UPPSALA UNIVERSITET

Marcin Kierczak

EXPLAINABLE(?) PREDICTIONS: INTRO TO BUILDING MODELS WITHIN THE SUPERVISED LEARNING FRAMEWORK

Overview

- A bit of history.
- Machine Learning (ML) in general.
- Supervised learning framework.
  - Feature engineering & selection.
  - Model selection.
  - Model validation.
  - Model interpretation.
  - Uncertainty.
- General remarks.

Artificial Intelligence & co.

Artificial Intelligence (AI)

The effort to automate intellectual tasks normally performed by humans.[1]

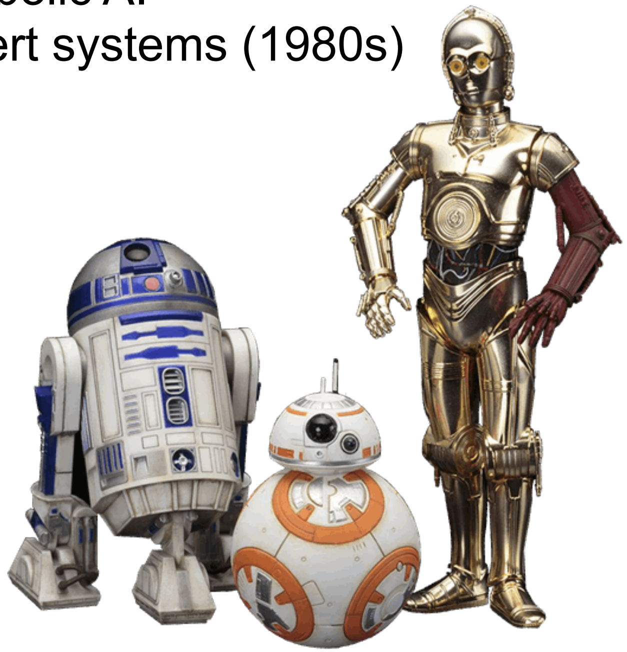
[1] *Chollet & Allaire. Deep learning with R.*

Artificial Intelligence & co.

Artificial Intelligence (AI)

The effort to automate intellectual tasks normally performed by humans.[1]

- 1950s
- symbolic AI
- expert systems (1980s)



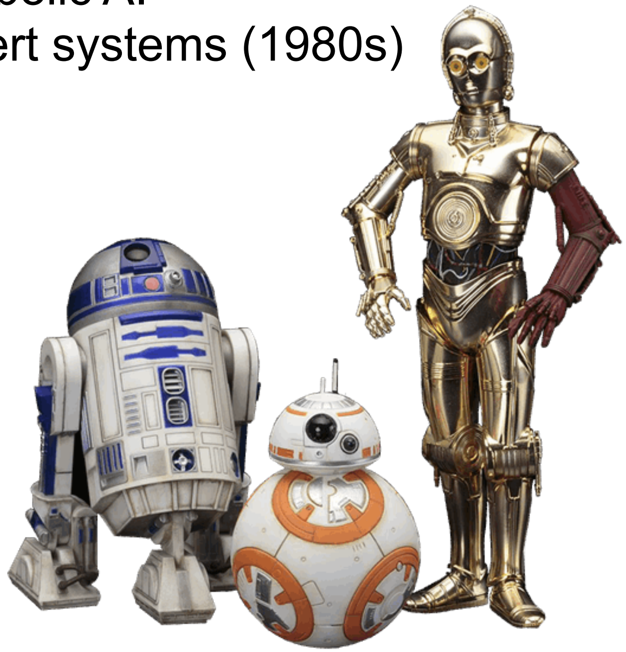
[1] *Chollet & Allaire. Deep learning with R.*

Artificial Intelligence & co.

Artificial Intelligence (AI)

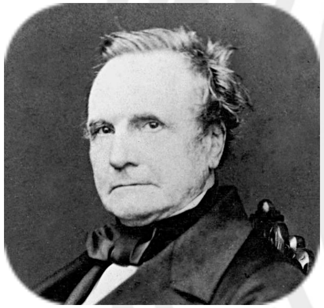
The effort to automate intellectual tasks normally performed by humans.[1]

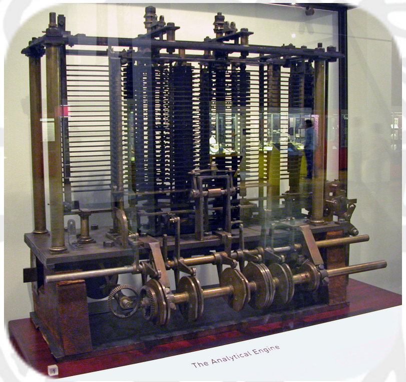
- 1950s
- symbolic AI
- expert systems (1980s)



The Analytical Engine has no pretentions whatever to originate anything. It can do whatever we know how to order it to perform…

Lady Ada Lovelace





[1] *Chollet & Allaire. Deep learning with R.*

images: commons.wikimedia.com

Machine Learning

Artificial Intelligence (AI)

The effort to automate intellectual tasks normally performed by humans.[1]

Allan Turing

- *Computing Machinery and Intelligence*
- Turing test

Data → PROGRAM → Answers
Rules → PROGRAM

[1] *Chollet & Allaire. Deep learning with R.*

images: commons.wikimedia.com

UPPSALA UNIVERSITET

Machine Learning

Artificial Intelligence (AI)

The effort to automate intellectual tasks normally performed by humans.[1]

Allan Turing

- *Computing Machinery and Intelligence*
- Turing test

Data → PROGRAM → **Rules**

Answers →

[1] *Chollet & Allaire. Deep learning with R.*

images: commons.wikimedia.com

Machine Learning

Artificial Intelligence (AI)

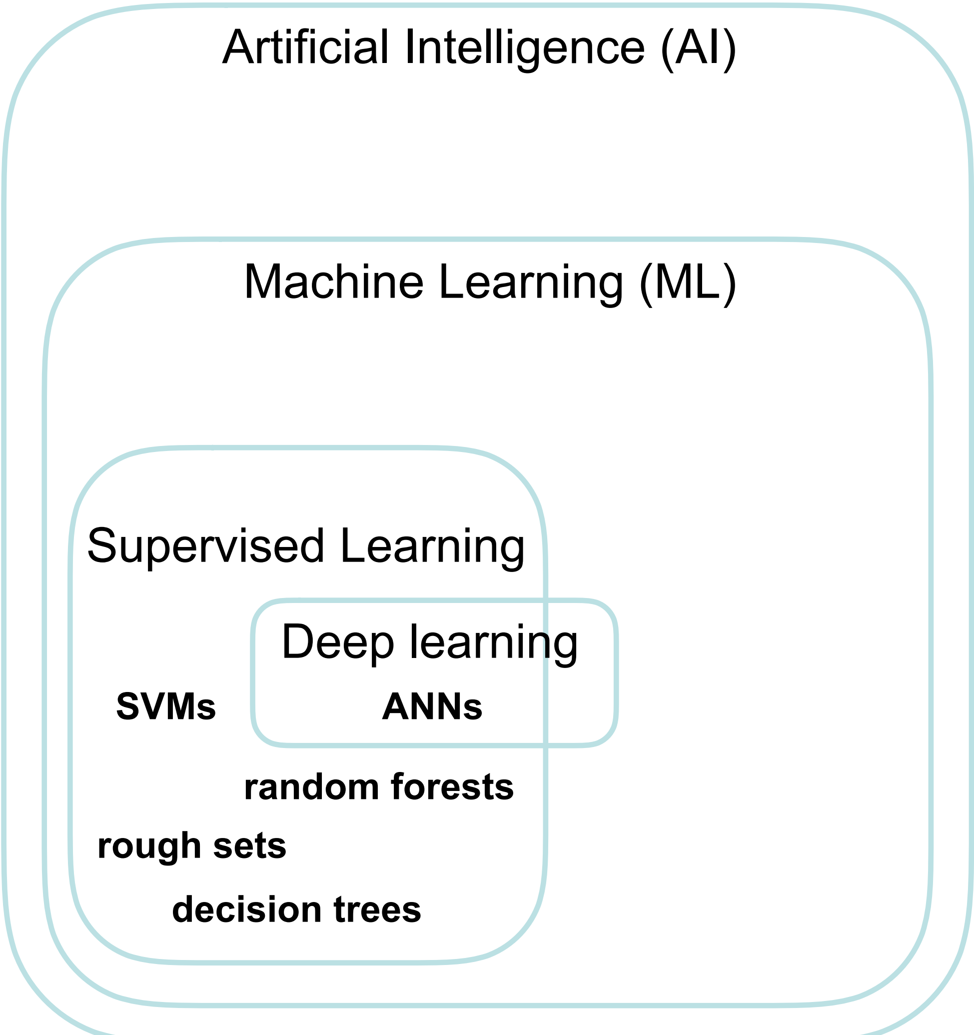
The effort to automate intellectual tasks normally performed by humans.

