Master's Internship

- February 2013 – May 2013
- Topic: Security of Machine Learning Processes applied to Multimedia
- Supervisors:

Laurent Amsaleg:

Ewa Kijak:

About me: PhD

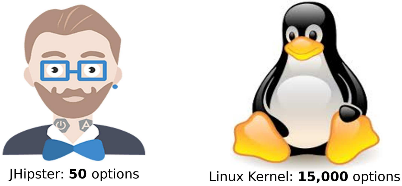
- Sept 2015 – Dec 2018
- Topic: Investigate the Matrix: Leveraging variability ...
- Supervisors:

Jean-Marc Jézéquel:

Mathieu Acher:

About me: PhD

Software Variability



$2^{15,000} \approx 10^{3,250} >> 10^{1,000} >>$ estimated # of particules

About me: PhD

options:

- no-mbtree (T or F)
- nr ([100..1000])
- qblur ([0; 1])

step = 0.0001

→ **18 millions** of configurations

⇒

About me: PhD

options:

- no-mbtree (T or F)
- nr ([100..1000])
- qblur ([0; 1])

step = 0.0001

$\rightarrow$ **18 millions** of configurations

$\Rightarrow$

no-mbtree = T
qblur = 0.33
nr = 400

no-mbtree = F
qblur = 0.90
nr = 200

About me: PhD

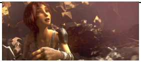
no-mbtree = F
qblur = 0.90
nr = 200

encoding time = 5 min

encoding time = 2 h

encoding time = 10 h

About me: PhD

	Program Variants				
				...	
Inputs		12	1	...	5
		1	348	...	10
	...				
		50	101	...	260

Can we and how to reduce either of those dimensions?

About me: current interests

Performance prediction of configurable systems

- Try to predict the performance based on the configuration
- Structure the configuration space for optimizing

About me: current interests

Performance prediction of configurable systems

- Try to predict the performance based on the configuration
- Structure the configuration space for optimizing

$\Rightarrow$ Use and adapt ML/DL techniques

About me: current interests

Performance prediction of configurable systems

- Try to predict the performance based on the configuration
- Structure the configuration space for optimizing

$\Rightarrow$ Use and adapt ML/DL techniques

Testing ML models

- Testing in the SE sense (check the behavior)
- Various performance measures (fairness, energy consumption, etc.)

About me: current interests

Performance prediction of configurable systems

- Try to predict the performance based on the configuration
- Structure the configuration space for optimizing

$\Rightarrow$ Use and adapt ML/DL techniques

Testing ML models

- Testing in the SE sense (check the behavior)
- Various performance measures (fairness, energy consumption, etc.)

$\Rightarrow$ Use and adapt SE testing techniques

About me: current interests

Performance prediction of configurable systems

- Try to predict the performance based on the configuration
- Structure the configuration space for optimizing

$\Rightarrow$ Use and adapt ML/DL techniques

Testing ML models

- Testing in the SE sense (check the behavior)
- Various performance measures (fairness, energy consumption, etc.)

$\Rightarrow$ Use and adapt SE testing techniques

ML models within systems

- ML models are part of systems
- ML models have a direct impact on the systems

About me: current interests

Performance prediction of configurable systems

- Try to predict the performance based on the configuration
- Structure the configuration space for optimizing

$\Rightarrow$ Use and adapt ML/DL techniques

Testing ML models

- Testing in the SE sense (check the behavior)
- Various performance measures (fairness, energy consumption, etc.)

$\Rightarrow$ Use and adapt SE testing techniques

ML models within systems

- ML models are part of systems
- ML models have a direct impact on the systems

$\Rightarrow$ how to safely update ML models?

About DiverSE

Diversity-centric Software Engineering (DiverSE)

- Modeling (of systems)
- Languages
- Evolution and Maintenance of systems
- Testing
- Security (malware, web, DevOps, ...?)
- Privacy
- ...

`https://www.diverse-team.fr/`

About DiverSE

Diversity-centric Software Engineering (DiverSE)

- Modeling (of systems)
- Languages
- Evolution and Maintenance of systems
- Testing
- Security (malware, web, DevOps, ...?)
- Privacy
- ...

https://www.diverse-team.fr/
https://www.diverse-team.fr/team/

About DiverSE

Diversity-centric Software Engineering (DiverSE)

- Modeling (of systems)
- Languages
- Evolution and Maintenance of systems
- Testing
- Security (malware, web, DevOps, ...?)
- Privacy
- ...

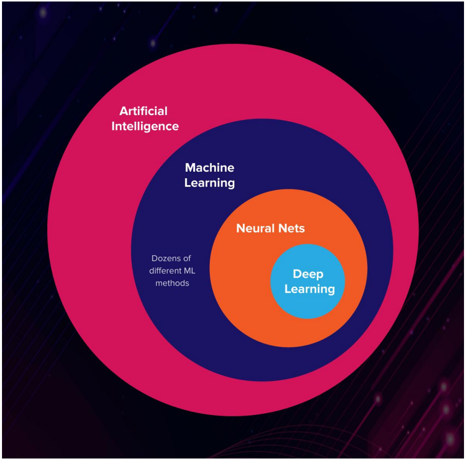
https://www.diverse-team.fr/
https://www.diverse-team.fr/team/
https://www.diverse-team.fr/positions/

Back to business
(ML and SE)

Intro: IA VS ML VS DL

Source: `https://serokell.io/blog/ai-ml-dl-difference`

Intro: IA VS ML VS DL



Source: `https://serokell.io/blog/ai-ml-dl-difference`

A gentle introduction to DL by Patrick Pérez

Patrick Pérez in short

- DR Inria in Rennes

A gentle introduction to DL by Patrick Pérez

Patrick Pérez in short

- DR Inria in Rennes
- Moved to industry (Technicolor)

A gentle introduction to DL by Patrick Pérez

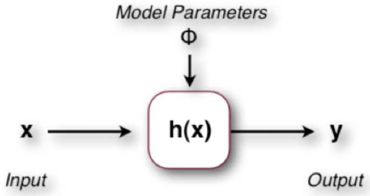
Patrick Pérez in short

- DR Inria in Rennes
- Moved to industry (Technicolor)
- Scientific director at Valeo.ai (and VP AI)

A gentle introduction to DL by Patrick Pérez

Patrick Pérez in short

- DR Inria in Rennes
- Moved to industry (Technicolor)
- Scientific director at Valeo.ai (and VP AI)



Source:

https://www.researchgate.net/publication/230691317_Gesture_Recognition_for_Musician_Computer_Interaction

Intro: Evaluate the performance of a ML model

How do you assess if your model performs well?

Intro: Evaluate the performance of a ML model

How do you assess if your model performs well?

