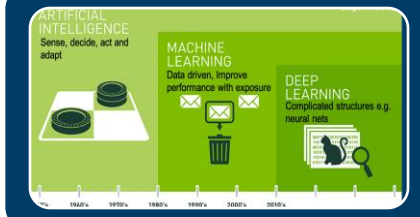
SINTEF

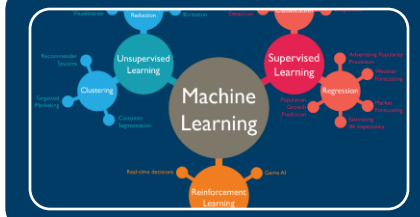
INTRODUCTION TO MACHINE LEARNING - KEY CONCEPTS AND TERMINOLOGY

Milan De Cauwer – SINTEF Digital – Norway, Oslo

Outline

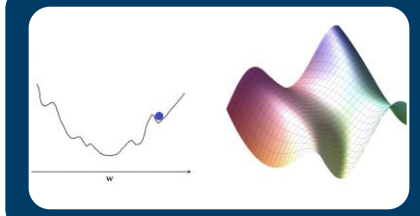


Discussion on Machine Learning and AI



Machine Learning – A definition

- Supervised learning
- Unsupervised learning
- Reinforcement learning



Concepts in machine learning

- Data – Features, labels, instance, dataset and data splitting
- Models – Training and inference
- Bias and variance trade-off

SINTEF

Who am I?

Milan
SINTEF Digital

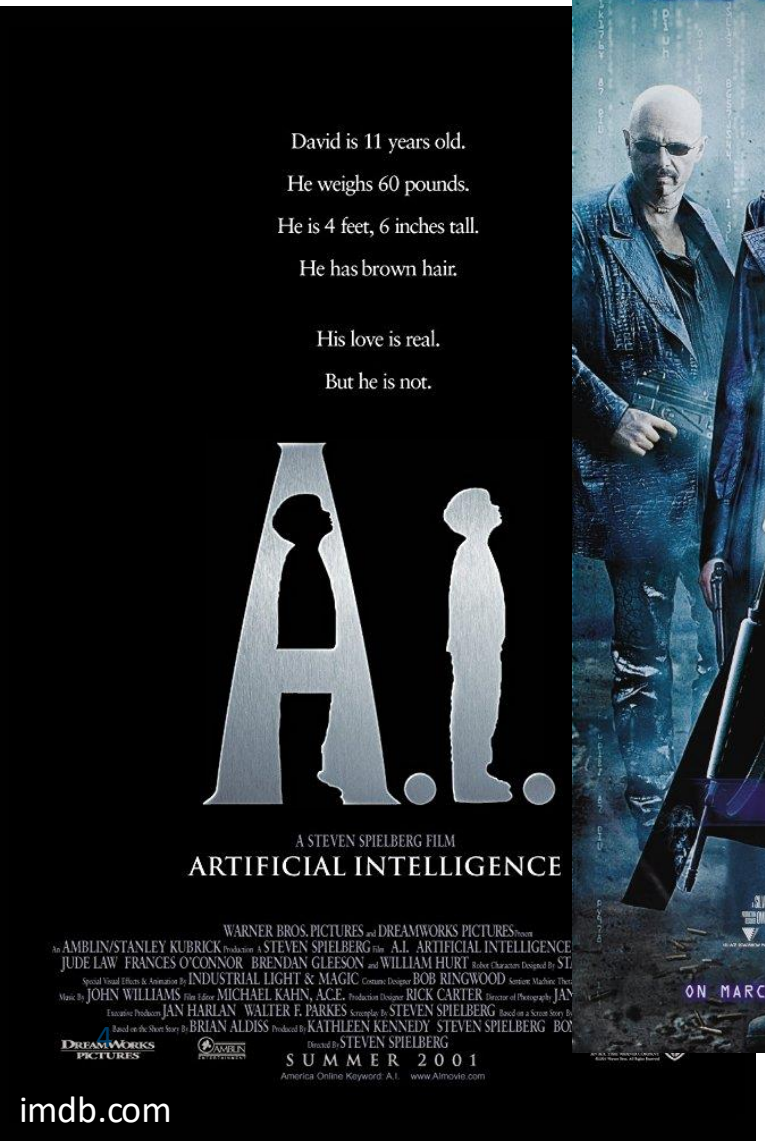
- Research scientist at SINTEF Digital, Norway
  - SINTEF is one of Europe's largest independent research organisations
  - Applied research in collaboration with industry

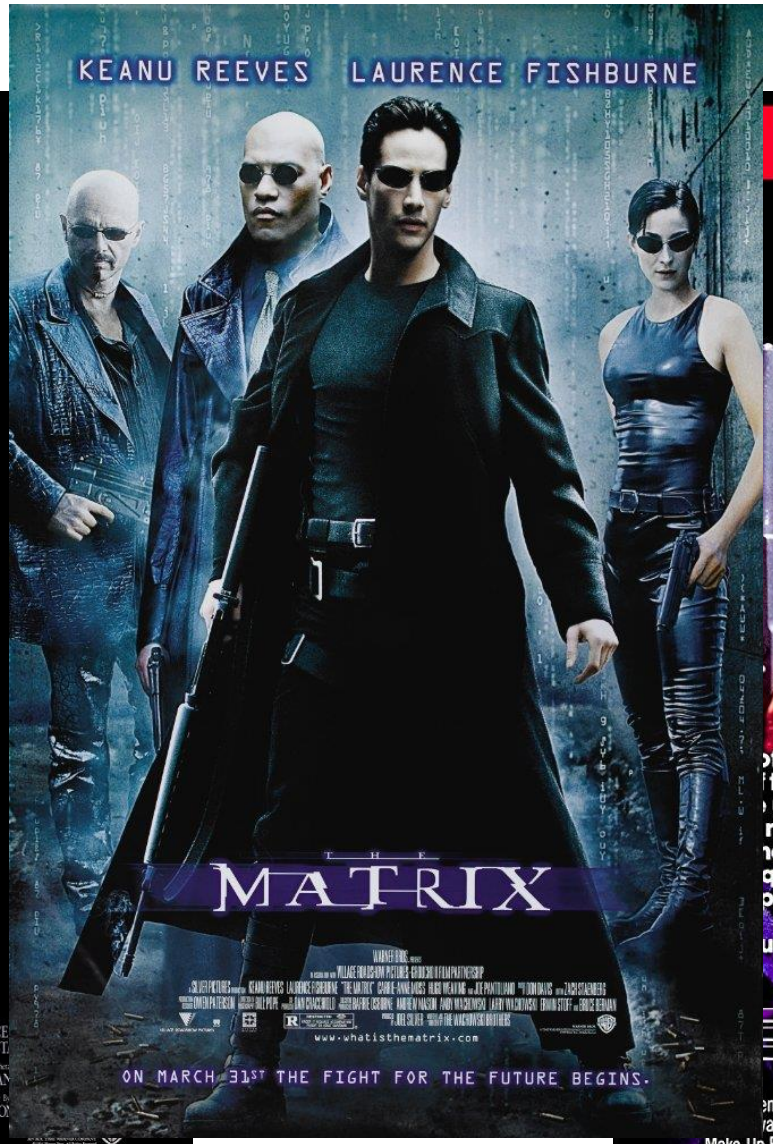
Building and construction
Climate and environment
Digitalisation
Food and agriculture
Health and medicine
Materials
Microsystems and nanotechnology
Ocean space
Renewable energy
Society and security
Sustainable production
Transport and mobility

- My research interests
  - Applications of **Machine Learning** for modeling and solving **combinatorial problems**

SINTEF

AI in pop culture





imdb.com

SINTEF

Examples of AI in daily life

Search engines

Filters

Media suggestions

Assistants

Online shopping

Examples of AI in daily life – ChatGPT (Large language models)

