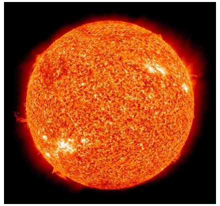
A gentle introduction to machine learning

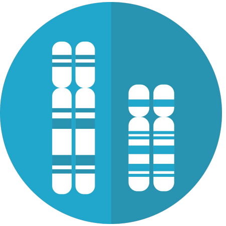
Hylke C. Donker, PhD
Medical statistics and decision making, Epidemiology, UMCG.
Computational Biology & Digital Sciences, Boehringer Ingelheim

umcg

Thinking inside the box



Measurement
Bayesian statistics
(posterior)



Discovery
Hypothesis tests
(*p*-values)

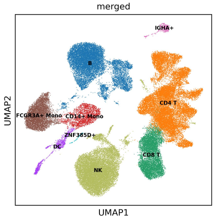
Explicit

Model of the world

Descriptive
(dimension reduction)

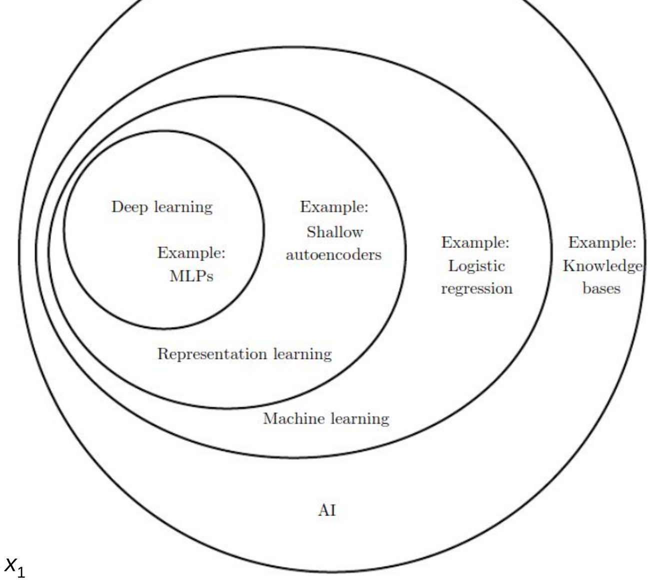
Implicit

Decision
Machine learning
(prediction)

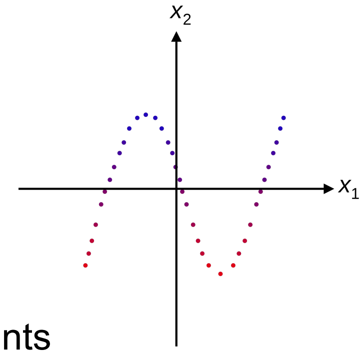


Machine Learning

Machine learning (**ML**) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks.

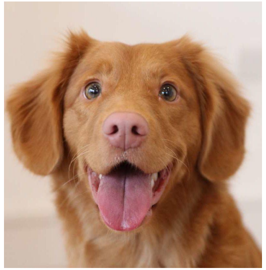


- **Expressive**: beyond linear regression
- **Versatile**: image, text, audio, etc.
- **Adaptible**: stand on the shoulders of giants