Confusion matrix

		Oracle (Y)	
		0	1
Prediction ML model ($\hat{Y}$)	0	234 ($\pm$57.899)	69.5 ($\pm$26.973)
	1	141.1 ($\pm$60.440)	3566.2 ($\pm$25.804)

Intro: Feature descriptors

ML algorithms **does not work on raw data**

Tabular data

Explicit or Implicit? On Feature Engineering for ML-based Variability-intensive Systems, Temple and Perrouin, VaMoS'23

ML algorithms **does not work on raw data**

Tabular data

- Iris dataset → does not report pictures of irises
- Measure 4 characteristics

Explicit or Implicit? On Feature Engineering for ML-based Variability-intensive Systems, Temple and Perrouin, VaMoS'23

Intro: Feature descriptors

ML algorithms **does not work on raw data**

Tabular data

- Iris dataset $\rightarrow$ does not report pictures of irises
- Measure 4 characteristics
- Characteristics are defined by domain experts

Explicit or Implicit? On Feature Engineering for ML-based Variability-intensive Systems, Temple and Perrouin, VaMoS'23

ML algorithms **does not work on raw data**

Image data

Explicit or Implicit? On Feature Engineering for ML-based Variability-intensive Systems, Temple and Perrouin, VaMoS'23

Intro: Feature descriptors

ML algorithms **does not work on raw data**

Image data

- Describe images based on changes of textures and contrasts
- A few descriptors:

Explicit or Implicit? On Feature Engineering for ML-based Variability-intensive Systems, Temple and Perrouin, VaMoS'23

Intro: Feature descriptors

ML algorithms **does not work on raw data**

Image data

- Describe images based on changes of textures and contrasts
- A few descriptors:
  - Harris corner

Explicit or Implicit? On Feature Engineering for ML-based Variability-intensive Systems, Temple and Perrouin, VaMoS'23

ML algorithms **does not work on raw data**

Image data

- Describe images based on changes of textures and contrasts
- A few descriptors:
  - Harris corner
  - SIFT
  - SURF

Explicit or Implicit? On Feature Engineering for ML-based Variability-intensive Systems, Temple and Perrouin, VaMoS'23

Intro: Feature descriptors

ML algorithms **does not work on raw data**

Image data

- Describe images based on changes of textures and contrasts
- A few descriptors:
  - Harris corner
  - SIFT
  - SURF
  - Visual Bag of Words
- All aggregate different information

Explicit or Implicit? On Feature Engineering for ML-based Variability-intensive Systems, Temple and Perrouin, VaMoS'23

Intro: Feature descriptors

ML algorithms **does not work on raw data**

Image data

- Describe images based on changes of textures and contrasts
- A few descriptors:
  - Harris corner
  - SIFT
  - SURF
  - Visual Bag of Words
- All aggregate different information
- Analyse (macro-)blocks of pixels

Explicit or Implicit? On Feature Engineering for ML-based Variability-intensive Systems, Temple and Perrouin, VaMoS'23

ML algorithms **does not work on raw data**

Audio data

Explicit or Implicit? On Feature Engineering for ML-based Variability-intensive Systems, Temple and Perrouin, VaMoS'23

Intro: Feature descriptors

ML algorithms **does not work on raw data**

Audio data

- Ups and downs
- Composition of sinusoidal waves

Explicit or Implicit? On Feature Engineering for ML-based Variability-intensive Systems, Temple and Perrouin, VaMoS'23

Intro: Feature descriptors

ML algorithms **does not work on raw data**

Audio data

- Ups and downs
- Composition of sinusoidal waves
- MFCC

$\rightarrow$ Change of representation space

Explicit or Implicit? On Feature Engineering for ML-based Variability-intensive Systems, Temple and Perrouin, VaMoS'23

Configurable systems? A few examples...

Intro: SPLs and Variability

Configurable systems? A few examples...

Intro: SPLs and Variability

Configurable systems? A few examples...

Intro: SPLs and Variability

Configurable systems? A few examples...

Intro: SPLs and Variability

Configurable systems? A few examples...

Intro: SPLs and Variability

What is variability?

- Software variability $\rightarrow$ ability to vary =)

Svahnberg *et al.*, A taxonomy of variability realization techniques: Research Articles, Soft. Pract. Exper. 2005

What is variability?

- Software variability $\rightarrow$ ability to vary =)
- Ability to be efficiently extended, changed, customized, or configured for use in a particular context

Svahnberg *et al.*, A taxonomy of variability realization techniques: Research Articles, Soft. Pract. Exper. 2005

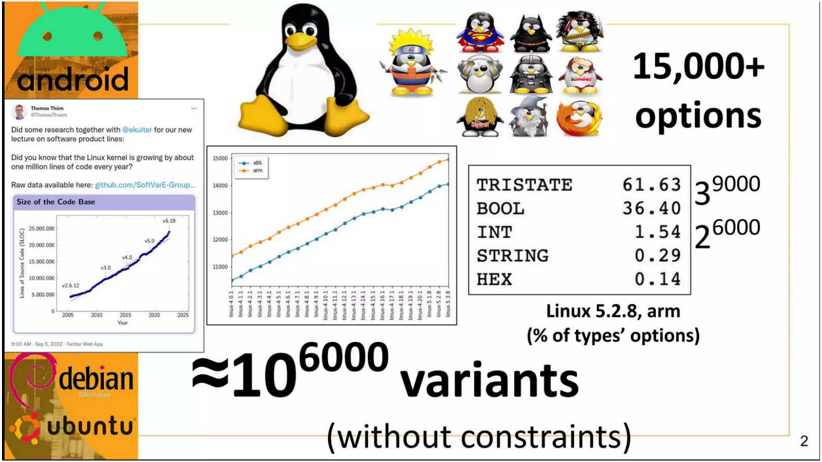
Intro: SPLs and Variability

What is variability?

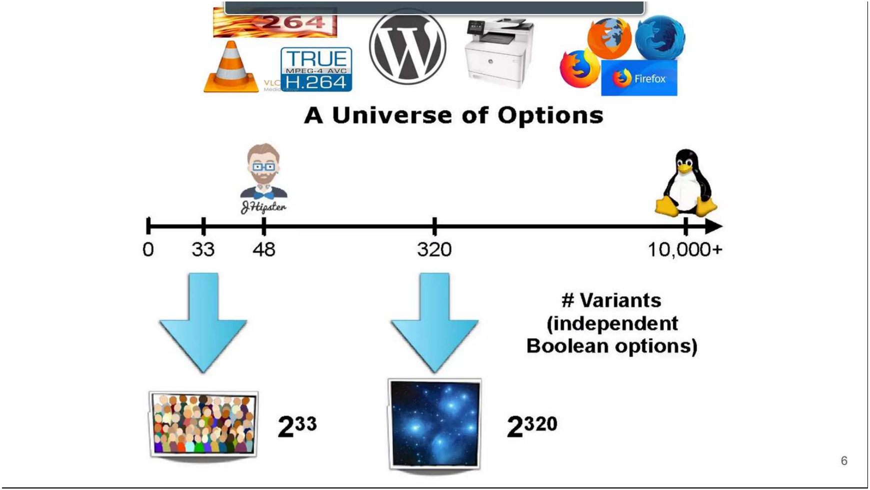
- Software variability $\rightarrow$ ability to vary =)
- Ability to be efficiently extended, changed, customized, or configured for use in a particular context
- Expected benefits
  - Lower production cost
  - Lower certification cost
  - Reduce time-to-market