Search engines

Filters

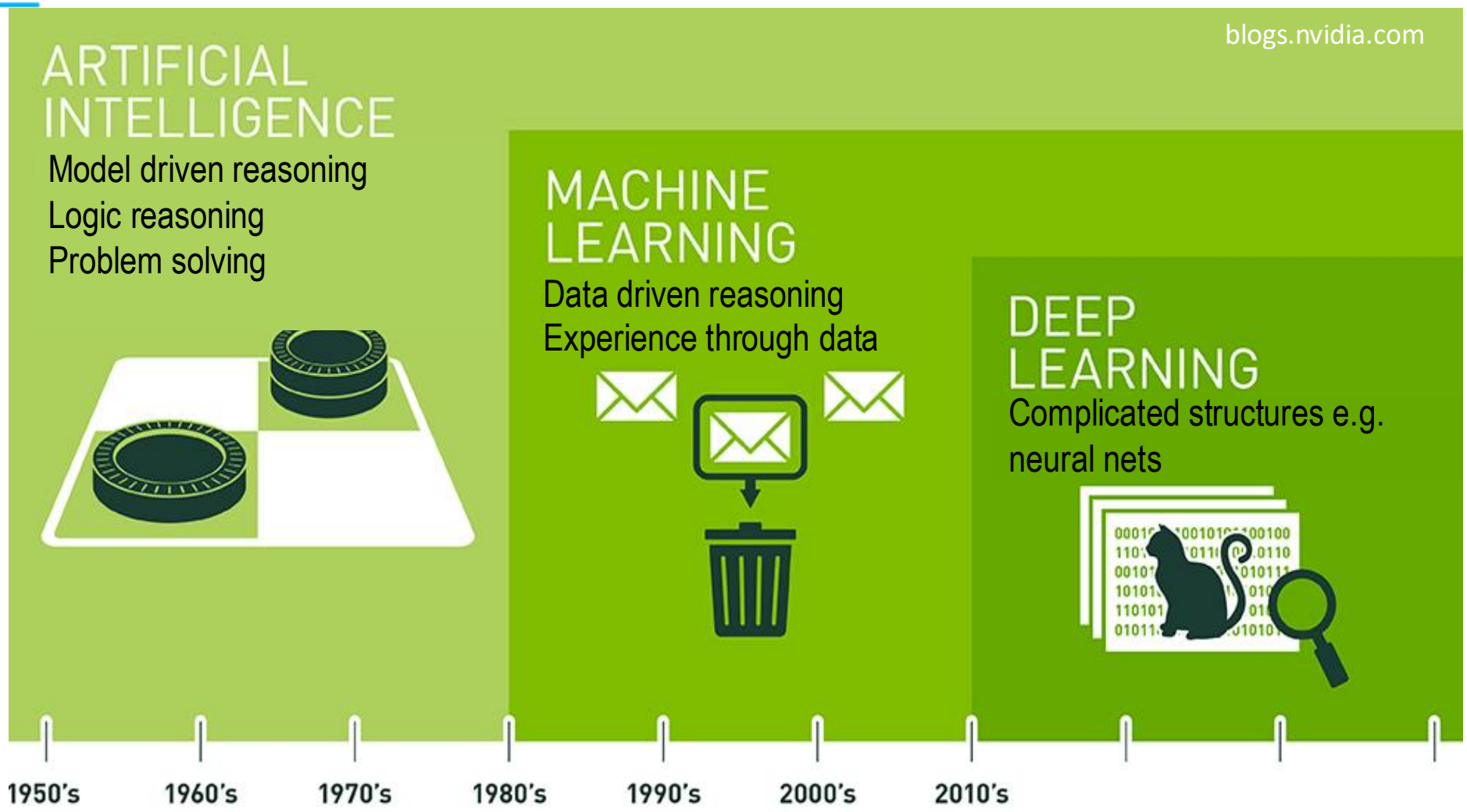
Media suggestions

Assistants

Online shopping

SINTEF

AI - Machine learning – Deep learning



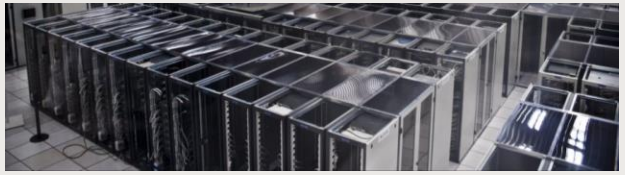
SINTEF

Why now?

Three paradigm shifts at once

Data

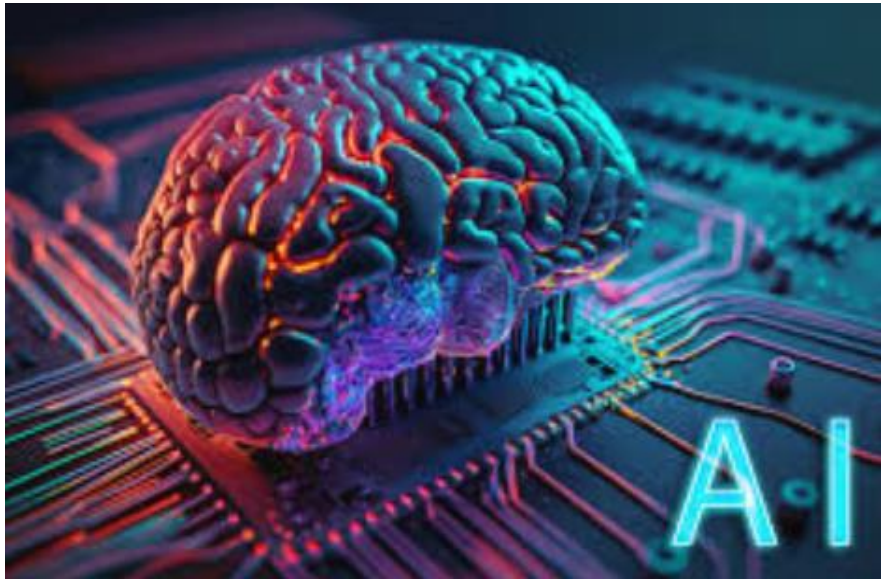
Technology

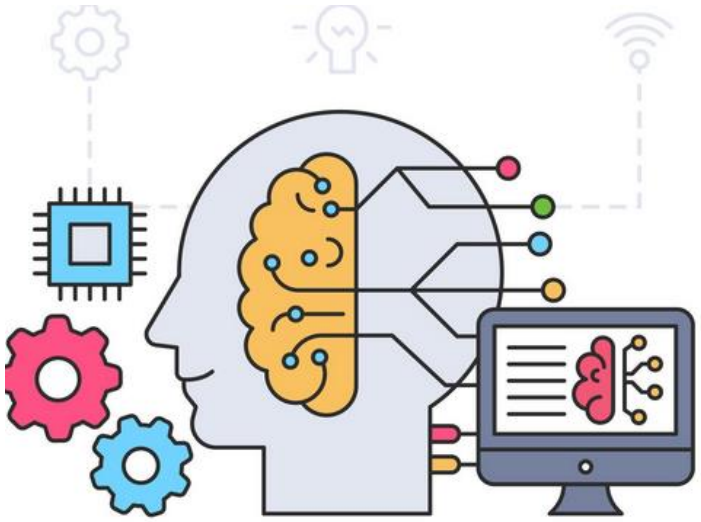


Analytics

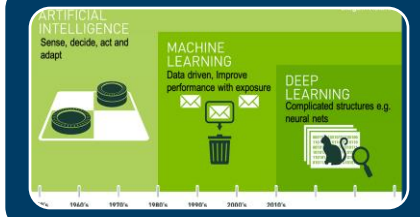
Machine learning

Quiz – Do all methods in AI rely on machine learning? Why?

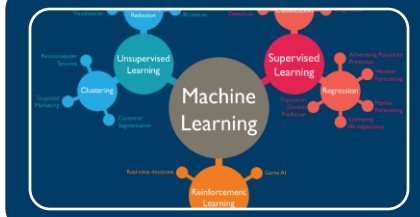




Outline

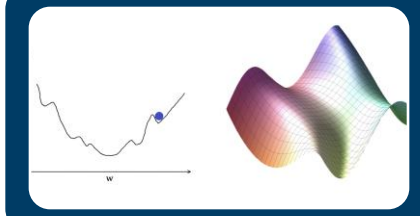


Discussion on Machine Learning and AI



Machine Learning – A definition

- Supervised learning
- Unsupervised learning
- Reinforcement learning



Concepts in machine learning

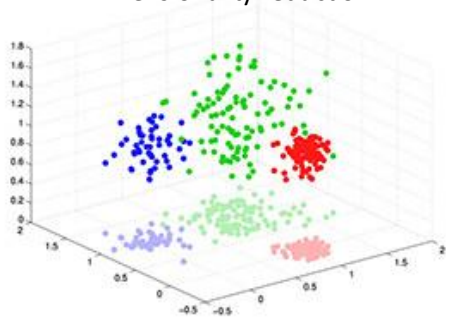
- Data – Features, labels, instance, dataset and data splitting
- Models – Training and inference
- Bias and variance trade-off

SINTEF

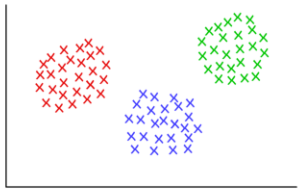
Machine-learning – A definition

- Definition:
  - A set of methods that enable computers to learn (something useful) from data
- Goal:
  - Build models that can make predictions or decisions based on data

Dimensionality reduction



Clustering



Machine Learning

Unsupervised Learning

- Dimensionality Reduction
  - Meaningful Compression
  - Structure Discovery
  - Big data alistaion
  - Feature Elicitation
- Clustering
  - Recommender Systems
  - Targetted Marketing
  - Customer Segmentation

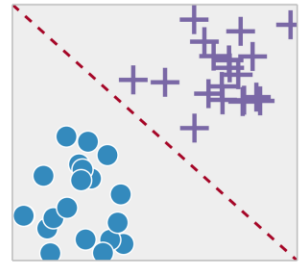
Supervised Learning

- Classification
  - Image Classification
  - Customer Retention
  - Idenity Fraud Detection
  - Diagnostics
- Regression
  - Advertising Popularity Prediction
  - Weather Forecasting
  - Population Growth Prediction
  - Market Forecasting
  - Estimating life expectancy

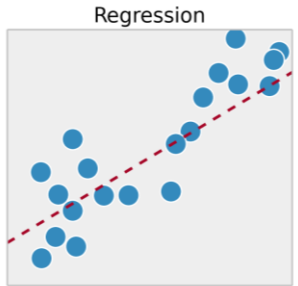
Reinforcement Learning

- Real-time decisions
- Game AI
- Robot Navigation
- Skill Acquisition
- Learning Tasks