Machine Learning (ML)

Machine Learning is a field of study that gives computers the ability to learn without being explicitly programmed.

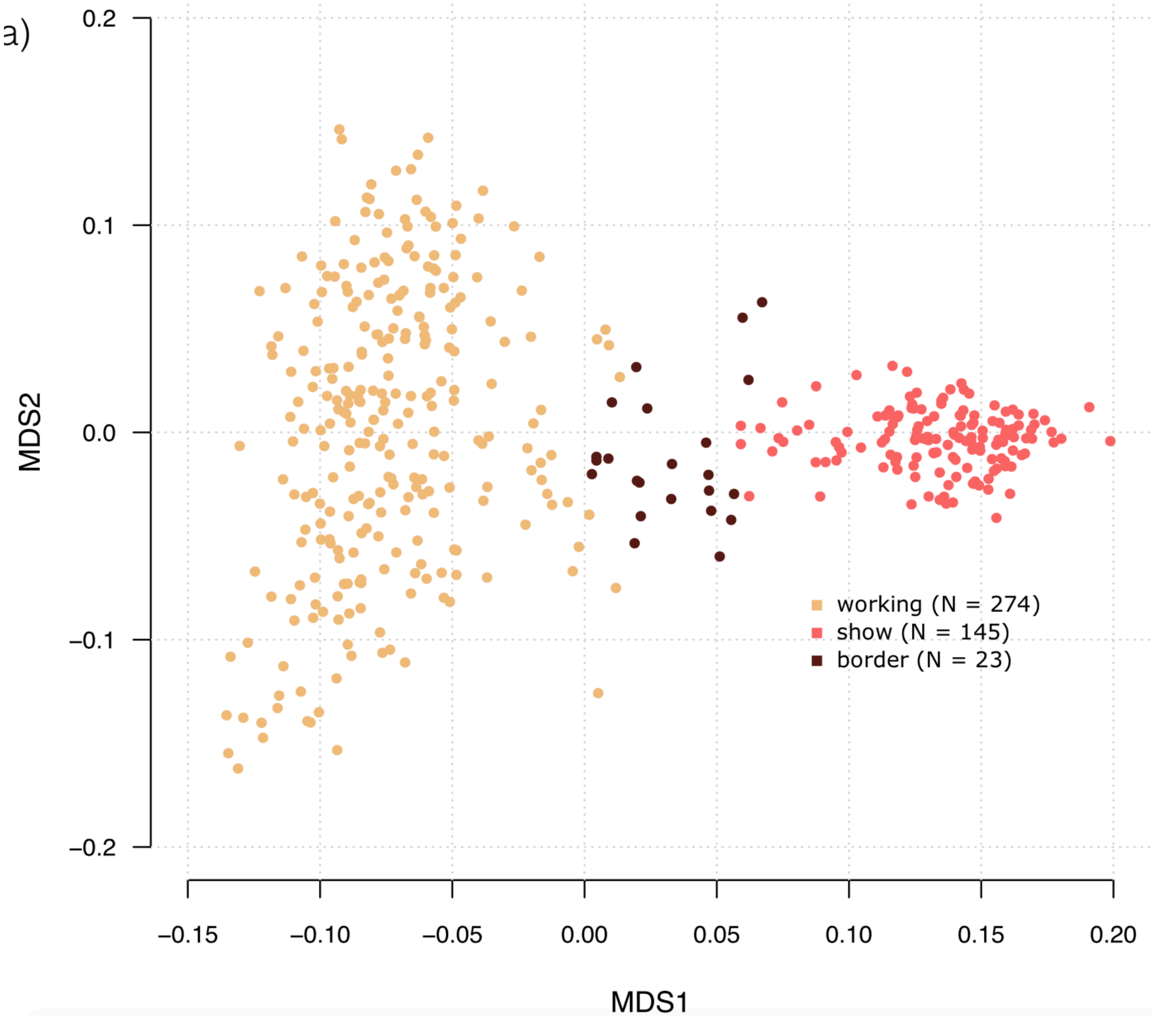
Arthur Samuel, 1959

Overview



SVM - support-vector machines
ANN - artificial neural network

Unsupervised learning



Kierczak et al. 2019 in press.

Supervised learning

A decision system

	Attr. 1	Attr. 2	…	Attr. M	d
Ex. 1	1	big	…	11.54	**case**
Ex. 2	4	small	…	-5.48	**control**
…	…	…	…	…	**…**
Ex. N	5	medium	…	7.26	**control**

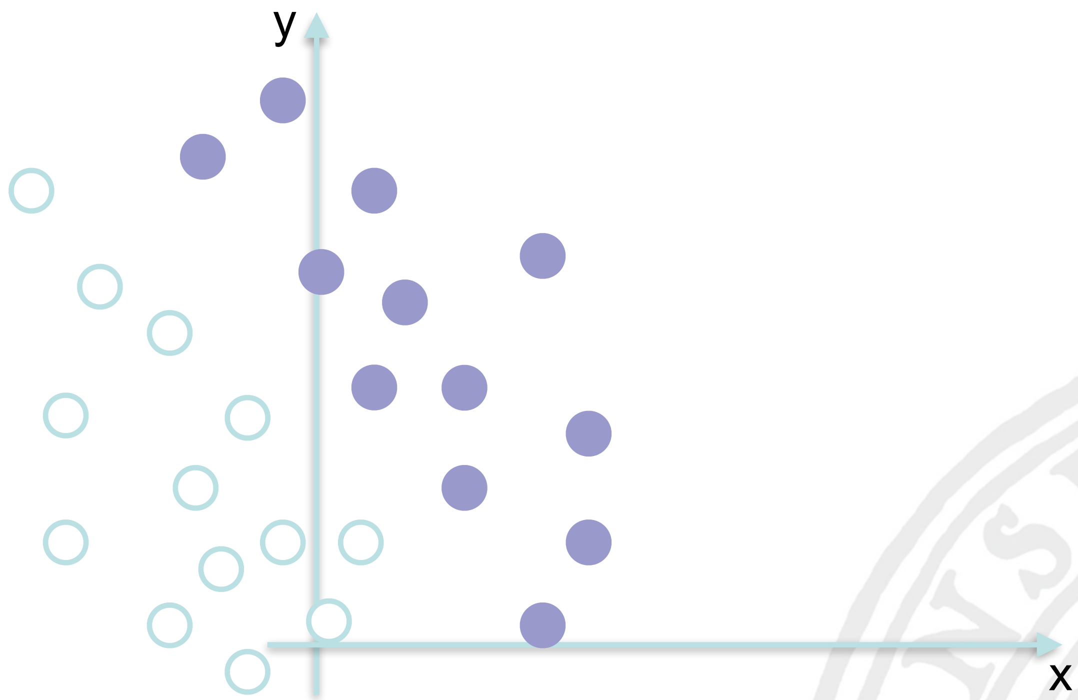
$$\mathbf{A}_d = (U, A \cup \{d\})$$

$$d \notin A$$

A – conditional attributes,
d – decision attribute(s)