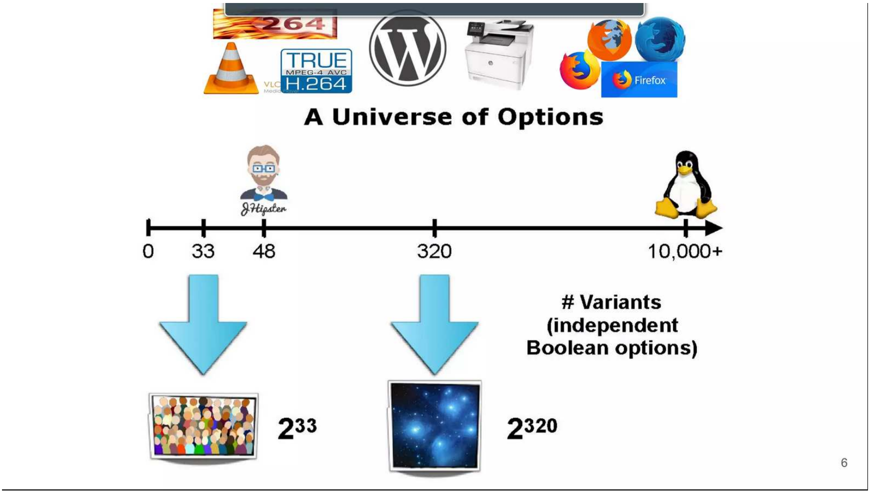
Svahnberg *et al.*, A taxonomy of variability realization techniques: Research Articles, Soft. Pract. Exper. 2005

Intro: SPLs and Variability



Intro: SPLs and Variability





Sébastien Mosser @petitroll · 25 févr.

"the number of atoms in the visible universe is 10^80. There are 2^15000 different versions of the Linux kernel. So astrophysicists works with things way simpler than software engineers". @jmjezequel options:

- no-mbtree (T or F)
- nr ([100..1000])
- qblur ([0; 1])

step = 0.0001

→ **18 millions** of configurations

⇒

no-mbtree = T
qblur = 0.33
nr = 400

no-mbtree = F
qblur = 0.90
nr = 200

Intro: SPLs and Variability

	no-mbtree	qbur	nr	exec time (s)
	T	0.33	400	24
	F	0.90	200	7

Intro: SPLs and Variability

	no-mbtree	qbur	nr	exec time (s)
	T	0.33	400	24
	F	0.90	200	7
	T	0.90	400	

How much time for the last configuration?

Intro: SPLs and Variability

	no-mbtree	qbur	nr	exec time (s)
	T	0.33	400	24
	F	0.90	200	7
	T	0.90	400	

How did we get the exeuction times?

Intro: SPLs and Variability

	no-mbtree	qbur	nr	exec time (s)
	T	0.33	400	24
	F	0.90	200	7
	T	0.90	400	

How did we get the exeuction times?

- Select a value for each option
- Generate/Build the configuration
- Run the system with a (or several) inputs
- Measure time for the execution

Intro: SPLs and Variability

	no-mbtree	qbur	nr	exec time (s)
264	T	0.33	400	24
264	F	0.90	200	7
264	T	0.90	400	

How did we get the exeuction times?

- Select a value for each option
- Generate/Build the configuration
- Run the system with a (or several) inputs
- Measure time for the execution

$\Rightarrow$ How to **measure without measuring?**

...

Intro: SPLs and Variability

	no-mbtree	qbur	nr	exec time (s)
	T	0.33	400	24
	F	0.90	200	7
	T	0.90	400	

How did we get the exeuction times?

- Select a value for each option
- Generate/Build the configuration
- Run the system with a (or several) inputs
- Measure time for the execution

$\Rightarrow$ How to **measure without measuring?**

...

(and **measure without building and running**)

Performance prediction

- Solely based on values of options
- "Guess" the value when executed

Intro: SPLs and Variability

Performance prediction

- Solely based on values of options
- "Guess" the value when executed
  - No builds
  - No executions

Performance prediction

- Solely based on values of options
- "Guess" the value when executed
  - No builds
  - No executions
- Generalize to unseen configurations

Performance prediction

- Solely based on values of options
- "Guess" the value when executed
  - No builds
  - No executions
- Generalize to unseen configurations

- Performance models
- Machine Learning (ML) models

Performance models

$Performance_{config}$ = 0.241 * no-mbtree - 0.316 * nr - 0.024 * qblur + 0.201 * no-mbtree x qblur + 0.129 no-mbtree x nr - ...

Siegmund *et al.*, Performance-influence models for highly configurable systems, ESEC/FSE 2015

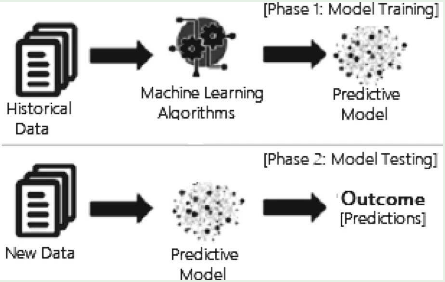
Performance models

$Performance_{config}$ = 0.241 * no-mbtree - 0.316 * nr - 0.024 * qblur + 0.201 * no-mbtree x qblur + 0.129 no-mbtree x nr - ...

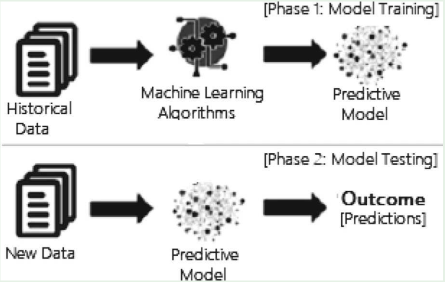
- Linear model
- More interactions → More precise
- Start with the minimal configuration and add influence of (tuple of) options

Siegmund *et al.*, Performance-influence models for highly configurable systems, ESEC/FSE 2015

ML models

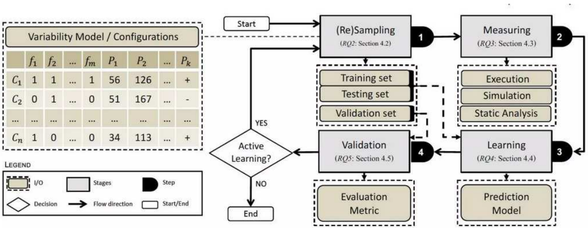

[Phase 1: Model Training]
Historical Data
Machine Learning Algorithms
Predictive Model
[Phase 2: Model Testing]
New Data
Predictive Model
Outcome
[Predictions]

ML models



- Can be more complex than a linear combination
- Start with a set of configurations (need to be measured)
- Can be interpretable (e.g., decision trees)

Sampling, Measuring, Learning



Learning Software Configuration Spaces: A Systematic Literature Review
Juliana Alves Pereira, Hugo Martin, Mathieu Acher, Jean-Marc Jézéquel, Goetz Botterweck, Anthony Ventresque https://arxiv.org/abs/1906.03018

Why performance prediction over configurable systems?

- Predict performance...