Classification



Regression



towardsdatascience.com

SINTEF

Supervised learning

Duck Duck Rabbit Hedgehog

Supervised learning

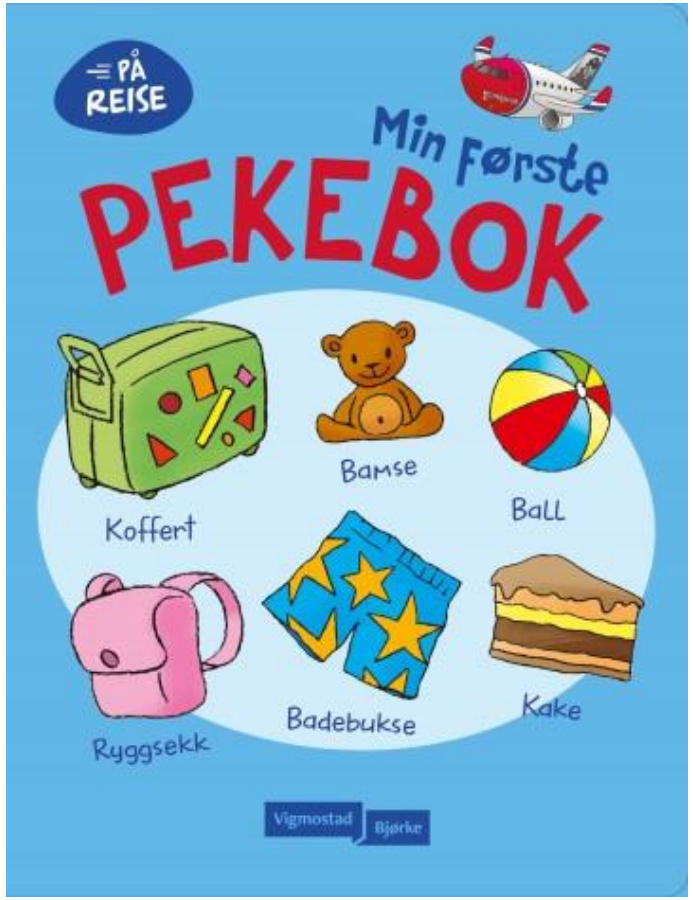
Predictive model

?