Geometric interpretation

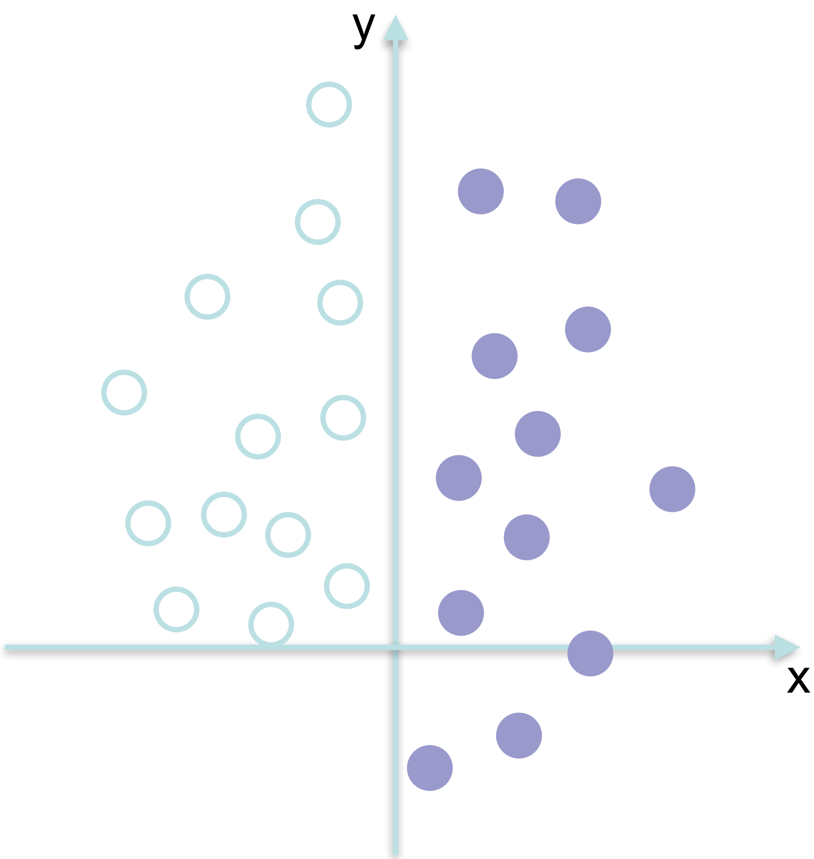
Geometric interpretation



IF x < 0 THEN white ELSE blue

Geometric interpretation

Geometric interpretation

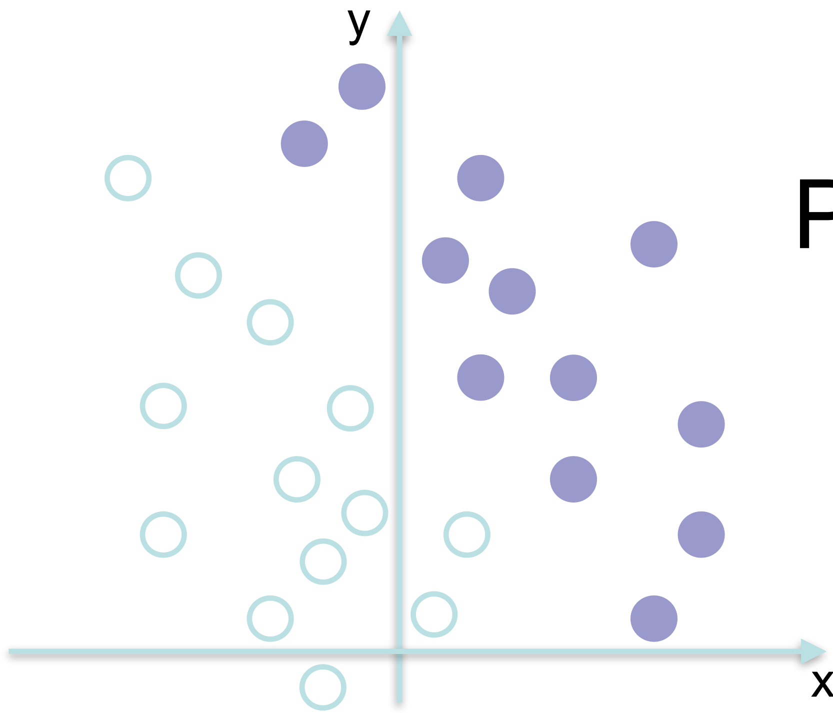


Formal definition of ML

A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with E.

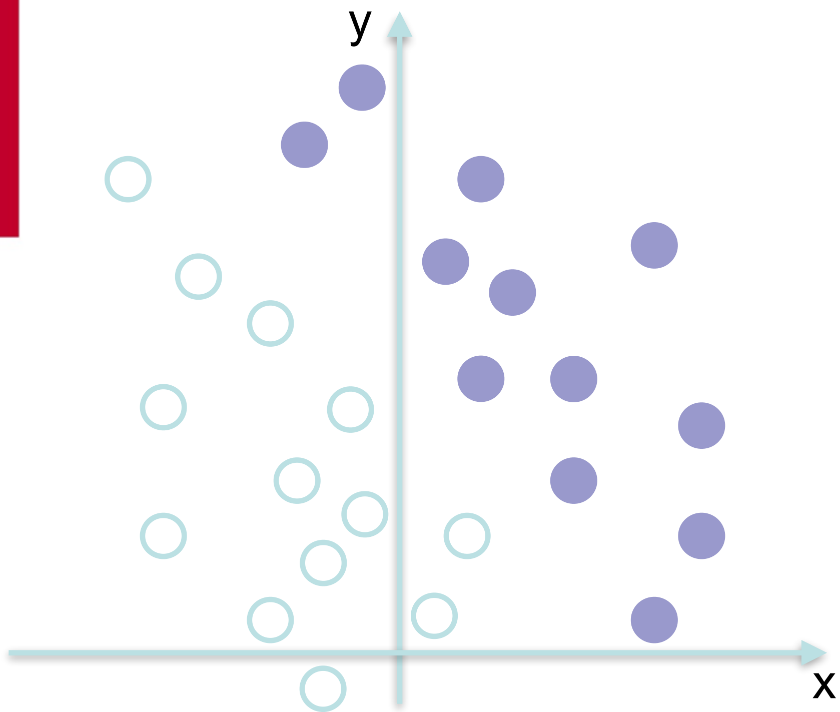
Tom Mitchell, 1997

Performance P



		Actual	
		●	○
Predicted	●	TP	FP
	○	FN	TN

Performance P



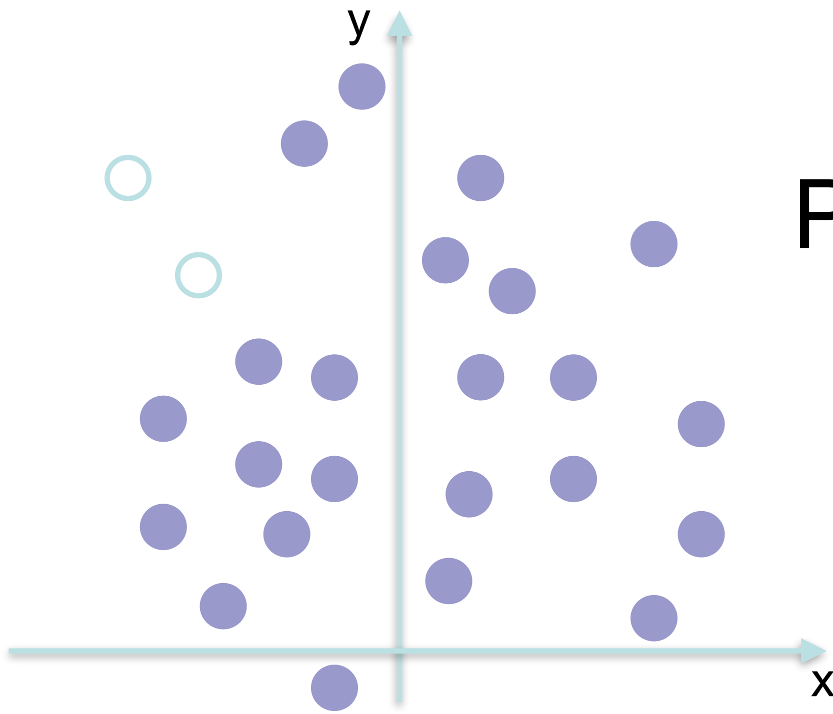
13 white
12 blue

- class bias
- misclassification cost

		Actual	
		●	○
Predicted	●	10	2
	○	2	11

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} = \frac{10 + 11}{10 + 11 + 2 + 2} = \frac{21}{25} = 0.84$$

Performance P



		Actual	
		●	○
Predicted	●	23	2
	○	0	0

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} = \frac{23 + 0}{23 + 0 + 2 + 0} = \frac{23}{25} = 0.92$$