Martin *et al.*, Learning very large configuration spaces: What matters for linux kernel sizes, https://inria.hal.science/hal-02314830/, 2019

Martin *et al.*, Learning from thousands of build failures of Linux kernel configurations, https://inria.hal.science/hal-02147012/, 2019

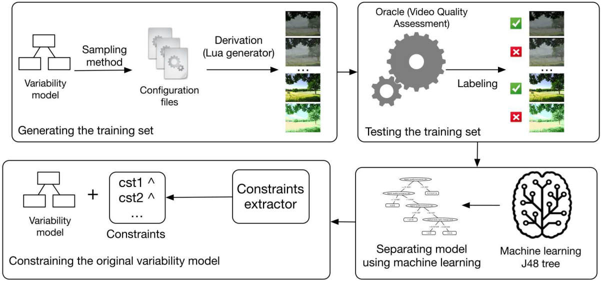
Why performance prediction over configurable systems?

- Predict performance... of unseen configurations
- Understand more things about the system (constraints, domain knowledge)
- Optimize
- Specialize the system
- ...

Martin *et al.*, Learning very large configuration spaces: What matters for linux kernel sizes, `https://inria.hal.science/hal-02314830/`, 2019

Martin *et al.*, Learning from thousands of build failures of Linux kernel configurations, `https://inria.hal.science/hal-02147012/`, 2019

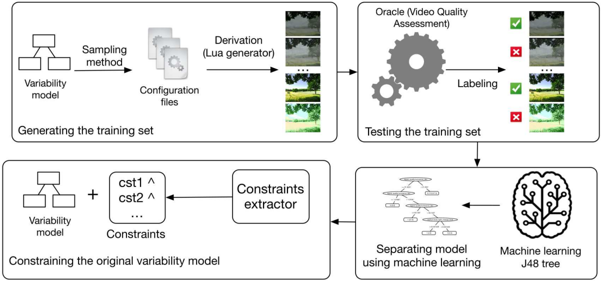
Specialize the system



Temple *et al.*, Using machine learning to infer constraints for product lines, SPLC 2016

Martin *et al.*, A comparison of performance specialization learning for configurable systems, SPLC 2021

Specialize the system



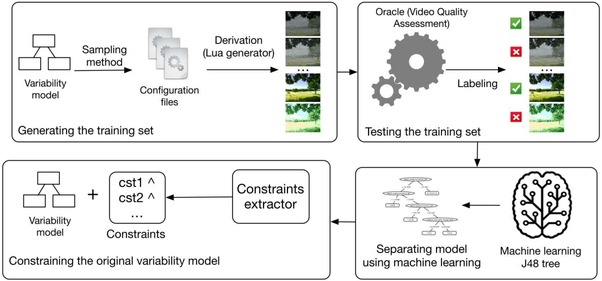
To keep or not to keep?

- (Binary) Classification problem

Temple *et al.*, Using machine learning to infer constraints for product lines, SPLC 2016

Martin *et al.*, A comparison of performance specialization learning for configurable systems, SPLC 2021

Specialize the system



To keep or not to keep?

- (Binary) Classification problem
- Regression → discretize → n-class problem

Temple *et al.*, Using machine learning to infer constraints for product lines, SPLC 2016

Martin *et al.*, A comparison of performance specialization learning for configurable systems, SPLC 2021

How to represent configurations for ML?

	no-mbtree	qbur	nr	exec time (s)
	T	0.33	400	24
	F	0.90	200	7
	T	0.90	400	

- Need a **homogeneous representation**

How to represent configurations for ML?

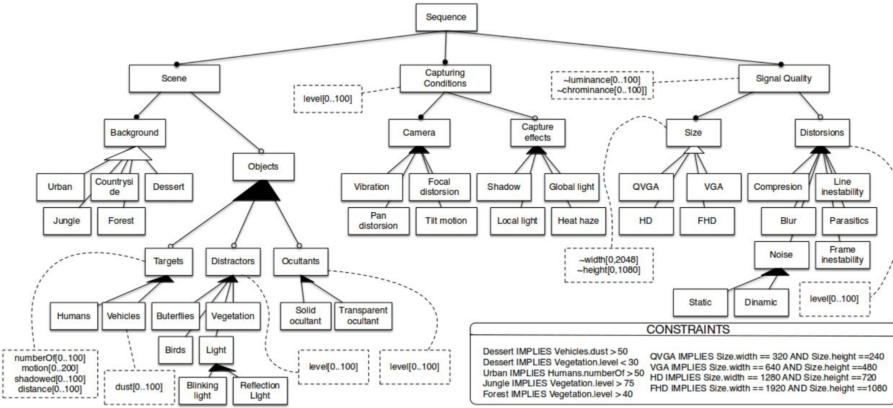


Figure 5: Feature model, excerpt of the model used in the MOTIV project, to represent variability (through feature attributes) of a video sequence.

How to represent configurations for ML?

	no-mbtree	qbur	nr	exec time (s)
	T	0.33	400	24
	F	0.90	200	7
	T	0.90	400	

- Need a **homogeneous representation**

How to represent configurations for ML?

	no-mbtree	qbur	nr	exec time (s)
	T	0.33	400	24
	F	0.90	200	7
	T	0.90	400	

- Need a **homogeneous representation**
- Configuration $\rightarrow$ set of options $\rightarrow$ vector
  Need all options (selected or not)

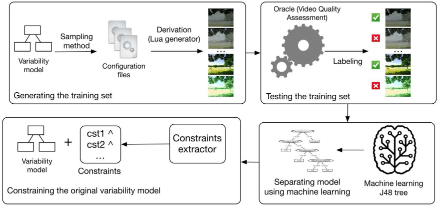
How to represent configurations for ML?

	no-mbtree	qbur	nr	exec time (s)
	T	0.33	400	24
	F	0.90	200	7
	T	0.90	400	

- Need a **homogeneous representation**
- Configuration $\rightarrow$ set of options $\rightarrow$ vector
  Need all options (selected or not)
- Heterogenous features in a homogeneous representation
  Feature importance?

Maybe we are wrong?

Are ML models good enough?

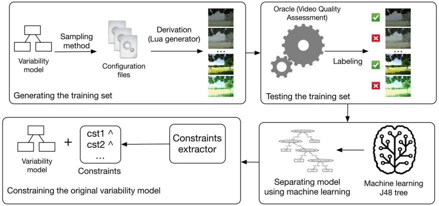


Results: more than 80% correctly classified

Temple *et al.*, Towards quality assurance of software product lines with adversarial configurations, SPLC 2019

Temple *et al.*, Empirical assessment of generating adversarial configurations for software product lines, EMSE 2021

Are ML models good enough?



Results: more than 80% correctly classified

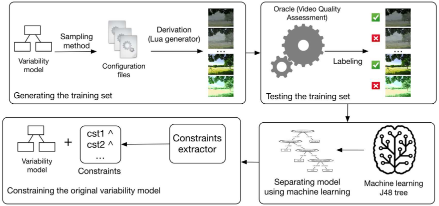
Can we do better?

- Generate more configs that are 'difficult"

Temple *et al.*, Towards quality assurance of software product lines with adversarial configurations, SPLC 2019

Temple *et al.*, Empirical assessment of generating adversarial configurations for software product lines, EMSE 2021

Are ML models good enough?