Predictive model

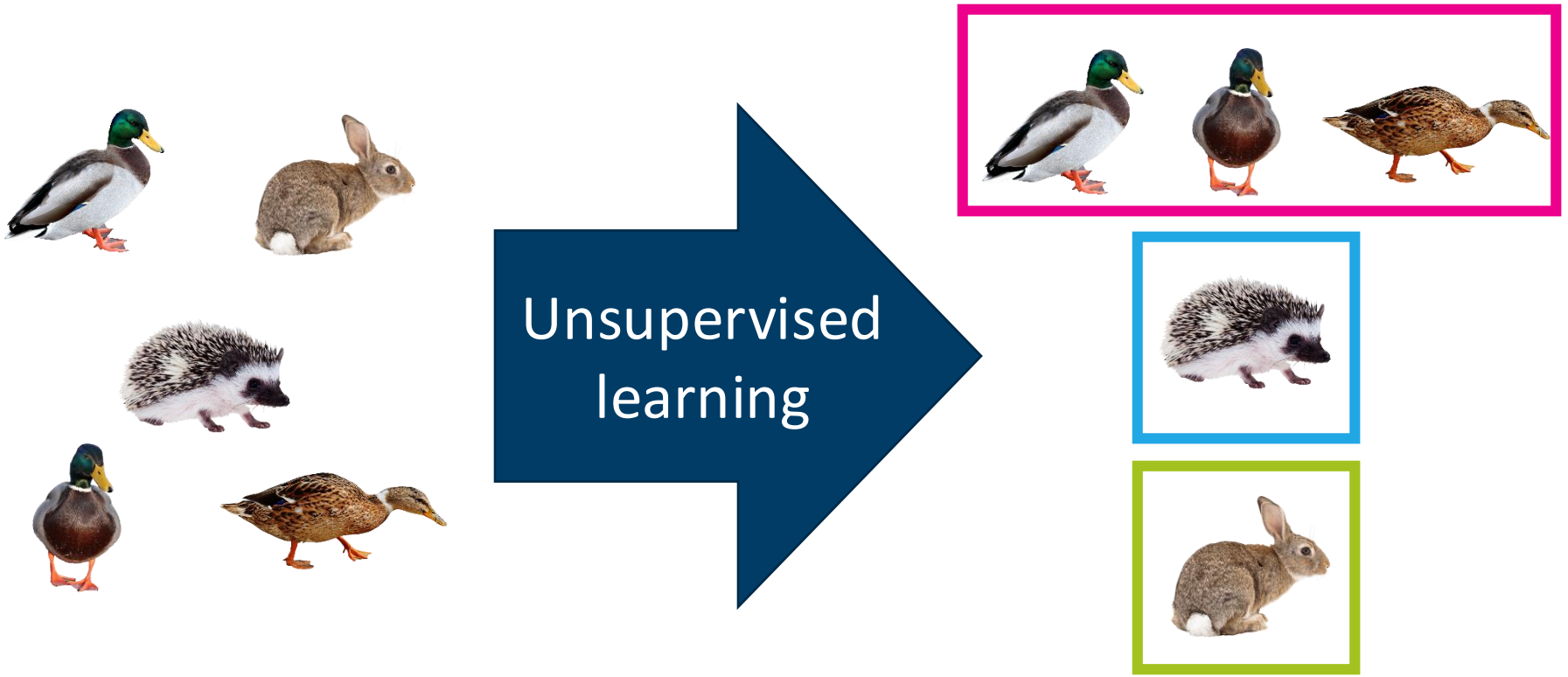
Duck

Example of supervised learning



tanum.no

Unsupervised learning



SINTEF

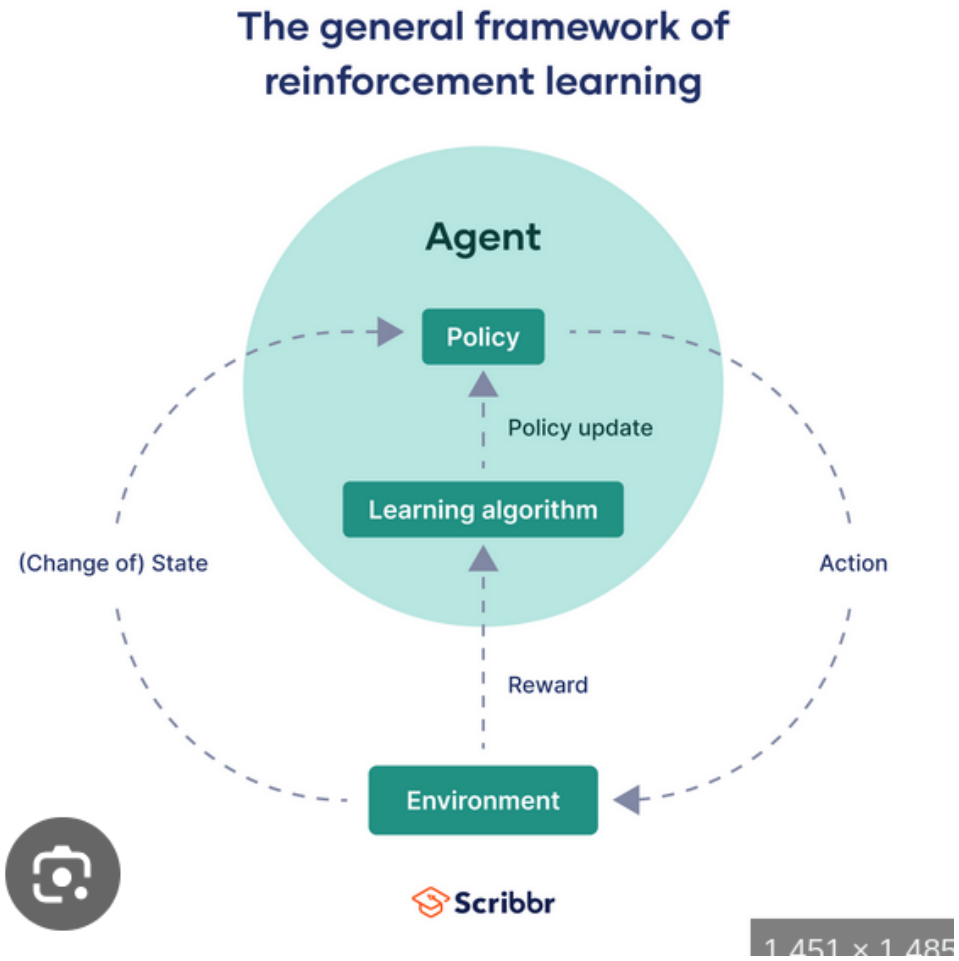
Example on un-supervised learning

slate.com

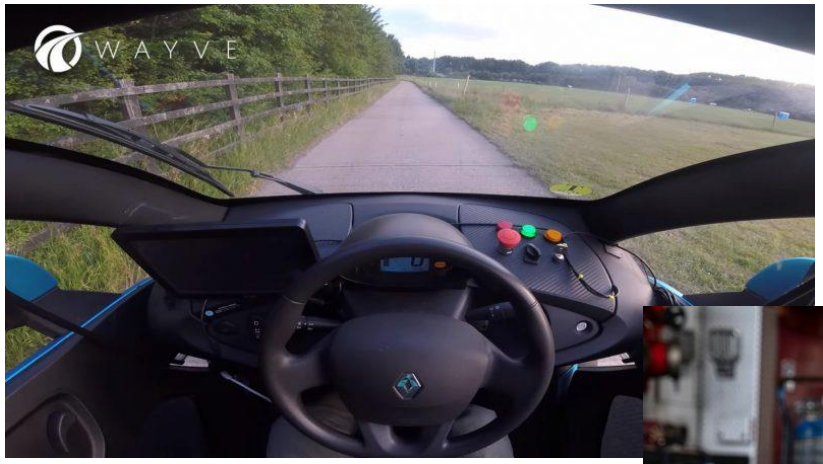
SINTEF

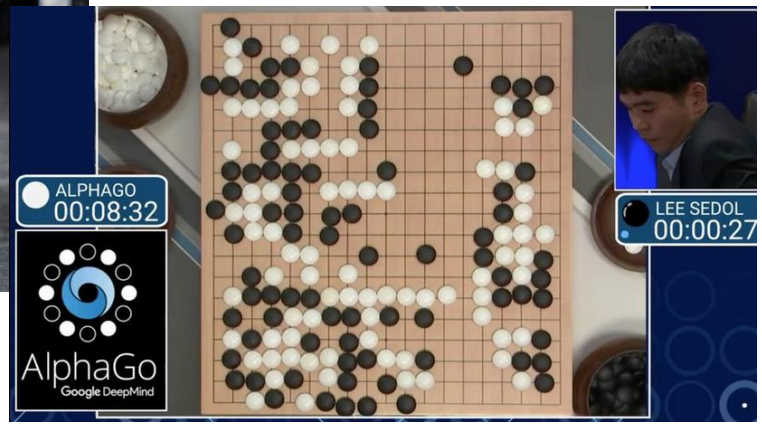
Reinforcement learning

- Goal: learn how to interact with an environment to achieve a specific objective.
- The agent receives rewards (and penalties) depending on the quality of its actions



Reinforcement learning examples: Self-driving cars, robotics and games

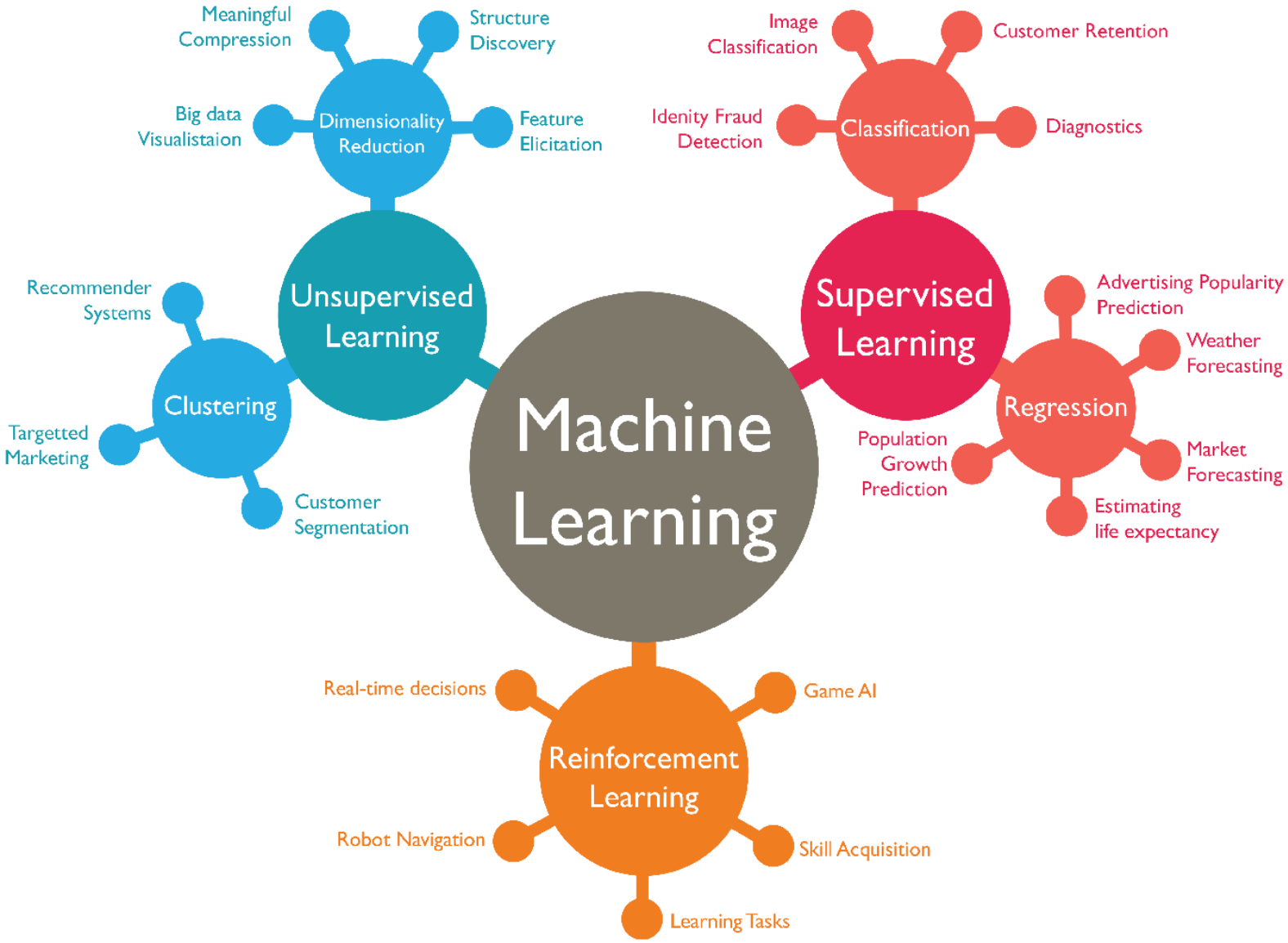




SINTEF

Quiz – Use-cases

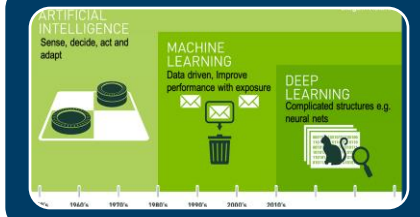
- **A:** Predicting whether the train will be late
- **B:** Teaching a computer to play angry birds
- **C:** Recommending your next dining experience.



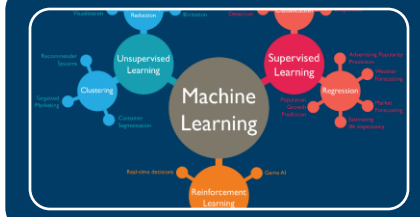
towardsdatascience.com

SINTEF

Outline

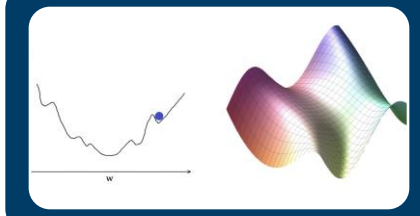


Discussion on Machine Learning and AI



Machine Learning – A definition

- Supervised learning
- Unsupervised learning
- Reinforcement learning



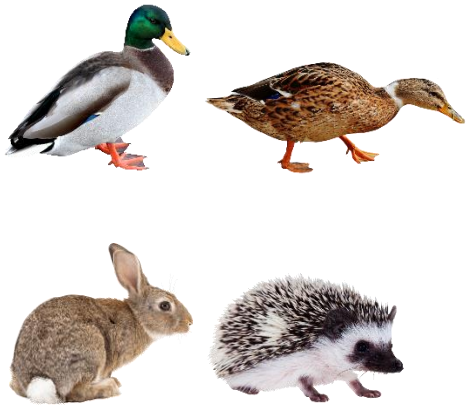
Concepts in machine learning

- Data – Features, labels, instance, dataset and data splitting
- Models – Training and inference
- Bias and variance trade-off