Performance P

		Actual	
		●	○
Predicted	●	10	2
	○	11	2

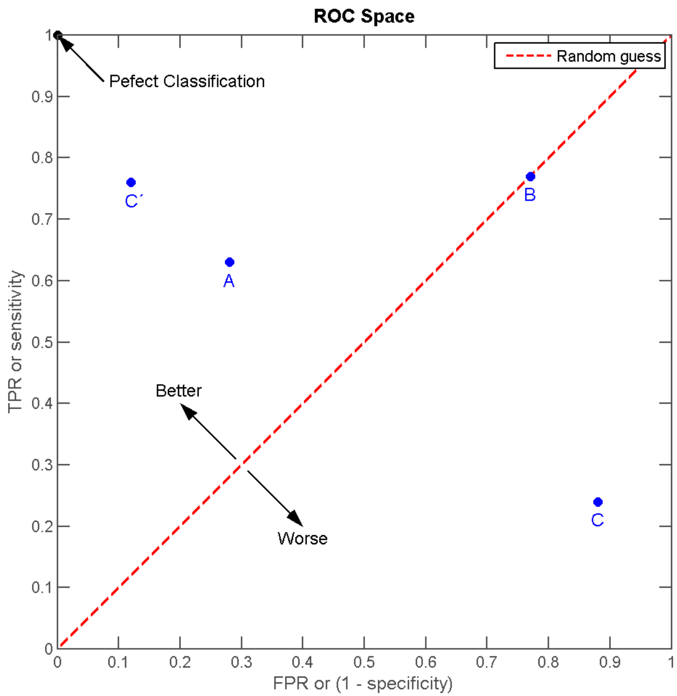
$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

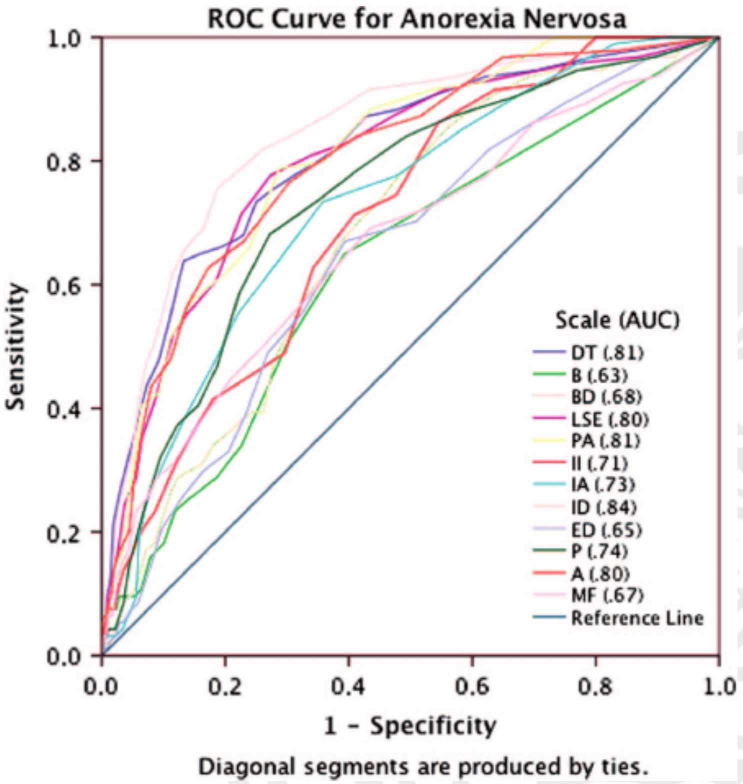
$$Precision = \frac{TP}{TP + FP}$$ fraction of correctly predicted sick to all predicted to be sick

$$Sensitivity = \frac{TP}{TP + FN}$$ fraction of correctly predicted sick to all sick

$$Specificity = \frac{TN}{TN + FP}$$ fraction of correctly predicted healthy to all healthy

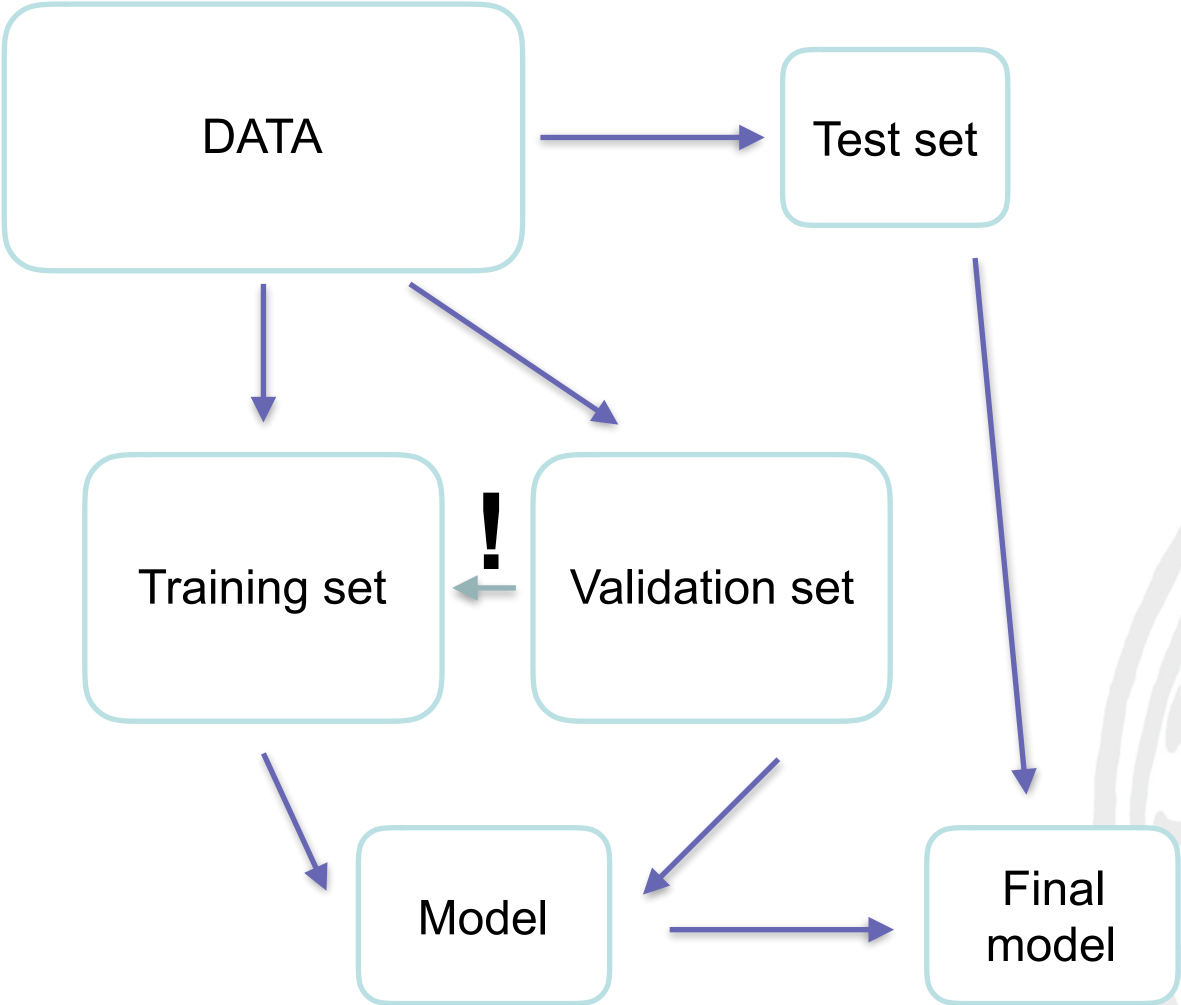
Performance P – ROC & AUC





Nyman-Carlsson, et al. (2014). Eating Disorder Inventory-3: Validation in Swedish patients with eating disorders, psychiatric outpatients and a normal control sample. Nordic Journal of Psychiatry. 68. 1-10.