Results: more than 80% correctly classified

Can we do better?

- Generate more configs that are ‘difficult”
- Use Adversarial Machine Learning

Temple *et al.*, Towards quality assurance of software product lines with adversarial configurations, SPLC 2019

Temple *et al.*, Empirical assessment of generating adversarial configurations for software product lines, EMSE 2021

Are ML models good enough?

About advML

- Went off around 2015
- Better understand ML models and their weaknesses
- 2 main directions: GANs, Evasion/Poisoning

Biggio and Roli, Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning, Pattern Recognition 2018

Are ML models good enough?

About advML

- Went off around 2015
- Better understand ML models and their weaknesses
- 2 main directions: GANs, Evasion/Poisoning

Are ML models good enough?

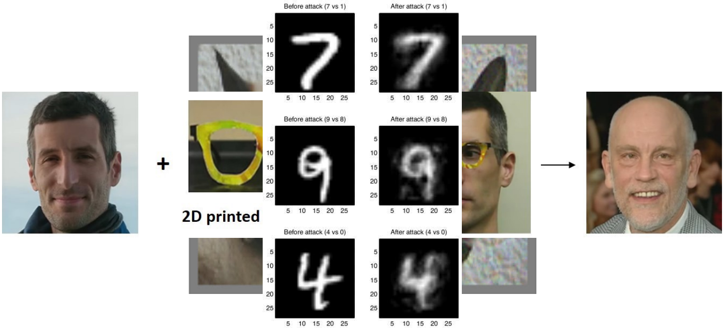
About advML

- Went off around 2015
- Better understand ML models and their weaknesses
- 2 main directions: GANs, Evasion/Poisoning

Are ML models good enough?

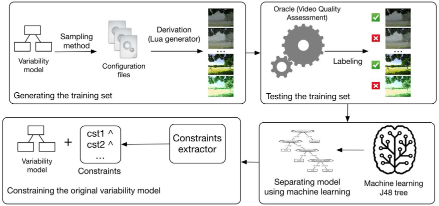
About advML

- Went off around 2015
- Better understand ML models and their weaknesses
- 2 main directions: GANs, Evasion/Poisoning



Biggio and Roli, Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning, Pattern Recognition 2018

Are ML models good enough?

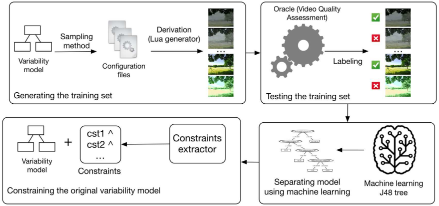


Results: more than 80% correctly classified

Use Adversarial Machine Learning (to get new configurations)

Intriguing properties https://arxiv.org/pdf/1312.6199.pdf

Are ML models good enough?



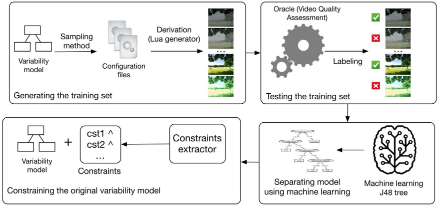
Results: more than 80% correctly classified

Use Adversarial Machine Learning (to get new configurations)

- NDSS'13: Intriguing properties of neural networks

Intriguing properties https://arxiv.org/pdf/1312.6199.pdf

Are ML models good enough?



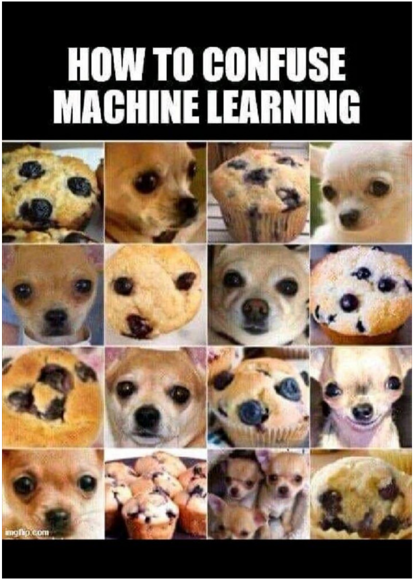
Results: more than 80% correctly classified

Use Adversarial Machine Learning (to get new configurations)

- NDSS'13: Intriguing properties of neural networks
- Present the notion of natural adversarial examples
- Data that are difficult to classify (*i.e.*, lay in between two classes)

Intriguing properties https://arxiv.org/pdf/1312.6199.pdf

Are ML models good enough?

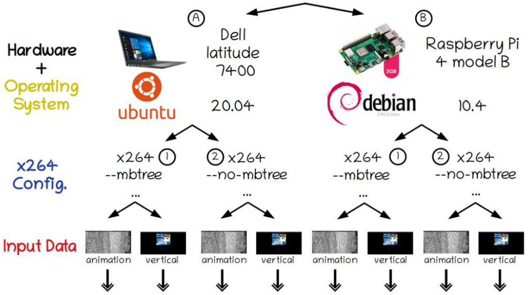


Source: `https://www.kaggle.com/datasets/samuelcortinhas/muffin-vs-chihuahua-image-classification/data`

Are we done?

	no-mbtree	qbur	nr	exec time (s)
	T	0.33	400	24
	F	0.90	200	7
	T	0.90	400	

Deep Software Variability



Duration (s)	6	22	6	25	73	351	72	359
Size (MB)	33	28	21	34	33	28	21	34

Martin *et al.*, Transfer learning across variants and versions: The case of linux kernel size, TSE 2021

Lesoil, Deep Software Variability, PhD thesis, `https://inria.hal.science/tel-04055573/`, 2023

Wrap-up Performance Prediction

- Configuration options quickly increase the complexity of a system
- Building/Executing all the different variants is not possible
- An example: the Linux Kernel and its 15,000 options

Wrap-up Performance Prediction

- Configuration options quickly increase the complexity of a system
- Building/Executing all the different variants is not possible
- An example: the Linux Kernel and its 15,000 options
- How can we try to "guess" the performance of a variant without measuring it?

Wrap-up Performance Prediction

- Configuration options quickly increase the complexity of a system
- Building/Executing all the different variants is not possible
- An example: the Linux Kernel and its 15,000 options
- How can we try to “guess” the performance of a variant without measuring it?
- Understand the interplay between configuration options

Wrap-up Performance Prediction

- Configuration options quickly increase the complexity of a system
- Building/Executing all the different variants is not possible
- An example: the Linux Kernel and its 15,000 options
- How can we try to “guess” the performance of a variant without measuring it?
- Understand the interplay between configuration options
  - Prediction models (linear and need to state all the interactions)
  - ML models (learn by themselves and co-occurence)

Wrap-up Performance Prediction

- ML models need a representative sample of configurations
- Configurations in the sample need to be measured $\rightarrow$ high entrance cost

Wrap-up Performance Prediction

- ML models need a representative sample of configurations
- Configurations in the sample need to be measured $\rightarrow$ high entrance cost
- The more the model is used, the more it pays for itself

Wrap-up Performance Prediction

- ML models need a representative sample of configurations
- Configurations in the sample need to be measured $\rightarrow$ high entrance cost
- The more the model is used, the more it pays for itself
- Be careful about the range of the values of options (ML like them homogeneous)