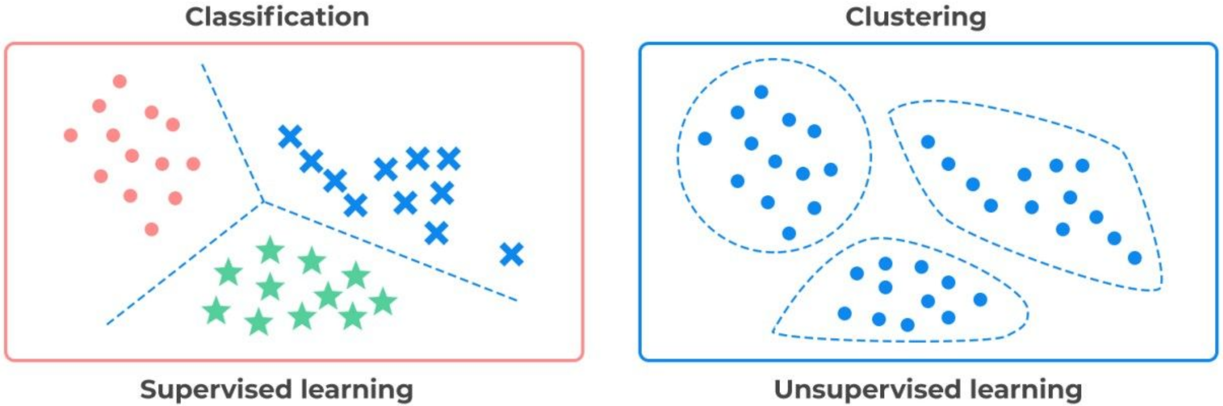
Machine learning

VS

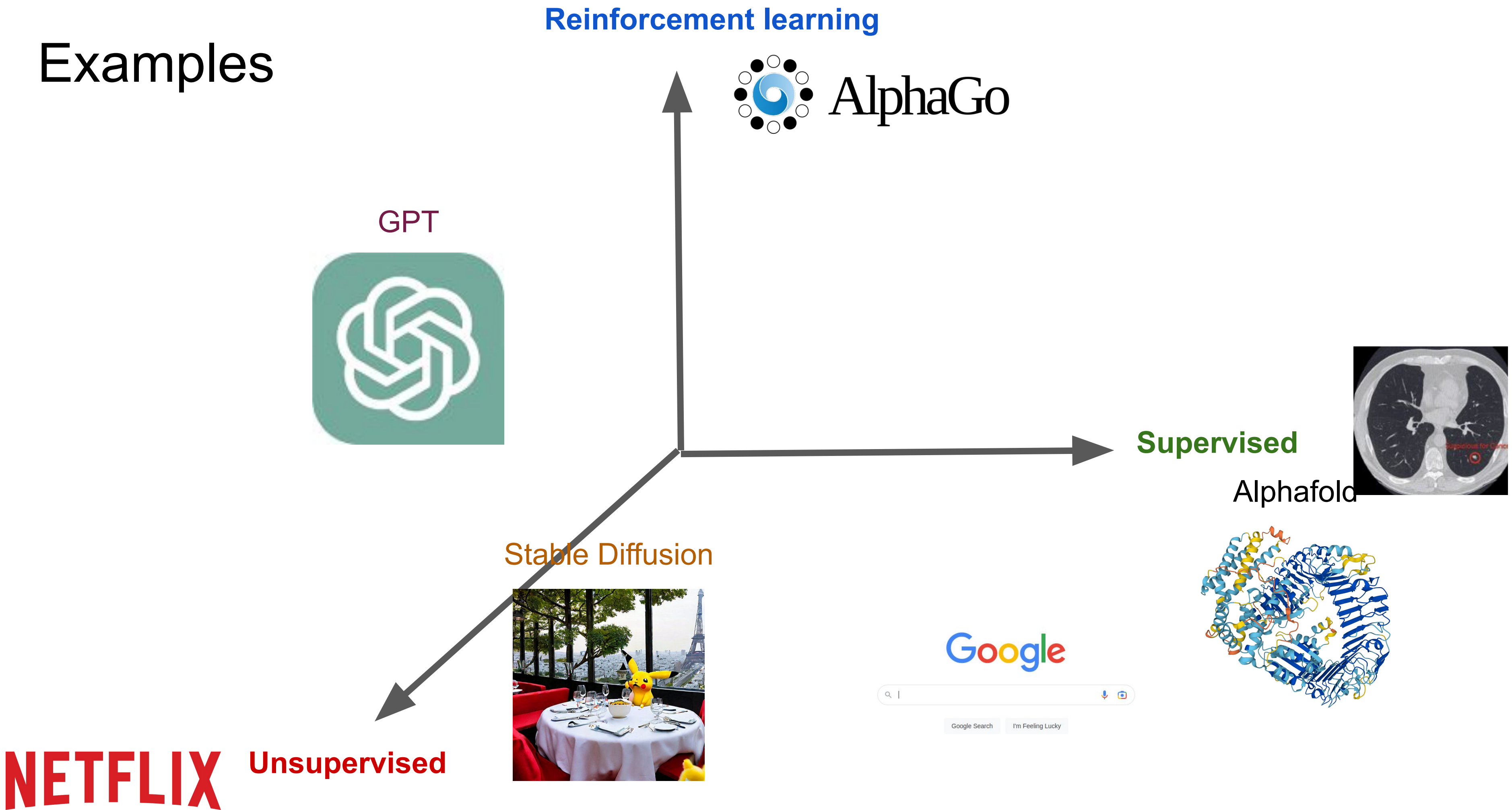


Types of learning:

- Supervised
- Unsupervised
- Reinforcement



Examples


Reinforcement learning
AlphaGo
GPT
Supervised
Alphafold
Stable Diffusion
Google
Google Search
I'm Feeling Lucky
NETFLIX
Unsupervised

Nomenclature

Number of samples / patients / subjects: $i = 1..m$.

Number of features: j = 1..n.

$x^{(i)}_j$: **Feature**, covariate, exogenous variable.

$y^{(i)}$: Target **label**, ground truth, class membership.

Binary classification: **$y^{(i)}$=1 positive** class; **$y^{(i)}$=0 negative** class.

Train - Dev - Test splits

- **Training set** (60 %)
  - Model fitting
- **Development or validation set** (20 %)
  - Modelling choices
- **Test or hold-out set** (20 %)
  - Performance (accuracy, ROC AUC, etc.)

Sample/feature	x_1	…	x_n
Patient 1			
…			
Patient *10*			

Demo Vertex AI

- No code AI training on Google Cloud Platform
- https://console.cloud.google.com/
- Build a dataset to train a classifier:
    - Cats (y=1) vs Dogs (y=0)

Confusion matrix

		Predicted	
		Cat (y'=1)	**Dog (y'=0)**
Actual	**Cat (y=1)**	**TP** (true positive)	**FN** (False negative)
	Dog (y=0)	**FP** (False positive)	**TN** (true negative)

Precision / positive predictive value: **TP** / (**TP** + **FP**)

Out of all predicted cats, how many did we get right?

Recall / sensitivity / true positive rate: **TP** / (**TP** + **FN**) = **TP**/**P**

Out of all cats, how many did we get right?

False positive rate / fall out: **FP** / (**TN** + **FP**) = **FP**/**N**

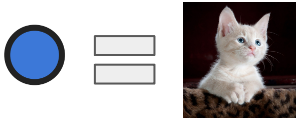
Out of all dogs, how many did we predict cat?

Precision recall curve

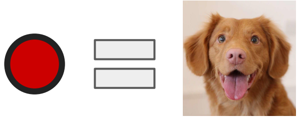
Precision: **TP** / (**TP** + **FP**)

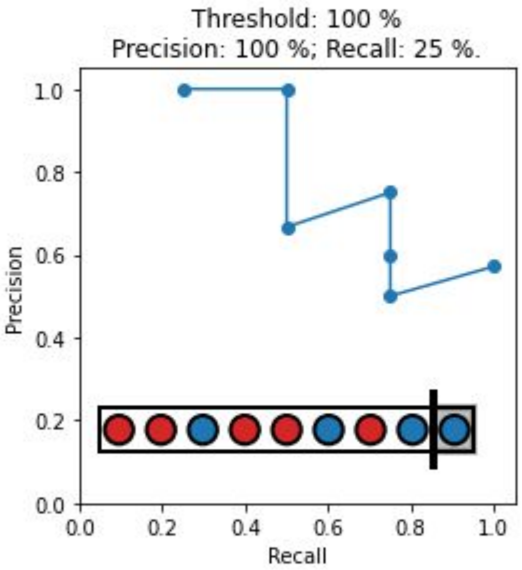
Recall: **TP** / (**TP** + **FN**) = **TP**/**P**

Positive label:



Negative label:

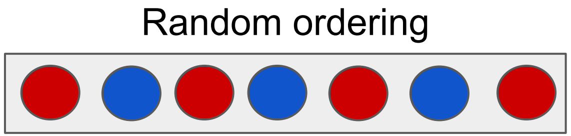


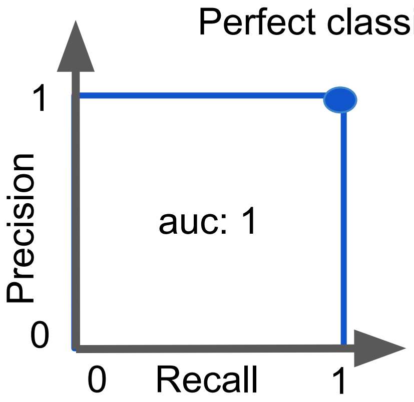


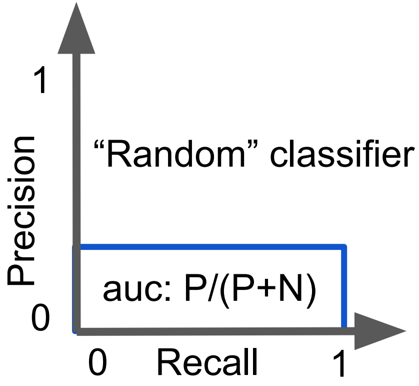
Precision recall curve

Precision: **TP** / (**TP** + **FP**)

Recall: **TP** / (**TP** + **FN**) = **TP**/**P**





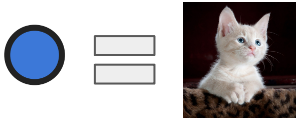


Receiver operating characteristic curve

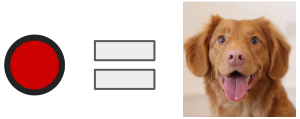
False positive rate: **FP** / (**TN** + **FP**) = **FP**/**N**

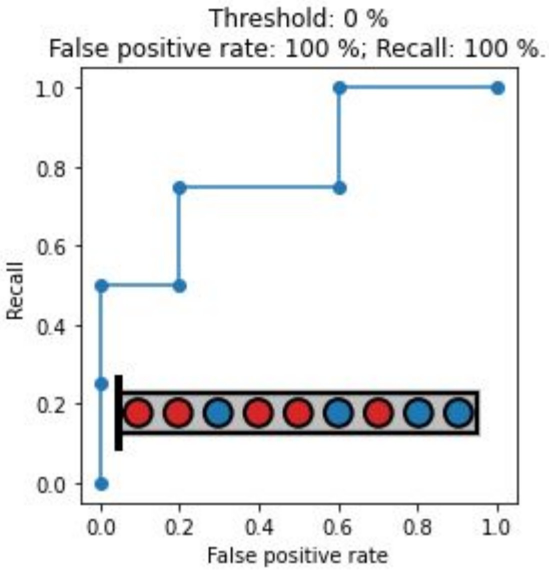
Recall: **TP** / (**TP** + **FN**) = **TP**/**P**

Positive label:



Negative label:





Demo Vertex AI

- https://console.cloud.google.com/
- Train a cats vs dogs classifier

Shapley values

- Attribute impact of each feature on final model prediction
- Concept from game theory

Example:
- Company, set *N* of three employees: {A, B, C}
- Profit outcome: *v(N)*
- How does each employee contribute to *v(N)*?

Shapley values example

Assume we can determine *v(N)* for the combinations:

$v(\{\}) = 0$
$v(\{A\}) = 10$
$v(\{B\}) = 20$
$v(\{C\}) = 30$
$v(\{A, B\}) = 60$
$v(\{B, C\}) = 70$
$v(\{A, C\}) = 90$
$v(\{A, B, C\}) = 100$

1. **$\{\} \rightarrow \{A\} \rightarrow \{A, B\} \rightarrow \{A, B, C\}$**
2. $\{\} \rightarrow \{A\} \rightarrow \{A, C\} \rightarrow \{A, B, C\}$
3. $\{\} \rightarrow \{B\} \rightarrow \{A, B\} \rightarrow \{A, B, C\}$
4. $\{\} \rightarrow \{B\} \rightarrow \{B, C\} \rightarrow \{A, B, C\}$
5. $\{\} \rightarrow \{C\} \rightarrow \{B, C\} \rightarrow \{A, B, C\}$
6. $\{\} \rightarrow \{C\} \rightarrow \{A, C\} \rightarrow \{A, B, C\}$

Trail 1:

- ΔA: $v(\{A\}) - v(\{\}) = 10 - 0 = 10$
- ΔB: $v(\{A, B\}) - v(\{A\}) = 60 - 10 = 50$
- ΔC: $v(\{A, B, C\}) - v(\{A, B\}) = 100 - 60 = 40$

Shapley values example

1. {} → {A} → {A, B} → {A, B, C} || A = 10, B = 50, C = 40
2. {} → {A} → {A, C} → {A, B, C} || A = 10, B = 10, C = 80
3. {} → {B} → {A, B} → {A, B, C} || A = 40, B = 20, C = 40
4. {} → {B} → {B, C} → {A, B, C} || A = 30, B = 20, C = 50
5. {} → {C} → {B, C} → {A, B, C} || A = 30, B = 40, C = 30
6. {} → {C} → {A, C} → {A, B, C} || A = 60, B = 10, C = 30

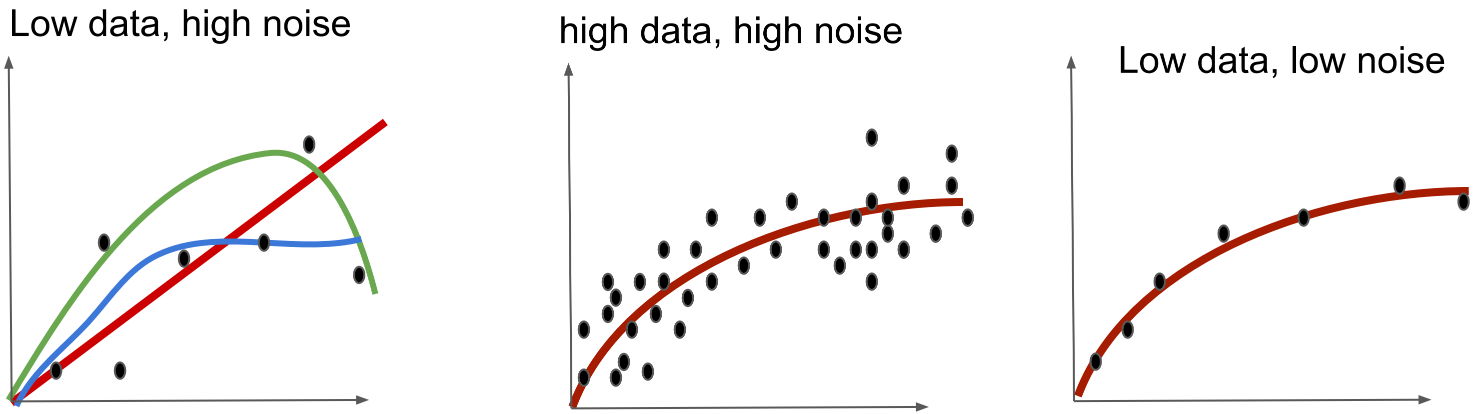
<A> = 30; <B> = 25; <C> = 45

Demo Vertex AI

- https://console.cloud.google.com/
- Evaluate classifier

Tips and Tricks

Data centric AI



Solution: Data polishing (!!only on **train** split!!)

- Take out bad samples
- Make synthetic data (scale, crop, rotate)
- Find slices of data that need improvement.

https://datacentricai.org/

Transparent reporting: TRIPOD

- TRIPOD [1]: Checklist for reporting prediction models in a scientific paper.

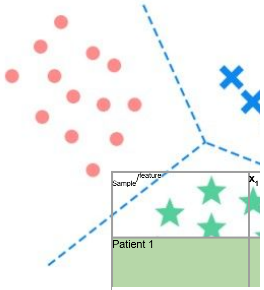
  https://www.tripod-statement.org/

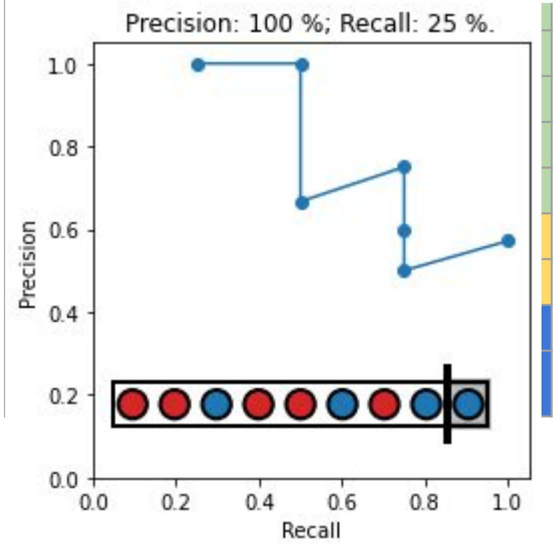
  - Source of data
  - Participants
  - Outcome
  - Performance
  - Limitations