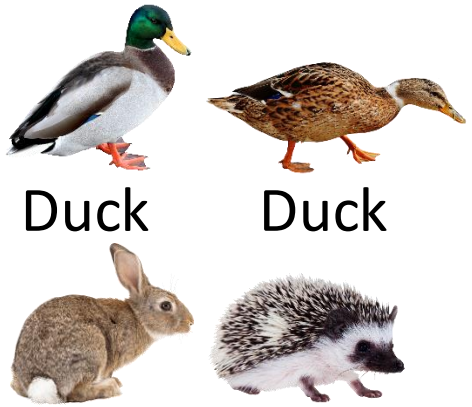
SINTEF

Different kind of data

Unlabelled data



Labelled data

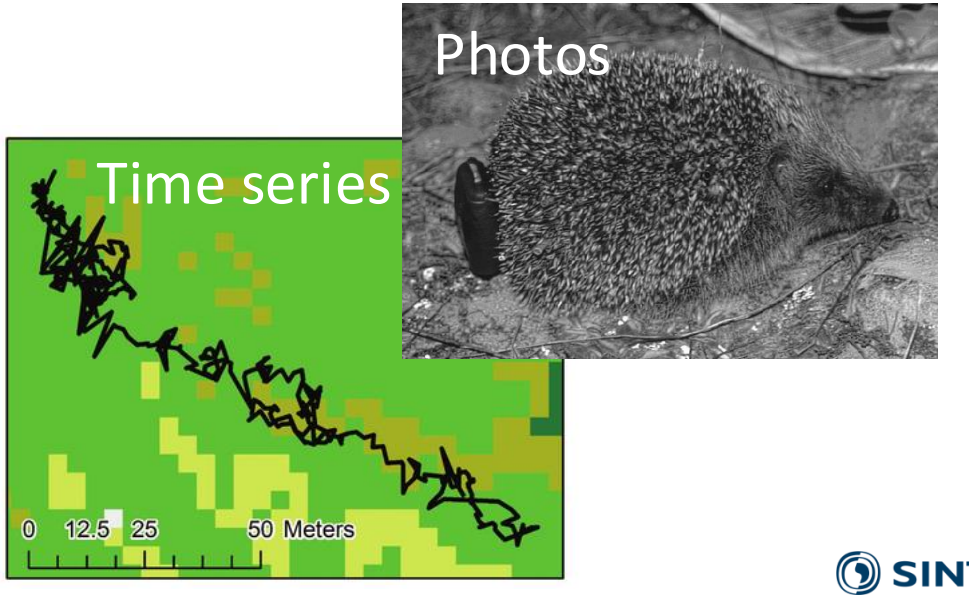


Duck Duck

Rabbit Hedgehog

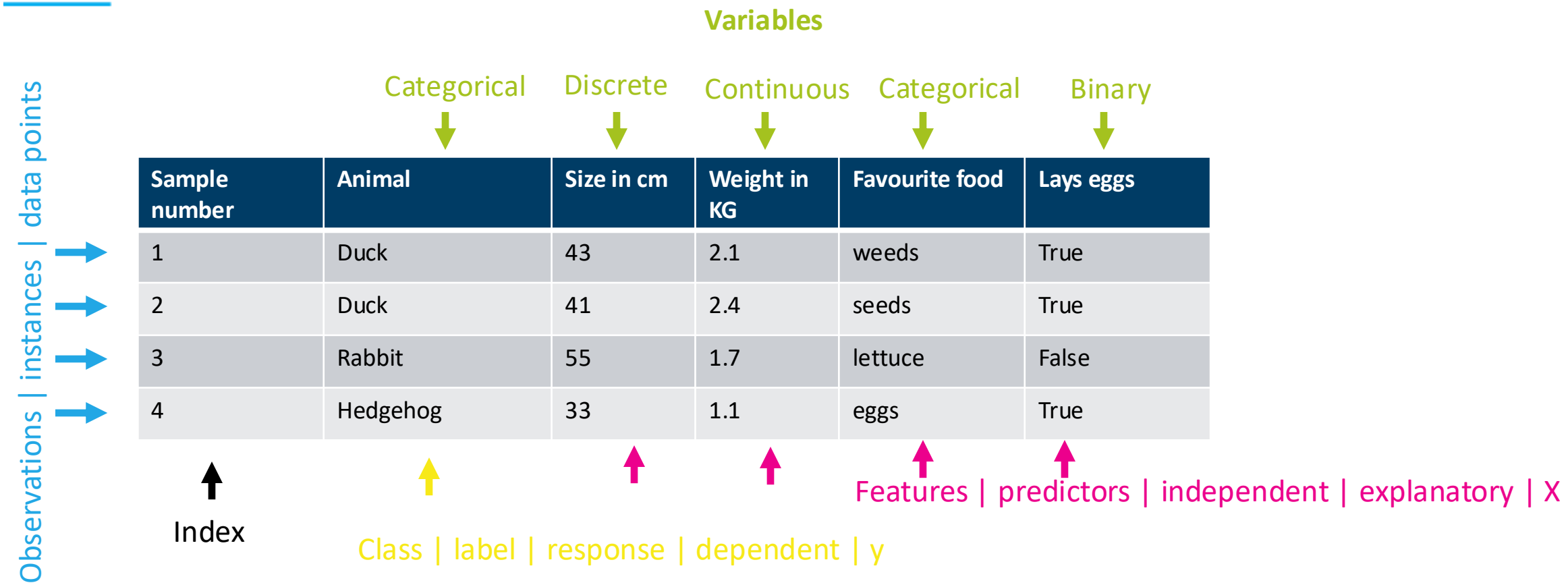
Tabular data

Image number	Animal	Size in cm	Favourite food	Lays eggs
1	Duck	43	weeds	True
2	Duck	41	seeds	True
3	Rabbit	55	lettuce	False
4	Hedgehog	33	eggs	True



SINTEF

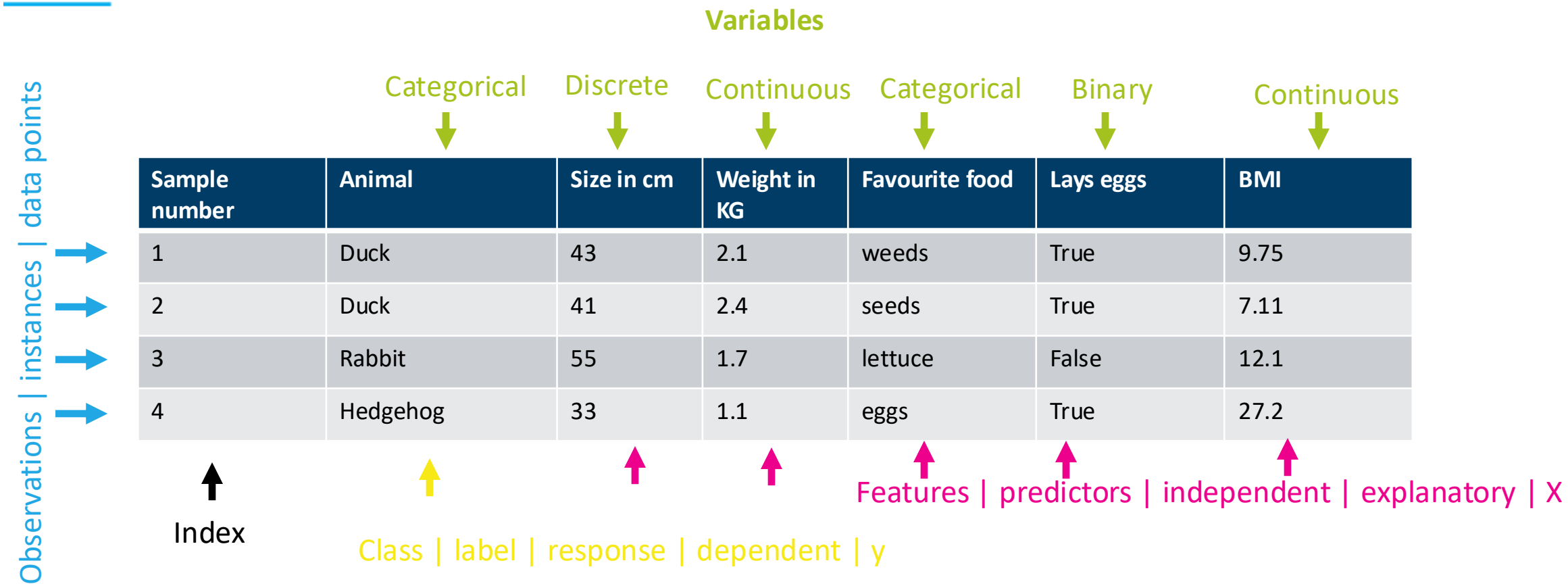
Variables and features



SINTEF

Feature Engineering

BMI = Weight / Size^2



Sample number	Animal	Size in cm	Weight in KG	Favourite food	Lays eggs	BMI
1	Duck	43	2.1	weeds	True	9.75
2	Duck	41	2.4	seeds	True	7.11
3	Rabbit	55	1.7	lettuce	False	12.1
4	Hedgehog	33	1.1	eggs	True	27.2

SINTEF

Quiz -- Variables

	age	income	marital_status	subscribed
0	56	5.99	divorced	0
1	69	3.72	married	0
2	41	7.30	married	1
3	58	3.53	single	1
4	21	4.53	divorced	0
5	18	7.54	divorced	1
6	64	4.53	married	0
7	68	3.54	married	0
8	68	5.31	single	1
9	47	4.02	married	0

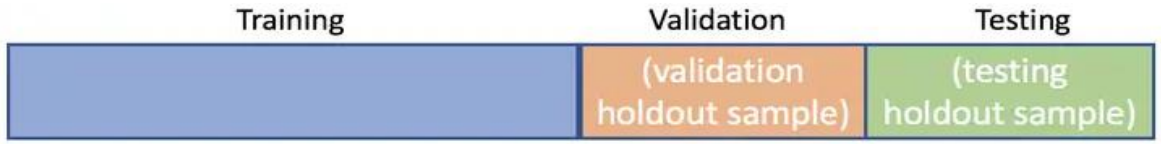
- **Application:** predicting whether a customer will subscribe to a marketing campain

A - Find the target variable

B – What are the features?

C – Name a categorical feature.