Measuring performance



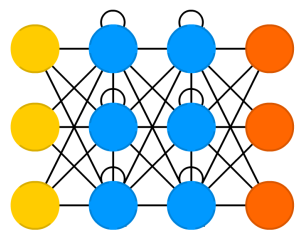
- class distribution
- information leaks
- temporal leaks
- duplication leaks

Performance baseline

IF x < 0 THEN white ELSE blue

Acc = 0.74

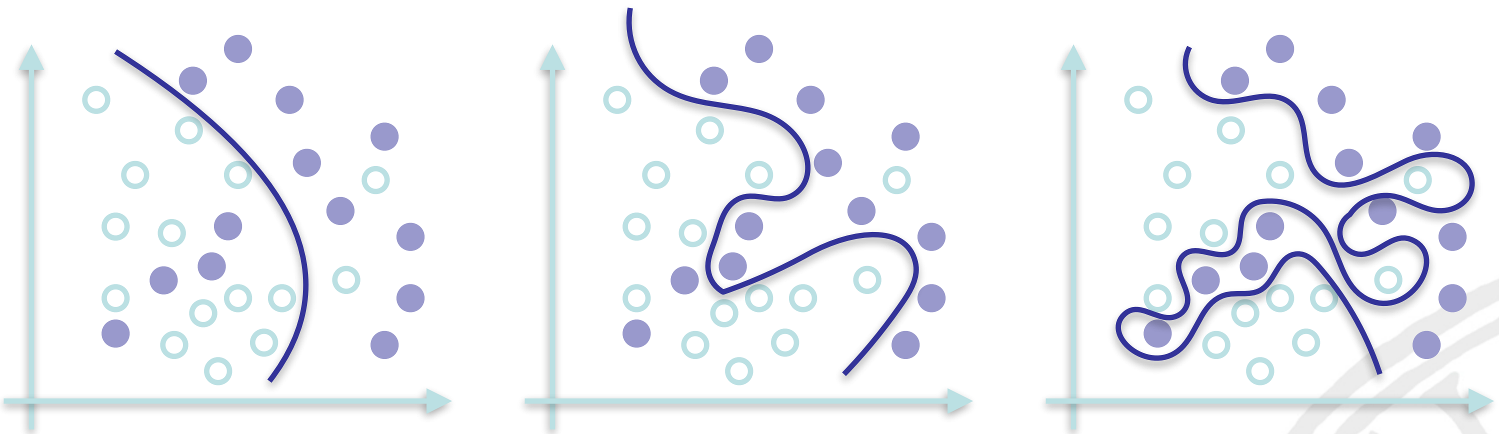
Acc = 0.54



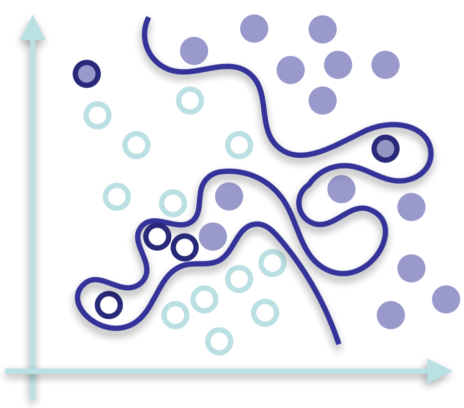
Acc = 0.72

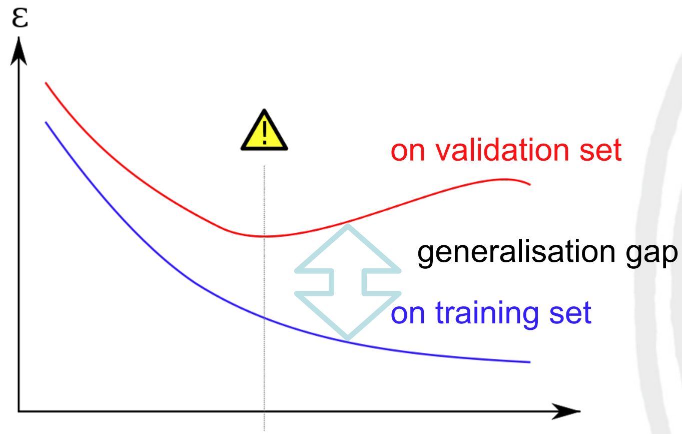
Acc = 0.80

Overfitting

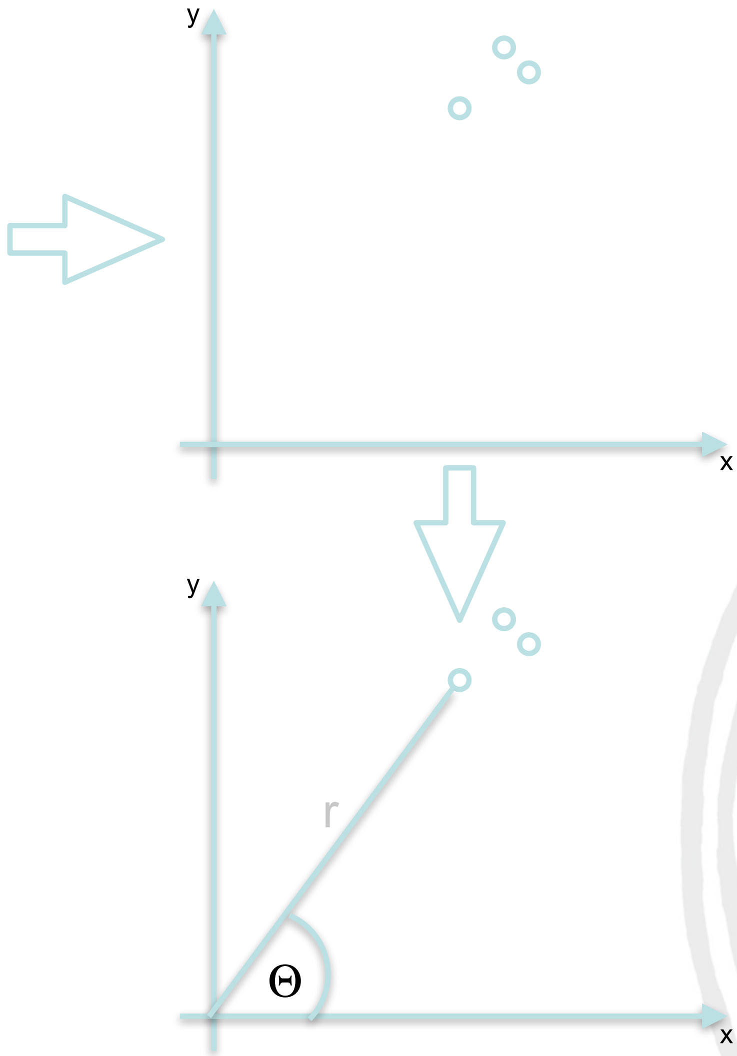


On new data

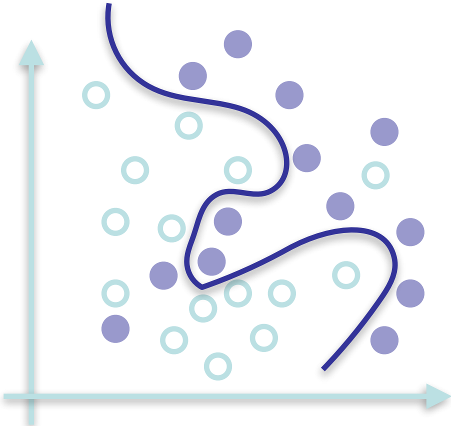


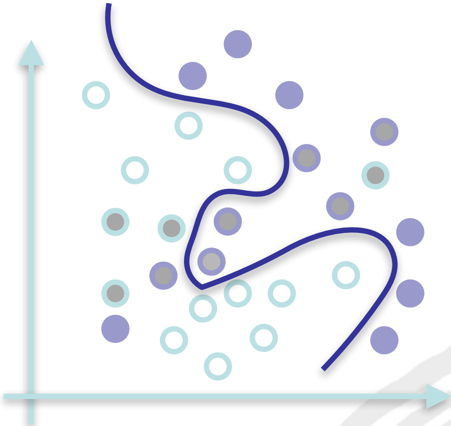


Feature engineering



Interpretability





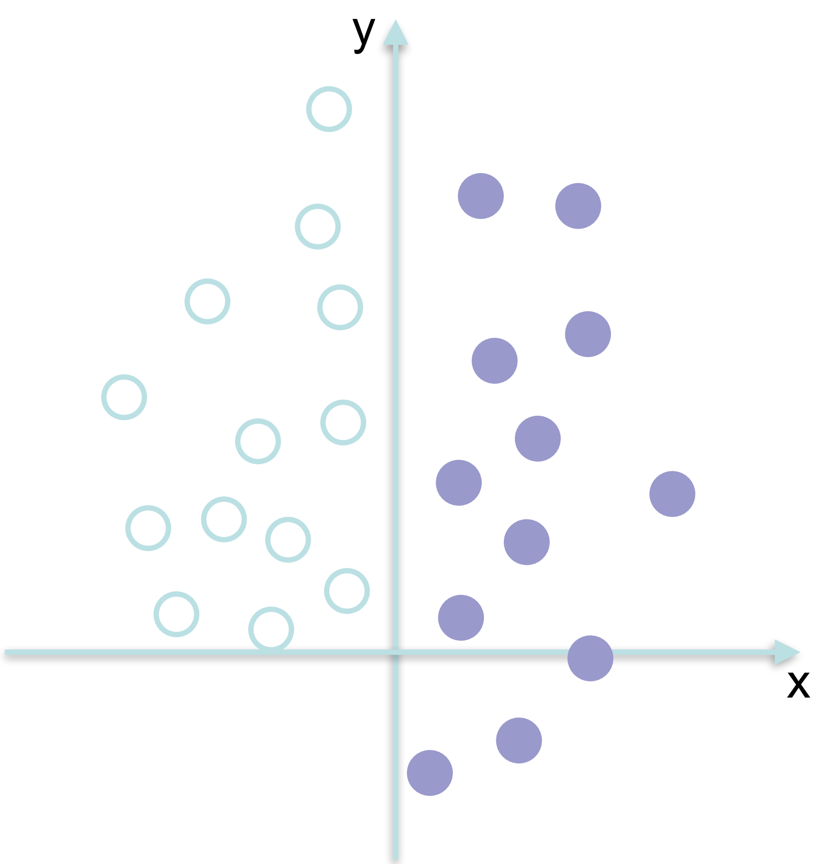
Input image

a couple of bears are
standing in a field

Output caption

Importance map

Feature selection