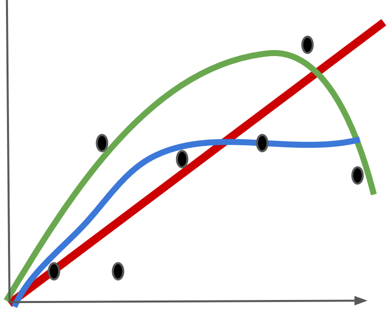
[1]: Collins, Gary S., et al. "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement." *Journal of British Surgery* 102.3 (2015): 148-158.

Conclusion

- Machine learning: statistics for decisions
- Training AI on Google Cloud Platform
- Train, dev, and test splits
- Precision-recall and ROC curves
- Shapley values
- Data centric AI & TRIPOD







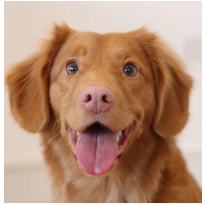
Questions

?

Exercises

Confusion matrix

VS



Probability(cat)	Is it a Cat?
0	0
0.125	0
0.25	1
0.375	0
0.5	0
0.625	1
0.75	0
0.875	1
1	1

Exercise (1)

Assume: probability threshold 0.5

- Predict cat when probability(cat) ≥ 0.5
- Compute precision: **TP** / (**TP** + **FP**)
- Compute recall: **TP** / (**TP** + **FN**) = **TP**/**P**
- Compute false positive rate: **FP** / (**TN** + **FP**) = **FP**/**N**

Exercise Shapley values

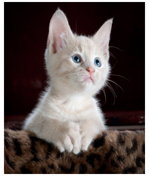
v ({}) = 0
v ({A}) = 10
v ({B}) = 20
v ({C}) = 30
v ({A, B }) = 60
v ({B, C }) = 70
v ({A, C }) = 90
v ({A, B , C }) = 100

Exercise (2)

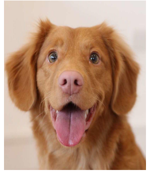
Trail 2: {} $\rightarrow$ {A} $\rightarrow$ {A, C} $\rightarrow$ {A, B, C}

- ΔA:
- ΔB:
- ΔC:

Decision curve analysis



VS



net benefit(π) = TP(π)/m — FP(π)/m · π/(1-π).

Exercise (3):

Assume: probability threshold π = 0.5,

- Compute net benefit

Probability(cat)	Is it a Cat?
0	0
0.125	0
0.25	1
0.375	0
0.5	0
0.625	1
0.75	0
0.875	1
1	1

Solutions

Solution (1)

TP(threshold: 0.5) = 3

FP(threshold: 0.5) = 2

P = 4

N = 5

Precision: **TP** / (**TP** + **FP**) = 3 / 5 = **60 %**

Recall: **TP** / (**TP** + **FN**) = **TP**/**P** = **75 %**

False positive rate: **FP** / (**TN** + **FP**) = **FP**/**N** = **40 %**

Probability(cat)	**Predict(cat)**	**Cat?**
0	**0**	**0**
0.125	**0**	**0**
0.25	0	1
0.375	**0**	**0**
0.5	1	0
0.625	**1**	**1**
0.75	1	0
0.875	**1**	**1**
1	**1**	**1**

Exercise Shapley values

$v(\{\}) = 0$
$v(\{A\}) = 10$
$v(\{B\}) = 20$
$v(\{C\}) = 30$
$v(\{A, B\}) = 60$
$v(\{B, C\}) = 70$
$v(\{A, C\}) = 90$
$v(\{A, B, C\}) = 100$