Wrap-up Performance Prediction

- ML models need a representative sample of configurations
- Configurations in the sample need to be measured $\rightarrow$ high entrance cost
- The more the model is used, the more it pays for itself
- Be careful about the range of the values of options (ML like them homogeneous)
- We can improve the model via adversarial retraining

ML and SE

(VaryMinions)

ML and SE

(VaryMinions)

Sophie Fortz's PhD

ML and SE

(VaryMinions)

Sophie Fortz's PhD
(thanks for the slides)

VaryMinions: Context

- A configurable process is being used
- Different evolutions came out
- You are managing several instances of the same process

VaryMinions: Context

- A configurable process is being used
- Different evolutions came out
- You are managing several instances of the same process
- Something wrong happens... (*e.g.*, a bug)
- You can **only monitor** the processes

VaryMinions: Context

- A configurable process is being used
- Different evolutions came out
- You are managing several instances of the same process
- Something wrong happens... (*e.g.*, a bug)
- You can **only monitor** the processes

What can you do to know which instances are impacted?

VaryMinions: Context

- A configurable process is being used
- Different evolutions came out
- You are managing several instances of the same process
- Something wrong happens... (*e.g.*, a bug)
- You can **only monitor** the processes

What can you do to know which instances are impacted?
What can you do to know where the bug comes from? (which feature or part of the process)

VaryMinions: Research Problem

Multiple instances of the same process

- Logs are produced so that the sytem can be monitored
- Multiple logs per instance of the process
- Multiple instances are running in parallel

VaryMinions: Research Problem

Multiple instances of the same process

- Logs are produced so that the sytem can be monitored
- Multiple logs per instance of the process
- Multiple instances are running in parallel

How can you **map** the logs to their instances?

VaryMinions: Research Problem

Multiple instances of the same process

- Logs are produced so that the sytem can be monitored
- Multiple logs per instance of the process
- Multiple instances are running in parallel

How can you **map** the logs to their instances?

Combinatorial explosions

- Lot of information per log
- Many variants
- Many logs per variants

VaryMinions: Research Problem

Need for automation

- Can ask domain expert which logs correspond to which variant(s)

VaryMinions: Research Problem

Need for automation

- Can ask domain expert which logs correspond to which variant(s)
  - Time consuming
  - Fatigue
  - Potential mistakes?

VaryMinions: Research Problem

Need for automation

- Can ask domain expert which logs correspond to which variant(s)
  - Time consuming
  - Fatigue
  - Potential mistakes?
- Can we use ML to help?

VaryMinions: Research Problem

Need for automation

- Can ask domain expert which logs correspond to which variant(s)
  - Time consuming
  - Fatigue
  - Potential mistakes?
- Can we use ML to help?

Representation of logs?

- Logged information follow the steps of the process
- Can be seen as an ordered sequence of events

VaryMinions: Research Problem

Need for automation

- Can ask domain expert which logs correspond to which variant(s)
  - Time consuming
  - Fatigue
  - Potential mistakes?
- Can we use ML to help?

Representation of logs?

- Logged information follow the steps of the process
- Can be seen as an ordered sequence of events
- A sentence is an ordered sequence of words

Can we make as if logs are text?

VaryMinions: ML to the rescue

Log annotation

- Let experts annotate a few logs
- Annotation = mapping from logs to **variant(S)**

VaryMinions: ML to the rescue

Log annotation

- Let experts annotate a few logs
- Annotation = mapping from logs to **variant(S)**
- Hope that ML models can do the rest...

VaryMinions: ML to the rescue

Log annotation

- Let experts annotate a few logs
- Annotation = mapping from logs to **variant(S)**
- Hope that ML models can do the rest...

Which ML model?

- NLP, Markov Chain, etc.

VaryMinions: ML to the rescue

Log annotation

- Let experts annotate a few logs
- Annotation = mapping from logs to **variant(S)**
- Hope that ML models can do the rest...

Which ML model?

- NLP, Markov Chain, etc.
- Models using text make some assumptions (entity recognition, etc.)

VaryMinions: ML to the rescue

Log annotation

- Let experts annotate a few logs
- Annotation = mapping from logs to **variant(S)**
- Hope that ML models can do the rest...

Which ML model?

- NLP, Markov Chain, etc.
- Models using text make some assumptions (entity recognition, etc.)
- What about DL models?

VaryMinions: ML to the rescue

Log annotation

- Let experts annotate a few logs
- Annotation = mapping from logs to **variant(S)**
- Hope that ML models can do the rest...

Which ML model?

- NLP, Markov Chain, etc.
- Models using text make some assumptions (entity recognition, etc.)
- What about DL models?

DL models for sequence

- Recurrent Neural Networks (RNNs)

VaryMinions: ML to the rescue

Log annotation

- Let experts annotate a few logs
- Annotation = mapping from logs to **variant(S)**
- Hope that ML models can do the rest...

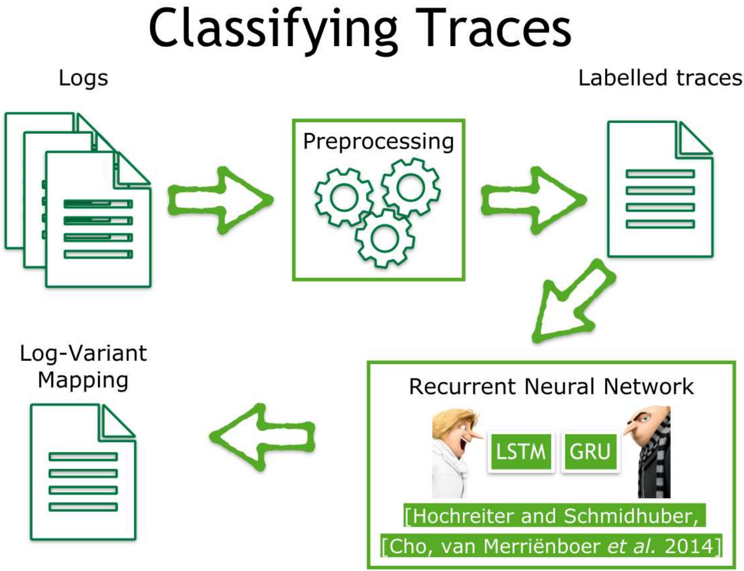
Which ML model?

- NLP, Markov Chain, etc.
- Models using text make some assumptions (entity recognition, etc.)
- What about DL models?

DL models for sequence

- Recurrent Neural Networks (RNNs)
  - GRU
  - LSTM

VaryMinions: Overview



VaryMinions: Does it work? Any problems?

Experimental setup

- 3 datasets
- Set different hyperparameters (including loss and activation functions)

VaryMinions: Does it work? Any problems?

Experimental setup

- 3 datasets
- Set different hyperparameters (including loss and activation functions)

Results

- Give at least 80% accuracy
- No prevalence of GRU nor LSTM

VaryMinions: Does it work? Any problems?