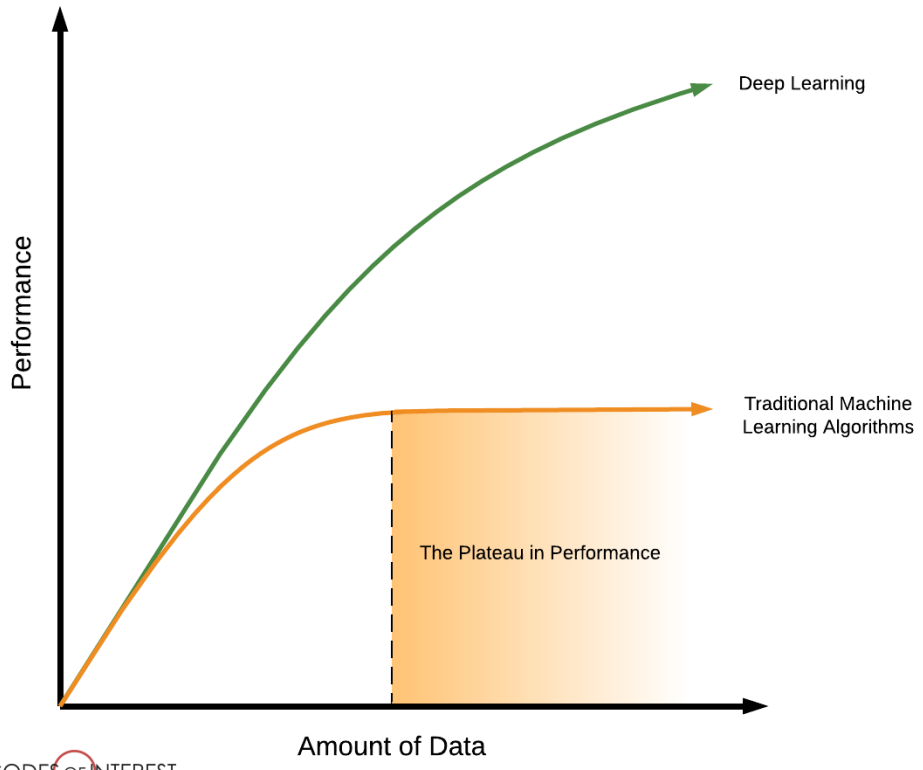
Splitting data – Test, Train and validation data

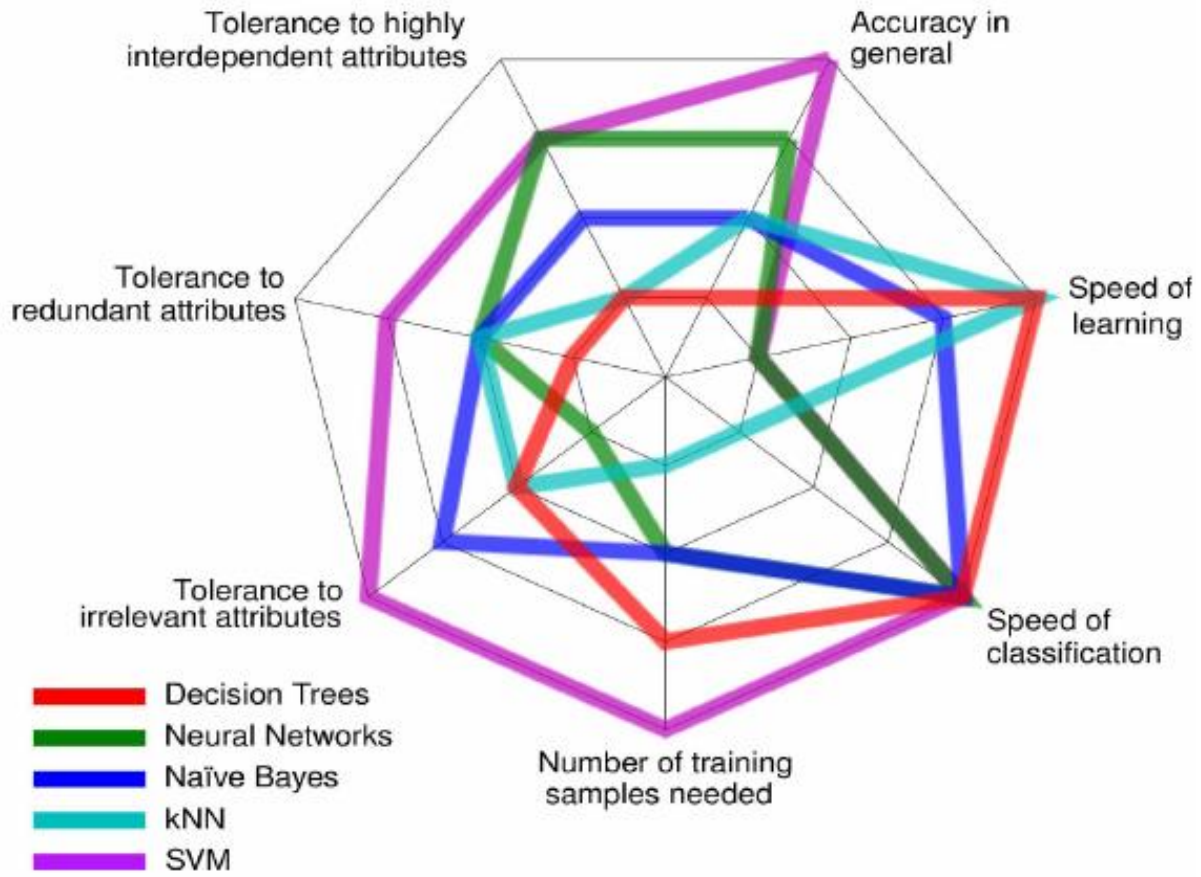


- ideally
  - Data which we use to design our models (Training set)
  - Data which we use to refine our models (Validation set)
  - Data which we use to test our models (Testing set)
- If we do not split our data, we might test our model with the same data that we use to train our model
  - Beware of *train-test contamination*