Exercise (2)

Trail 2: $\{\} \to \{A\} \to \{A, C\} \to \{A, B, C\}$

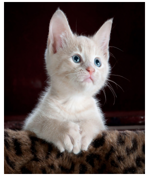
- ΔA:
- ΔB:
- ΔC:

Solution (2)

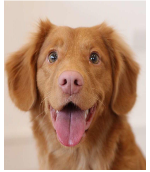
Trail 2: $\{\} \to \{A\} \to \{A, C\} \to \{A, B, C\}$

- ΔA: $v(\{A\}) - v(\{\}) = 10 - 0 = 10$
- ΔB: $v(\{A, B, C\}) - v(\{A, C\}) = 100 - 90 = 10$
- ΔC: $v(\{A, C\}) - v(\{A\}) = 90 - 10 = 80$

Decision curve analysis



vs



net benefit(π) = TP(π)/m — FP(π)/m · π/(1-π).

Exercise (3):

Assume: probability threshold 0.5

- Compute net benefit

Solution (3):

TP(π=0.5) = 3

FP(π=0.5) = 2

m = 9

net benefit(π=0.5) = 3/9 - 2/9 · 1 = **1/9**

Probability(cat)	Is it a Cat?
0	0
0.125	0
0.25	1
0.375	0
0.5	0
0.625	1
0.75	0
0.875	1
1	1