Experimental setup

- 3 datasets
- Set different hyperparameters (including loss and activation functions)

Results

- Give at least 80% accuracy
- No prevalence of GRU nor LSTM

Problems?

- Are the losses right? $\rightarrow$ need to define custom losses?

VaryMinions: Does it work? Any problems?

Experimental setup

- 3 datasets
- Set different hyperparameters (including loss and activation functions)

Results

- Give at least 80% accuracy
- No prevalence of GRU nor LSTM

Problems?

- Are the losses right? $\rightarrow$ need to define custom losses?
- Are our architectures right? $\rightarrow$ how much layers is too many layers?

VaryMinions: Does it work? Any problems?

Experimental setup

- 3 datasets
- Set different hyperparameters (including loss and activation functions)

Results

- Give at least 80% accuracy
- No prevalence of GRU nor LSTM

Problems?

- Are the losses right? $\rightarrow$ need to define custom losses?
- Are our architectures right? $\rightarrow$ how much layers is too many layers?
- ...

Wrap-up VaryMinions

- Variants of the process are deployed, ran, and monitored
- A problem occur
- Is it a problem for a single variant or multiple ones?

Wrap-up VaryMinions

- Variants of the process are deployed, ran, and monitored
- A problem occur
- Is it a problem for a single variant or multiple ones?
- Tons of logs can be available (¿ 10... =))
- Cannot be manually managed

Wrap-up VaryMinions

- Variants of the process are deployed, ran, and monitored
- A problem occur
- Is it a problem for a single variant or multiple ones?
- Tons of logs can be available (¿ 10... =))
- Cannot be manually managed
- Can ML help after experts have labelled a few? (help in mapping a log to potential variants that could have generated it)

Wrap-up VaryMinions

- Logs can be seen as text (the sequence is not completely arbitrary)
- RNNs are able to deal with sequences (GRU and LSTM)

Wrap-up VaryMinions

- Logs can be seen as text (the sequence is not completely arbitrary)
- RNNs are able to deal with sequences (GRU and LSTM)
- Evaluation show that RNNs can help in the mapping (at least 80% accuracy)

Wrap-up VaryMinions

- Logs can be seen as text (the sequence is not completely arbitrary)
- RNNs are able to deal with sequences (GRU and LSTM)
- Evaluation show that RNNs can help in the mapping (at least 80% accuracy)
- Cannot favor GRU nor LSTM

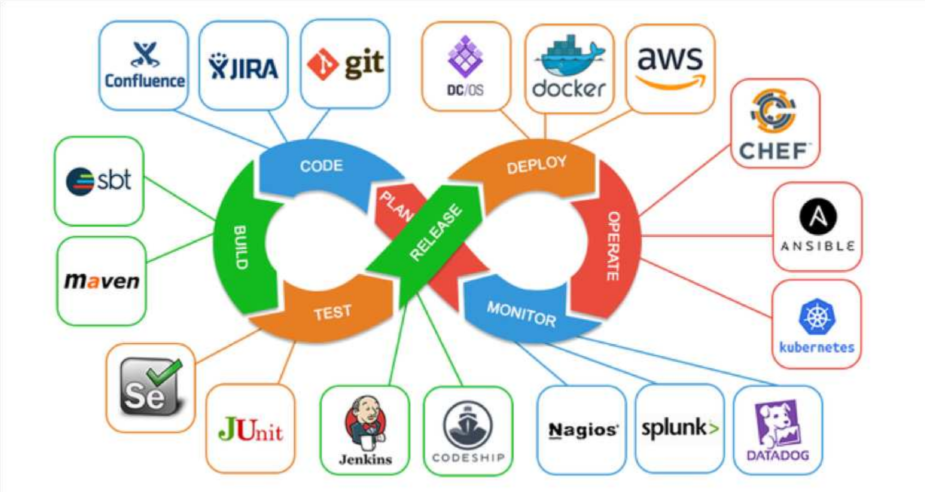
Wrap-up VaryMinions

- Logs can be seen as text (the sequence is not completely arbitrary)
- RNNs are able to deal with sequences (GRU and LSTM)
- Evaluation show that RNNs can help in the mapping (at least 80% accuracy)
- Cannot favor GRU nor LSTM
- potential evolutions: custom losses, bigger models?, much more

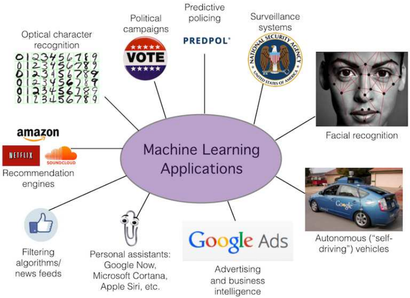
ML and SE

(When ML meets DevOps)

Overview of DevOps



ML in DevOps



Source:https://www.redshiftzero.com/img/ml_applications.png

- Modern system are complex (multiple components, configurable, etc.)
- Need to adapt to the environment
- ML decisions impact the system (adaptations, outputs, etc.)

ML in DevOps

ML Assumptions

- Training(, validation,) and test sets $\rightarrow$ same distribution
- Distributions do not evolve

ML in DevOps

ML Assumptions

- Training(, validation,) and test sets $\rightarrow$ same distribution
- Distributions do not evolve

Do not hold

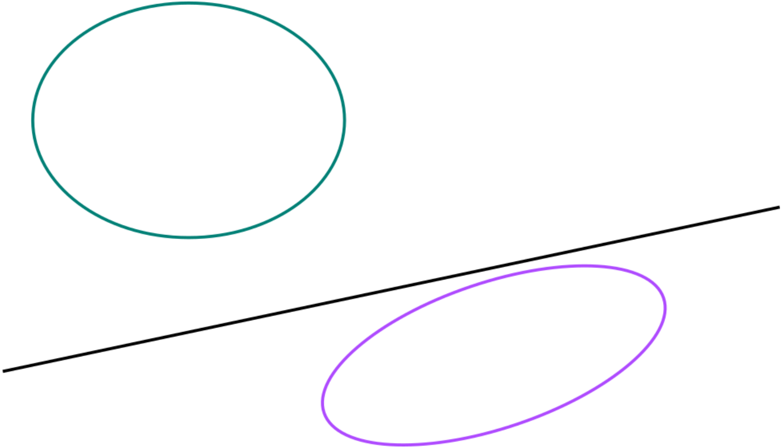
ML in DevOps

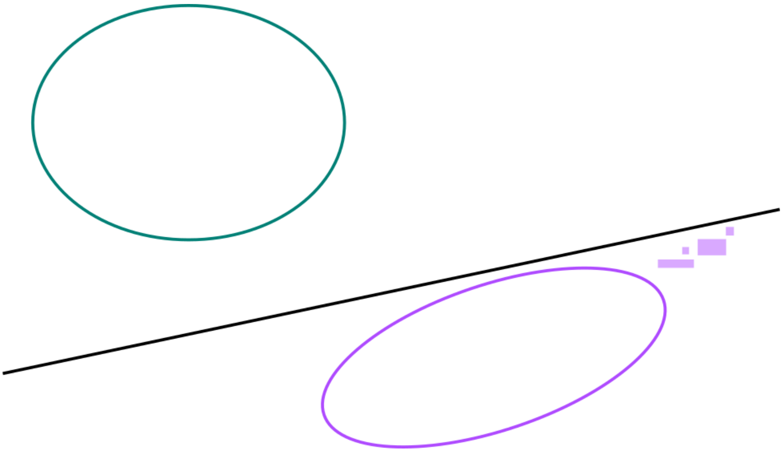
ML Assumptions

- Training(, validation,) and test sets $\rightarrow$ same distribution
- Distributions do not evolve