https://algotrading101.com/learn/train-test-split/

Model Performance – No magic bullet





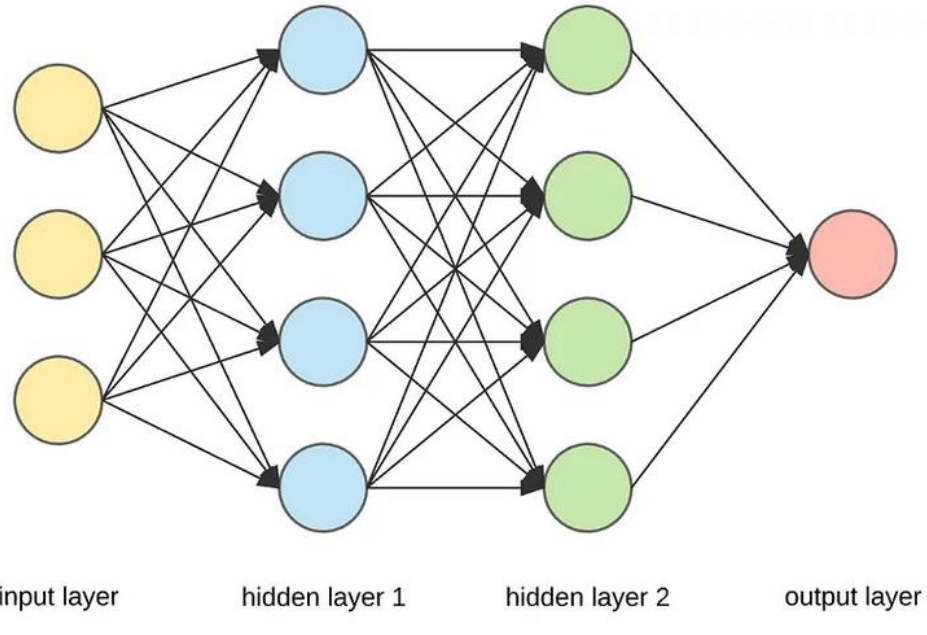
Widanapathirana et al. 2012

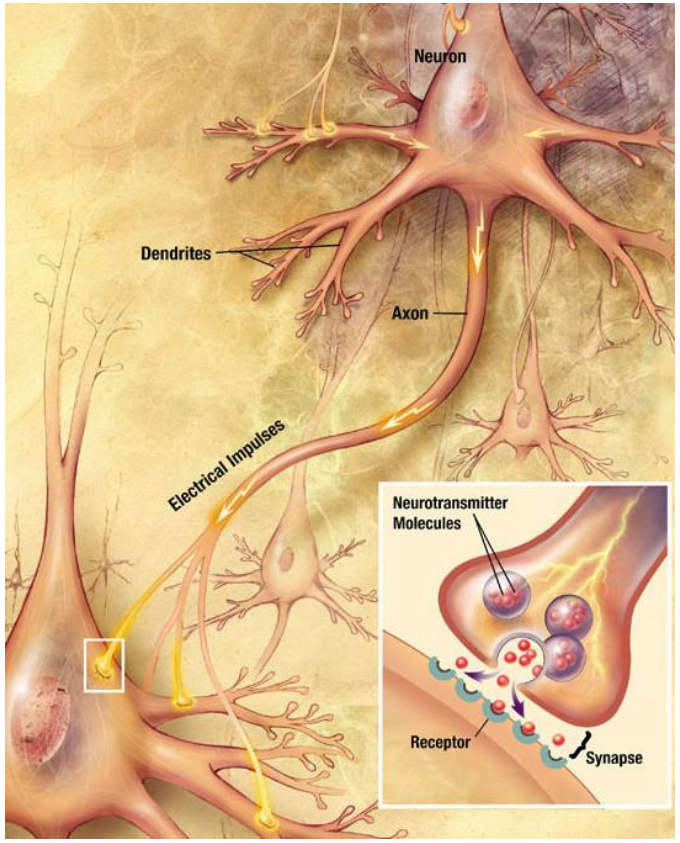
CODES OF INTEREST

SINTEF

Choosing an architecture and tuning hyperparameters

A simple building block



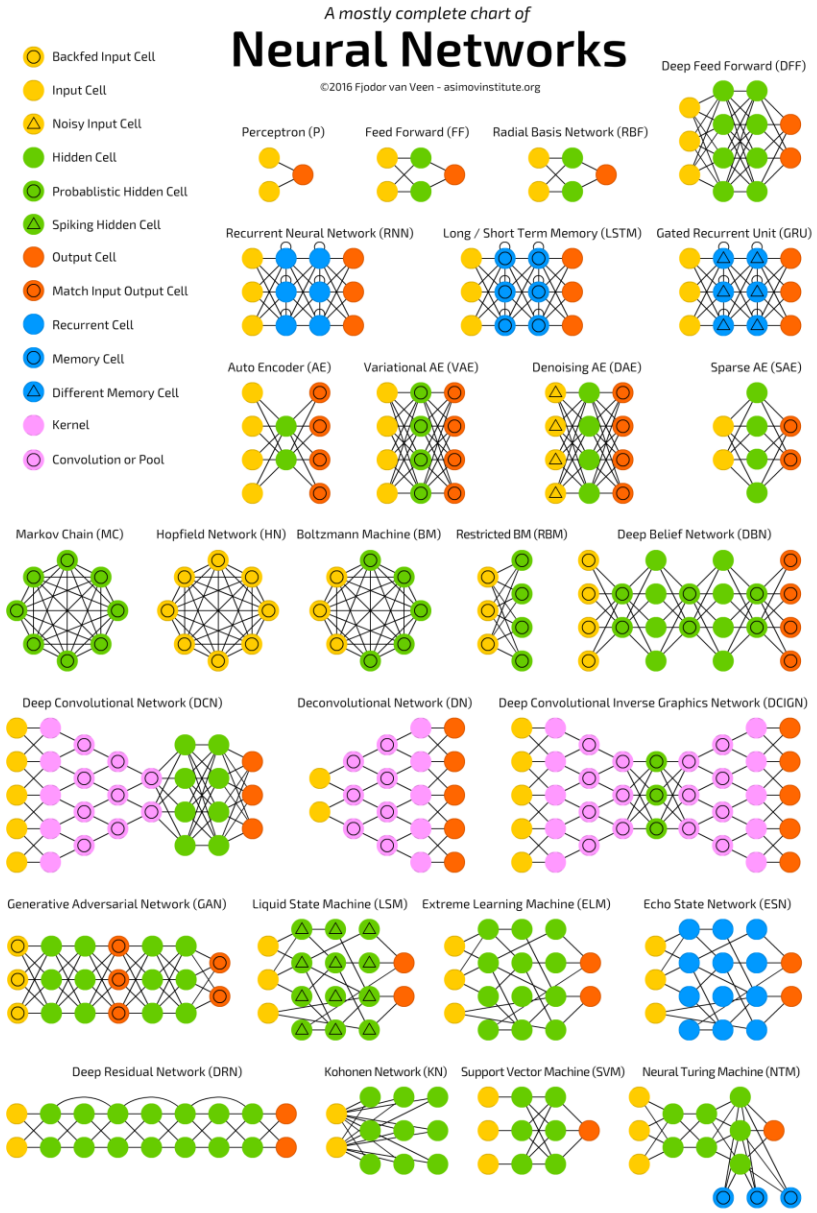


SINTEF

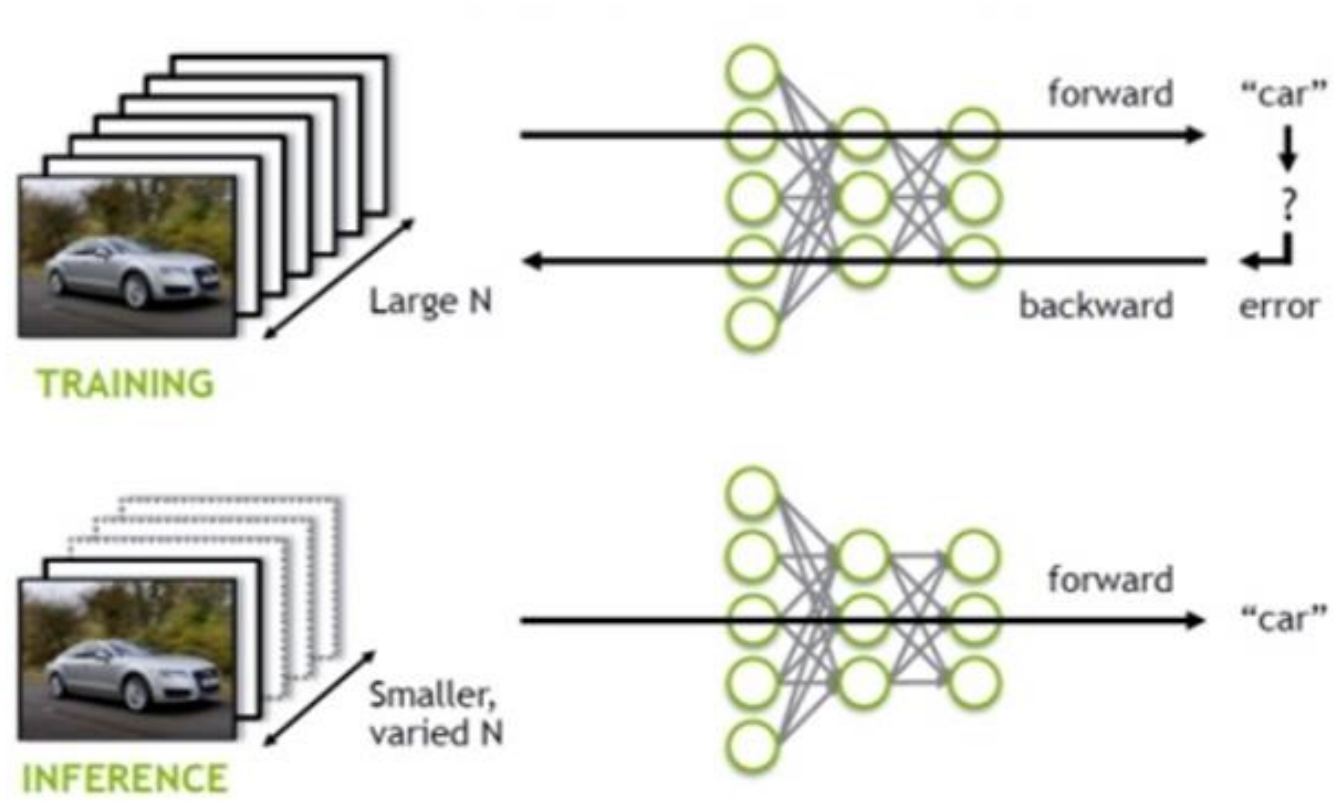
Choosing an architecture and tuning hyperparameters

For larger systems, we need to decide:

- How many layers, how many unit?
- Activation function
- Global architecture
  - Convolutional neural nets (CNN) - Good choice for spatial data
  - Recurent neural nets (RNN) - Good choice for temporal data
- How to connect the layers
- How to initialize weights
- Learning Rate
- Regularization Strength and technique



Training a model VS using a model



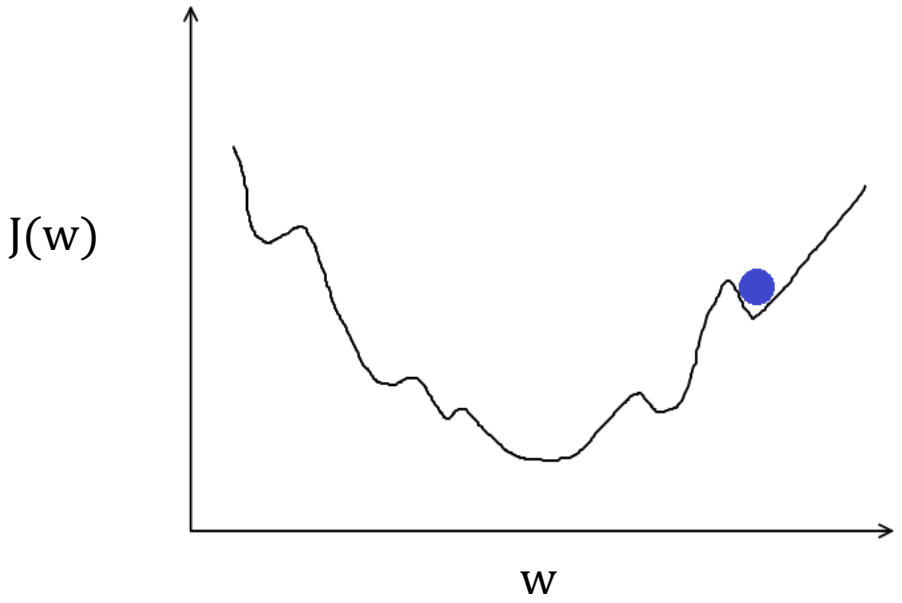
https://www.exxactcorp.com/blog/HPC/discover-the-difference-between-deep-learning-training-and-inference