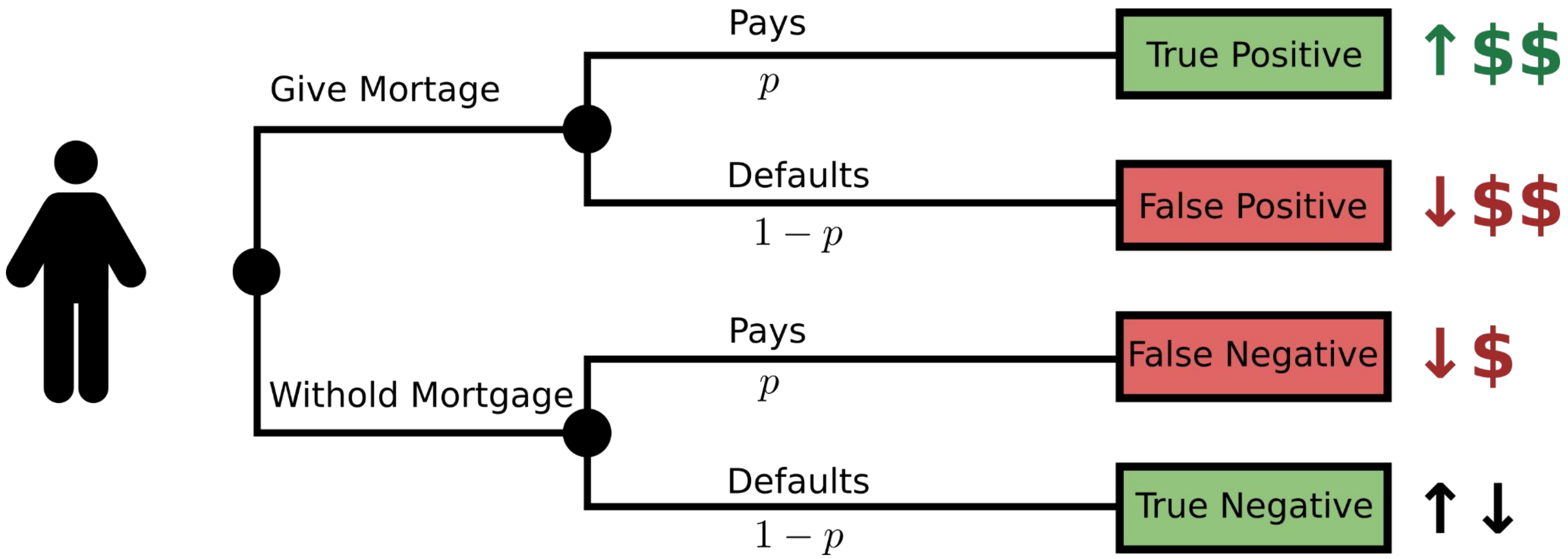
Backup slides

Meta learning

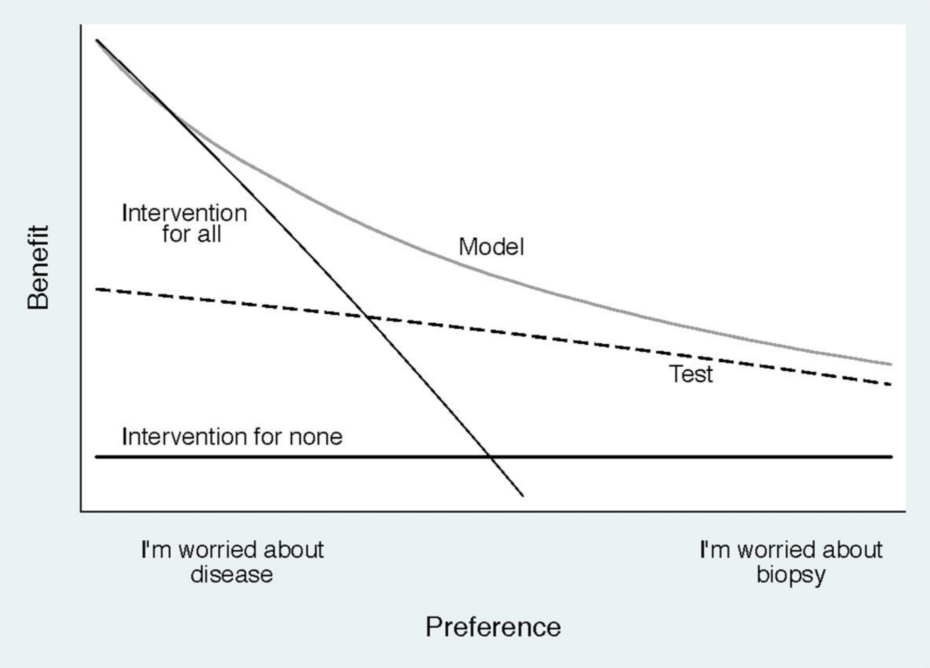
- Meta learning: models that create models
- Neural architecture search (NAS)
  - Search space
  - Search strategy

Decision curve analysis

- Measures **clinical utility**



net benefit(π) = TP(π)/m — FP(π)/m · π/(1-π).



Vickers, Andrew J., Ben van Calster, and Ewout W. Steyerberg. "A simple, step-by-step guide to interpreting decision curve analysis."

Diagnostic and prognostic research 3.1 (2019): 1–8.

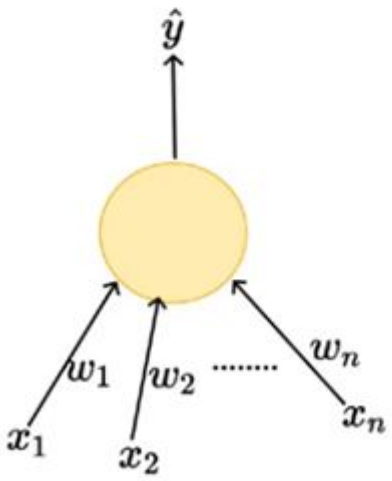
Transparent reporting (1/2): Model cards [1]

- Model cards [1]: documentation procedure for prediction models
    - Model details
    - Intended use
    - Factors (slices of the data)
    - Metrics
    - Evaluation data
    - Training data
    - Quantitative analyses
    - Ethical considerations
    - Caveats and Recommendations

[1]: Mitchell, Margaret, et al. "Model cards for model reporting." Proceedings of the conference on fairness, accountability, and transparency. 2019.

Neural network intuition

https://playground.tensorflow.org/



$$\hat{y} = 1 \text{ if } \sum_{i=1}^{n} w_i x_i \geq b$$

$$\hat{y} = 0 \text{ otherwise}$$

Explainable AI

XAI (explainable AI): “*ability to explain or to present in understandable terms to a human*” [1]

[1]: Doshi-Velez, Finale, and Been Kim. "Towards a rigorous science of interpretable machine learning." arXiv preprint arXiv:1702.08608 (2017).