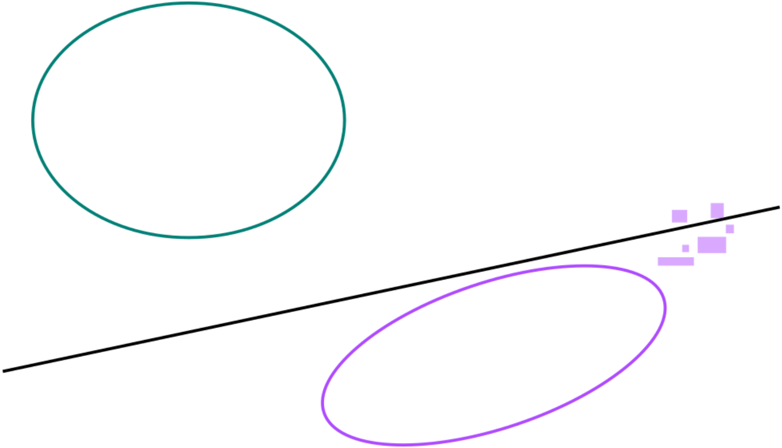
Do not hold

Machine Learning in evolving context

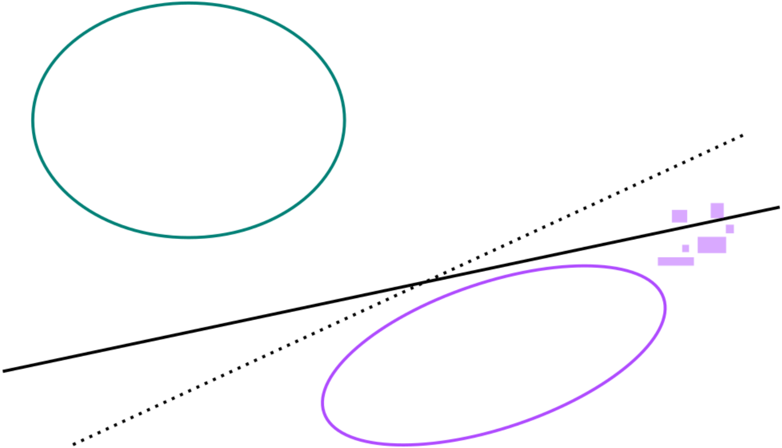




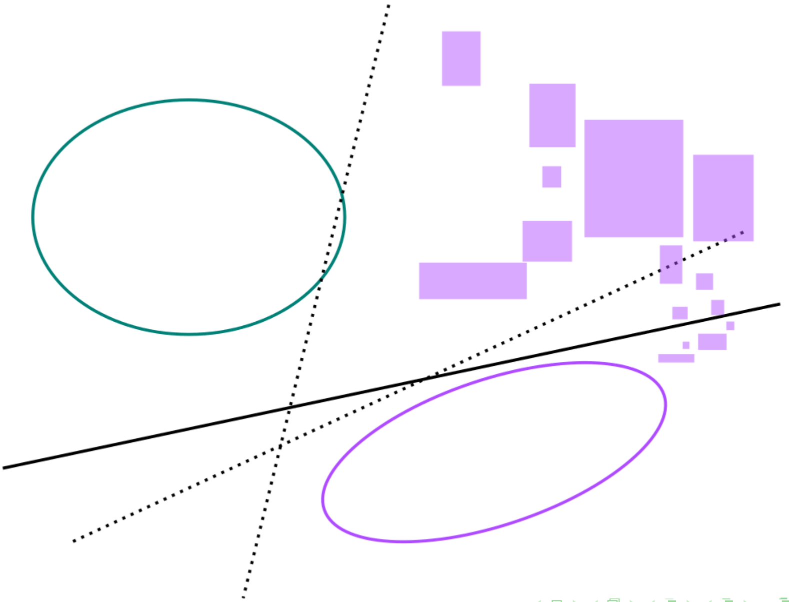
Machine Learning in evolving context



Machine Learning in evolving context



Machine Learning in evolving context



How to prevent from shifting?

retraining from scratch

- Original training set + some new data
- Manual check of new data to add
- Relabel if needed

How to prevent from shifting?

retraining from scratch

- Original training set + some new data
- Manual check of new data to add
- Relabel if needed

Problems

- Check and relabeling can be tedious
- Model can be hard to optimize

How to prevent from shifting?

retraining from scratch

- Original training set + some new data
- Manual check of new data to add
- Relabel if needed

Problems

- Check and relabeling can be tedious
- Model can be hard to optimize

→ Compliant with DevOps?
How often should we retrain?

How to prevent from shifting?

Update the model

- Add new data to update partially the model
- Convergence should be fast
- Check and relabel if needed

How to prevent from shifting?

Update the model

- Add new data to update partially the model
- Convergence should be fast
- Check and relabel if needed

Problems

- Check and relabeling can be tedious
- Lot of data might be necessary to have an impact
- Consecutive changes over time $\Rightarrow$ **model shifting**

How to prevent from shifting?

Adding data

- Relabeling might be enough
- Mixing new data with older data
- data sanitization (e.g., out of distribution)

How to prevent from shifting?

Adding data

- Relabeling might be enough
- Mixing new data with older data
- data sanitization (e.g., out of distribution)

Mixing update and retraining

- Updating for fast adaptations
- Regularly retrain from scratch

How to prevent from shifting?

Adding data

- Relabeling might be enough
- Mixing new data with older data
- data sanitization (e.g., out of distribution)

Mixing update and retraining

- Updating for fast adaptations
- Regularly retrain from scratch

Federated learning?

- Keep a "master" version
- Deploy customized versions for customers
- customized versions are upgraded & "master" version is retrained

Extremely hot topic in research

Wrap-up DevOps and ML

ML models are part of systems

- With DevOps $\rightarrow$ Models need to be updated
- Beware of Model Shifting

Wrap-up DevOps and ML

ML models are part of systems

- With DevOps $\rightarrow$ Models need to be updated
- Beware of Model Shifting
- Different strategy for update
  - Update training set and retrain from scratch

Wrap-up DevOps and ML

ML models are part of systems

- With DevOps $\rightarrow$ Models need to be updated
- Beware of Model Shifting
- Different strategy for update
  - Update training set and retrain from scratch
  - Update the model with data sanitization
  - Update the model with adversarial retraining

Wrap-up DevOps and ML

ML models are part of systems

- With DevOps $\rightarrow$ Models need to be updated
- Beware of Model Shifting
- Different strategy for update
  - Update training set and retrain from scratch
  - Update the model with data sanitization
  - Update the model with adversarial retraining
- Problem: How often updating is needed?
- Problem: Consequences on Federated learning?