Feature selection

	a1	a2	d
Ex. 1	1	7.45	**case**
Ex. 2	0	3.24	**control**
…	…	…	**…**
Ex. N-1	1	8.72	**case**
Ex. N	0	11.5	**control**

Acc = 0.98

Ex. 1	1	7.45	**case**
Ex. 2	1	3.24	**control**
…	…	…	**…**
Ex. N-1	0	8.72	**case**
Ex. N	0	11.5	**control**

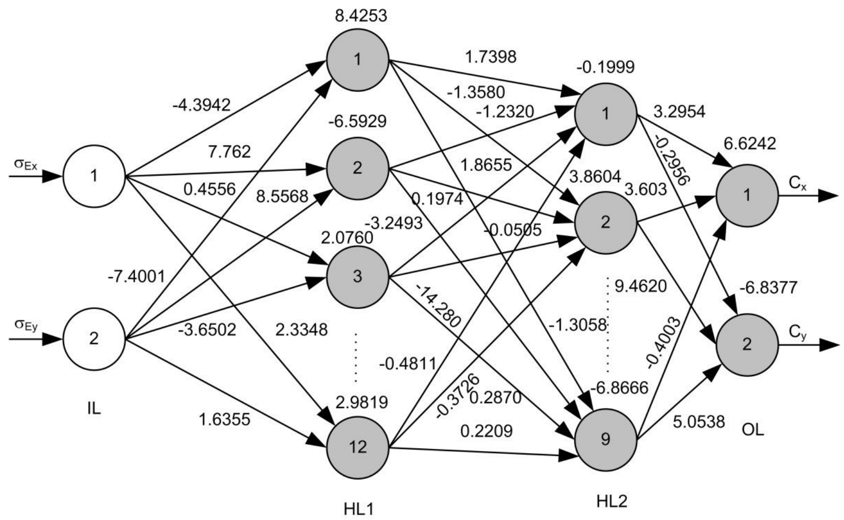
Acc = 0.95

I(a2) > I(a1)

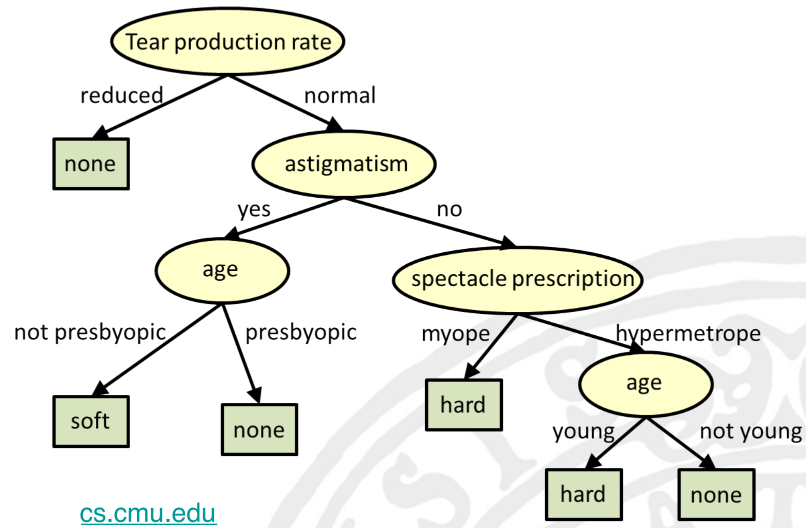
Ex. 1	1	8.72	**case**
Ex. 2	0	11.5	**control**
…	…	…	**…**
Ex. N-1	1	7.45	**case**
Ex. N	0	3.24	**control**

Acc = 0.75

Explainability



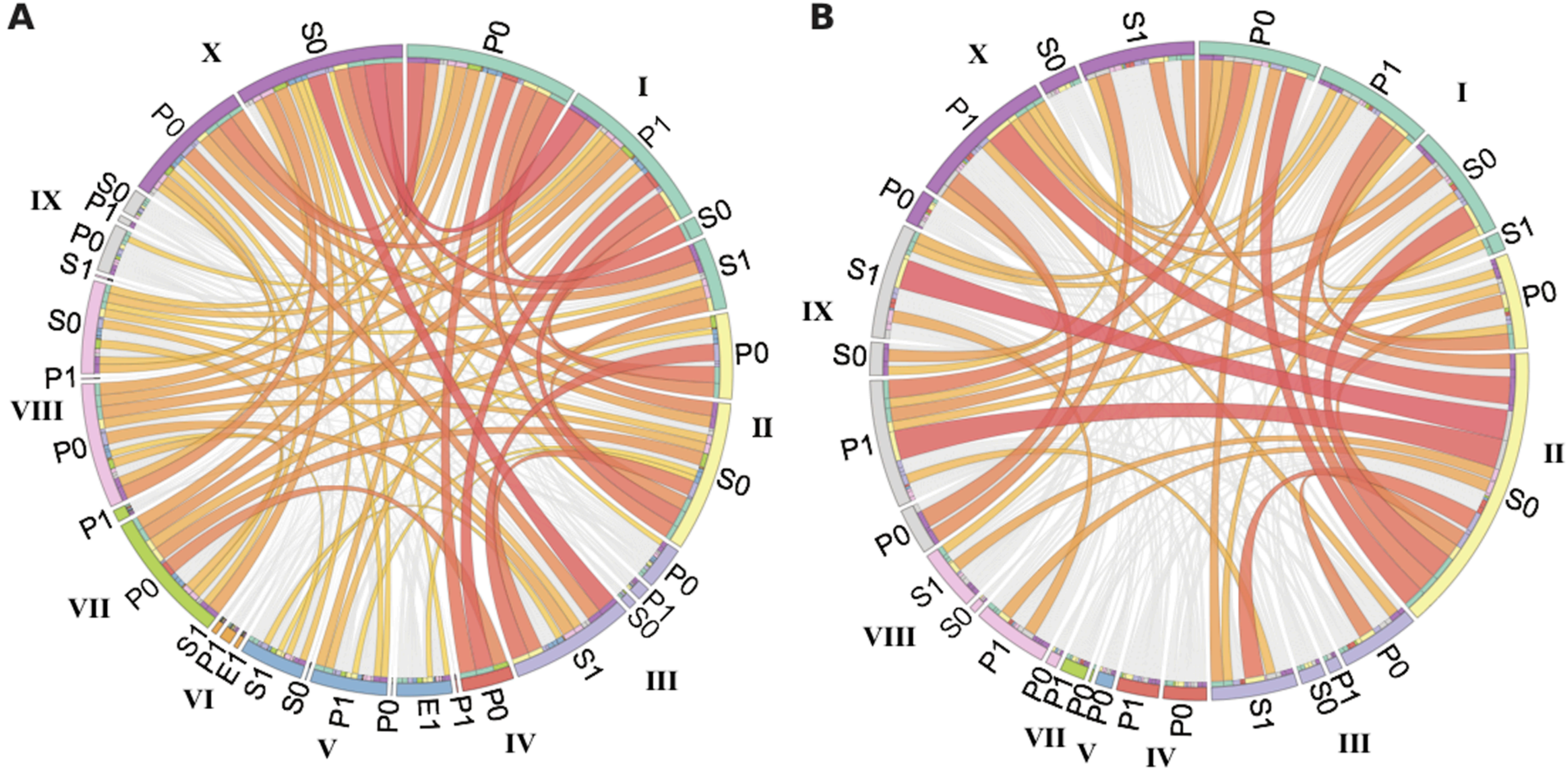
Arockia et al. Journal on Wireless Communications and Networking 2014:160



cs.cmu.edu

IF tear prod. rate = normal & astigmatism = yes & age != presbyopic THEN lenses = soft

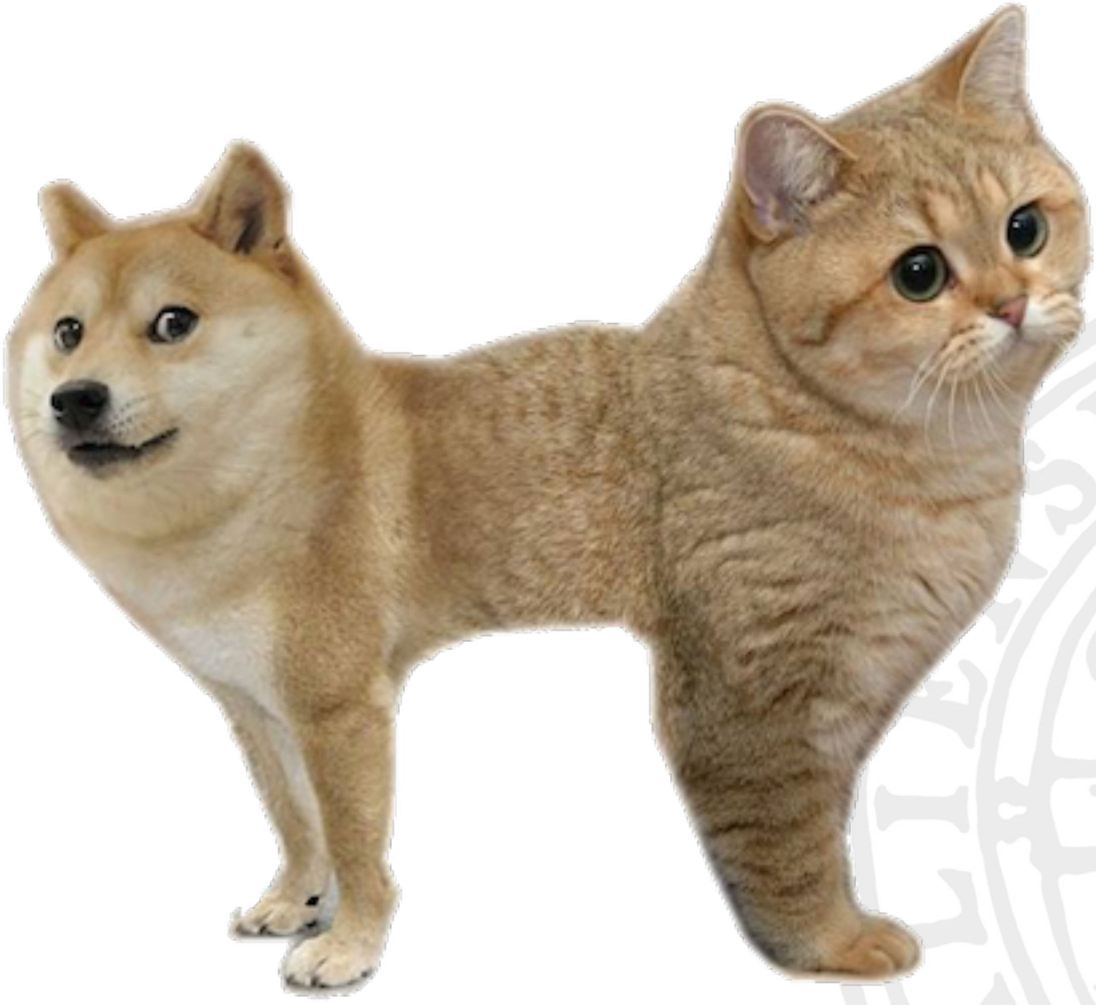
Explainability



“I P0” is interpreted as “H2BK5me1 preceding the exon is absent”

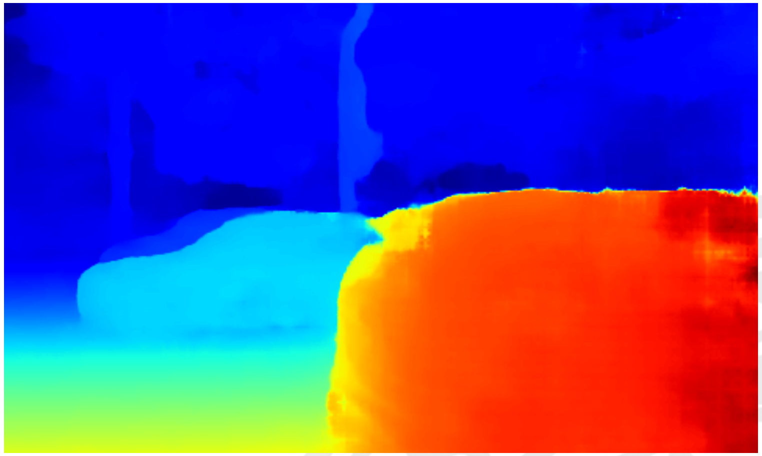
S. Enroth et. al. (2012), Combinations of Histone Modifications Mark Exon Inclusion Levels, PLoS ONE

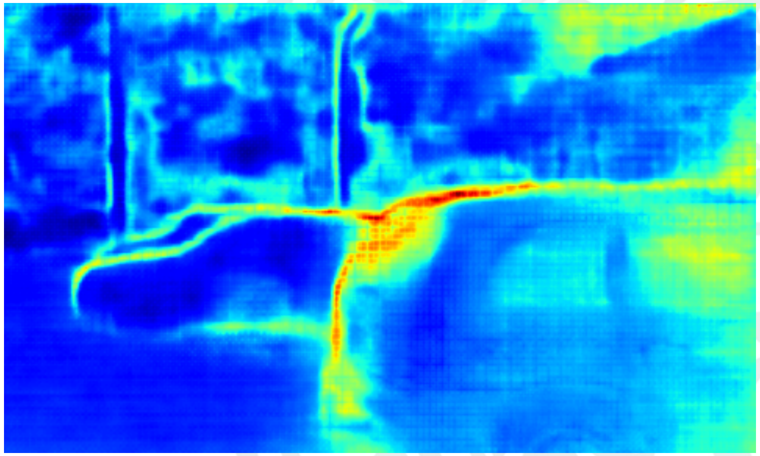
Dealing with uncertainty



Uncertainty

(1)


(2)


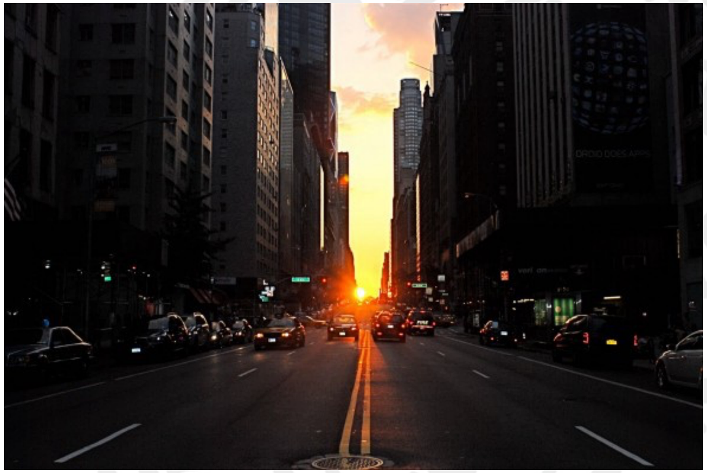
(3)


Uncertainty of depth estimation:

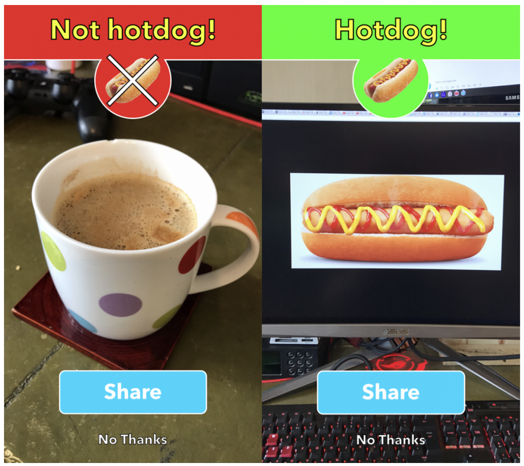
1. Original picture.
2. Estimated depth.
3. Estimation uncertainty.

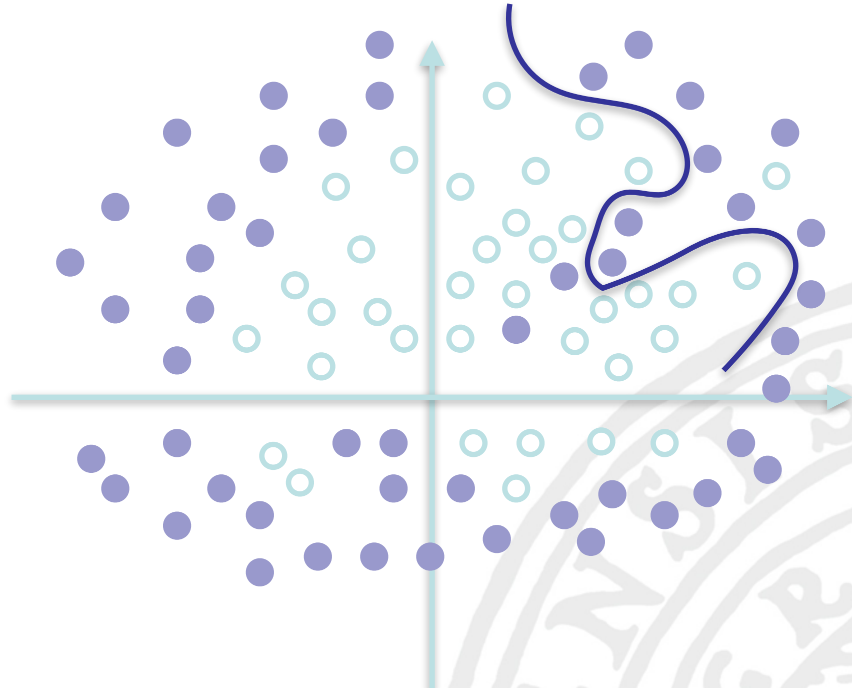
Aleatoric uncertainty

Aleatoric uncertainty is the sensing uncertainty.



Epistemic uncertainty

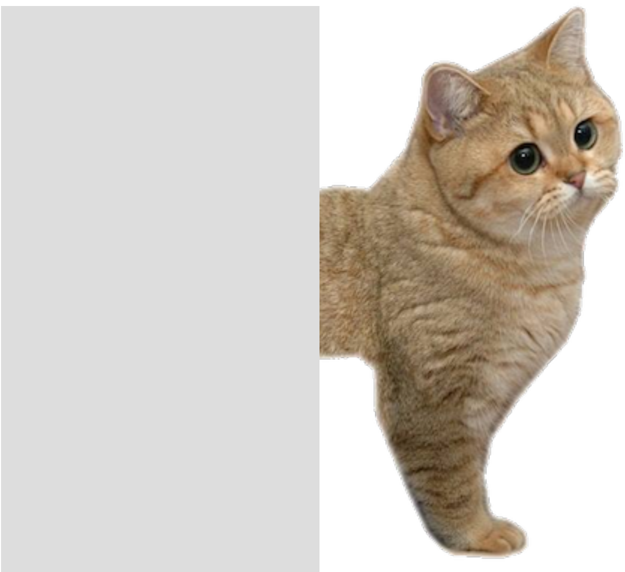


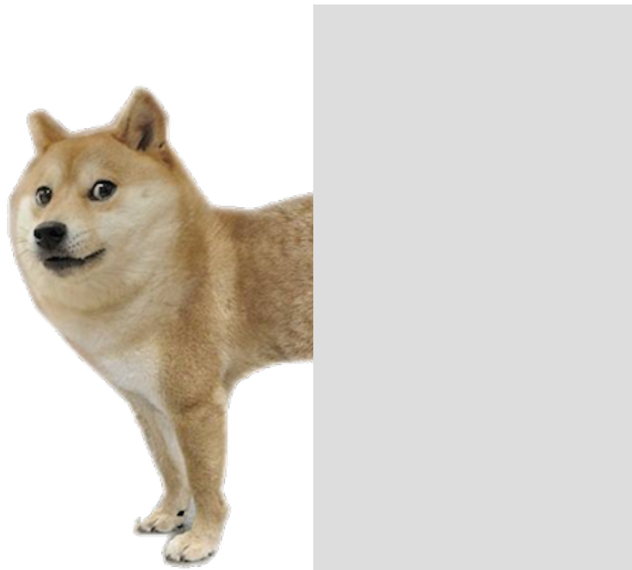


Epistemic uncertainty – related to the model.

- can be fixed given more training data

Assessing uncertainty



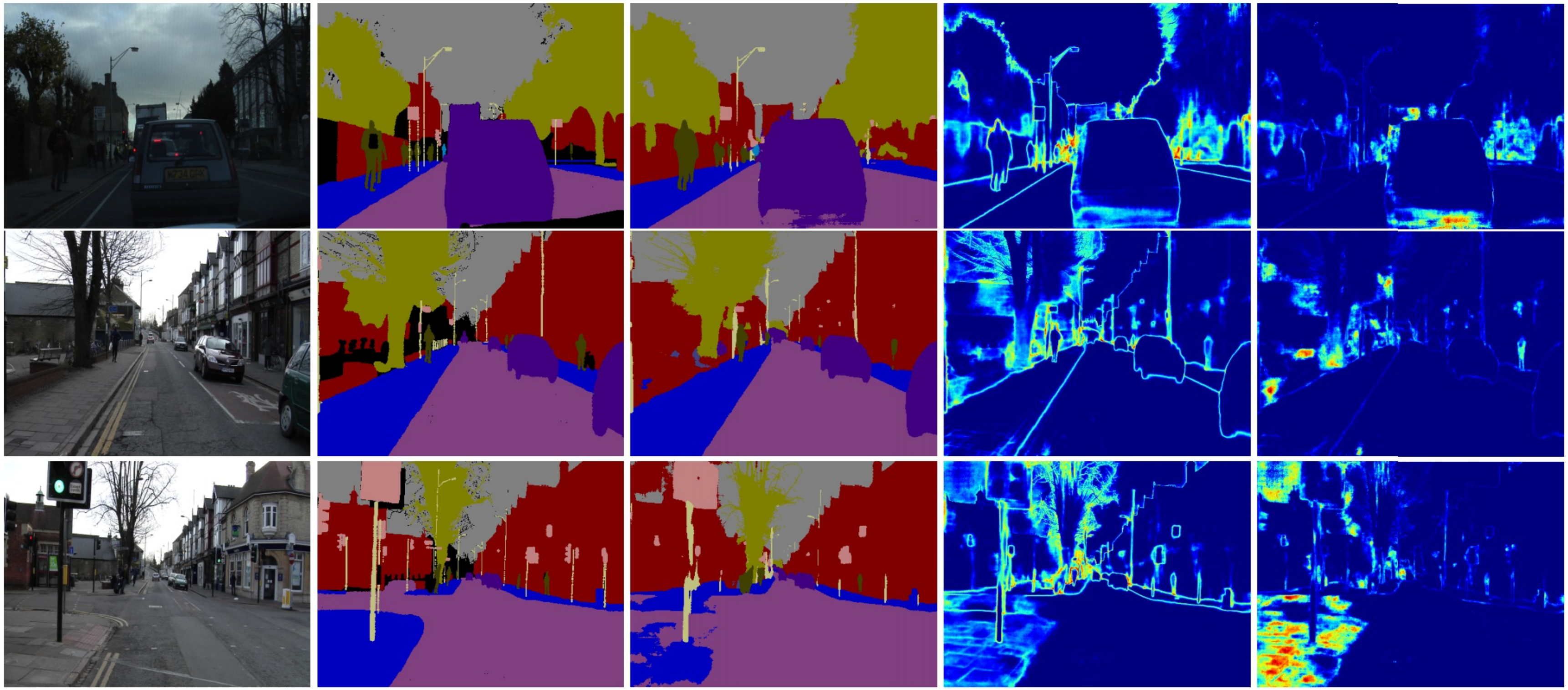


100% cat

100% dog

high epistemic uncertainty

Uncertainty – applications



(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

Final remarks

History

Machine Learning

Feature engineering

Feature selection

Overfitting

Performance

Interpretability

Explainability

Uncertainty

UPPSALA UNIVERSITET

Thank You!