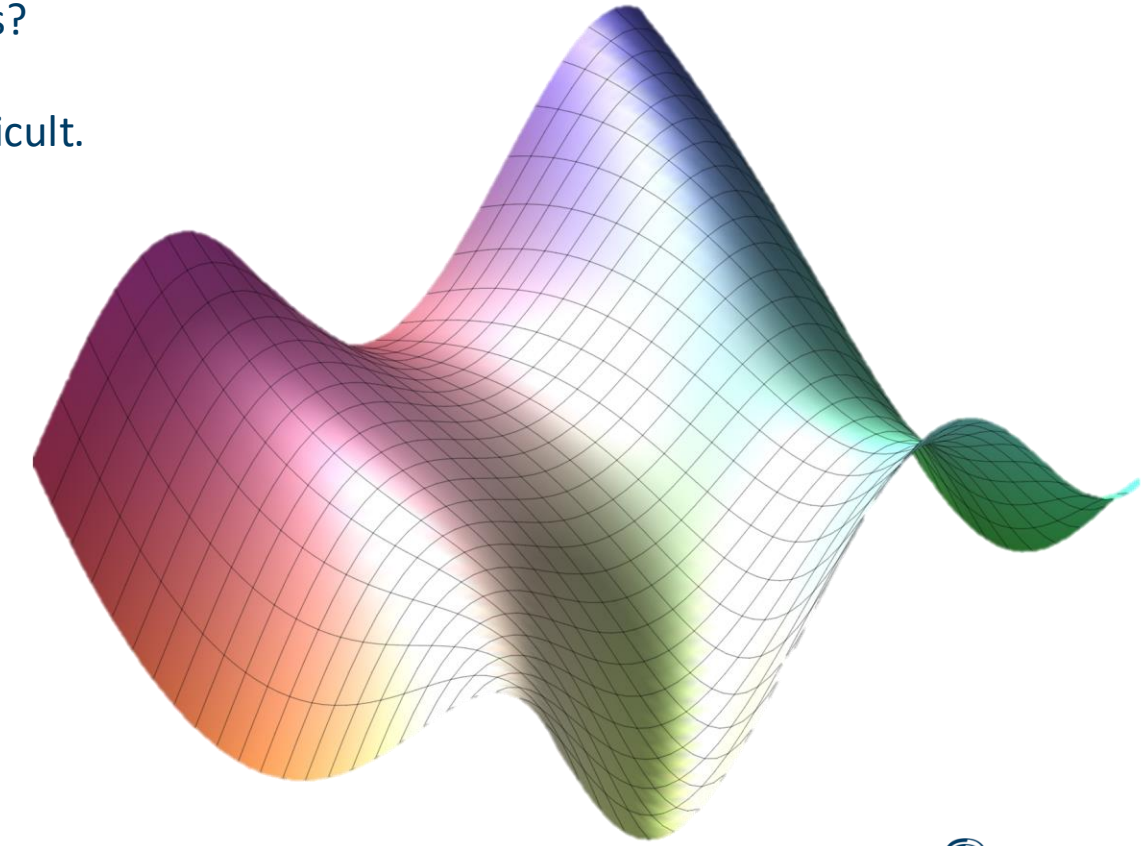
Loss/cost/value function

The loss function measures:

- how far is the model from the data/desired properties?

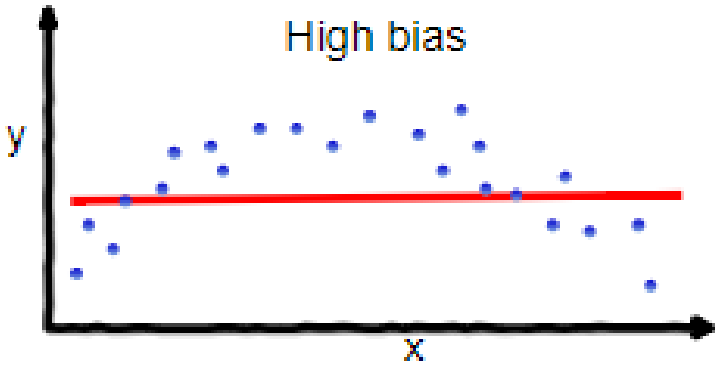
Finding the parameters leading to minimal loss can be difficult.



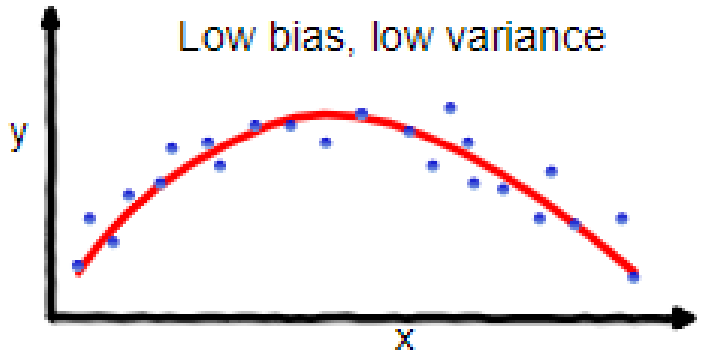


SINTEF

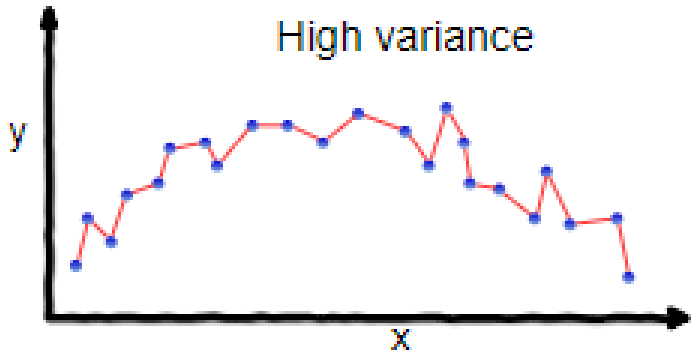
Bias-variance tradeoff



Under fitting



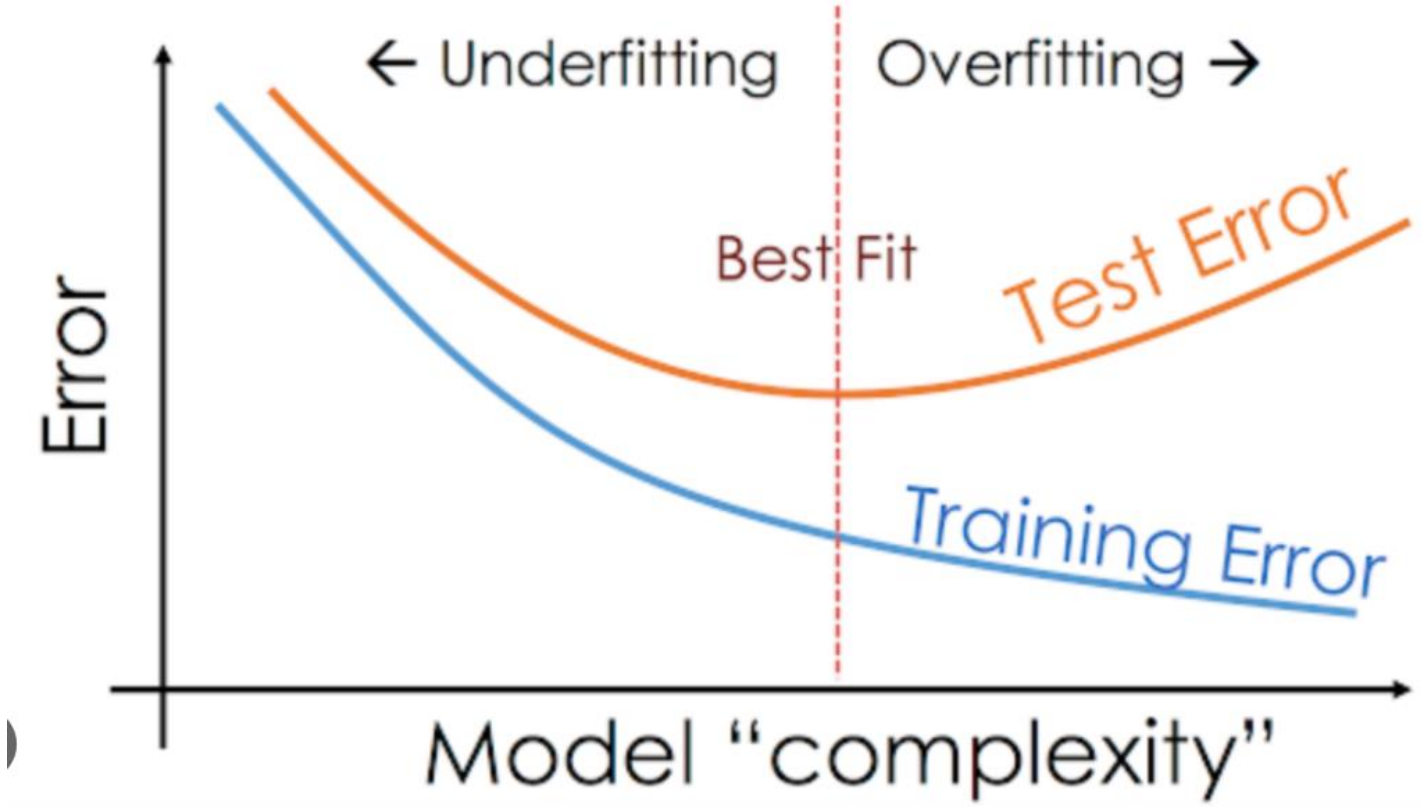
Good balance



Over fitting

https://en.wikipedia.org/wiki/Bias%E2%80%93variance_tradeoff

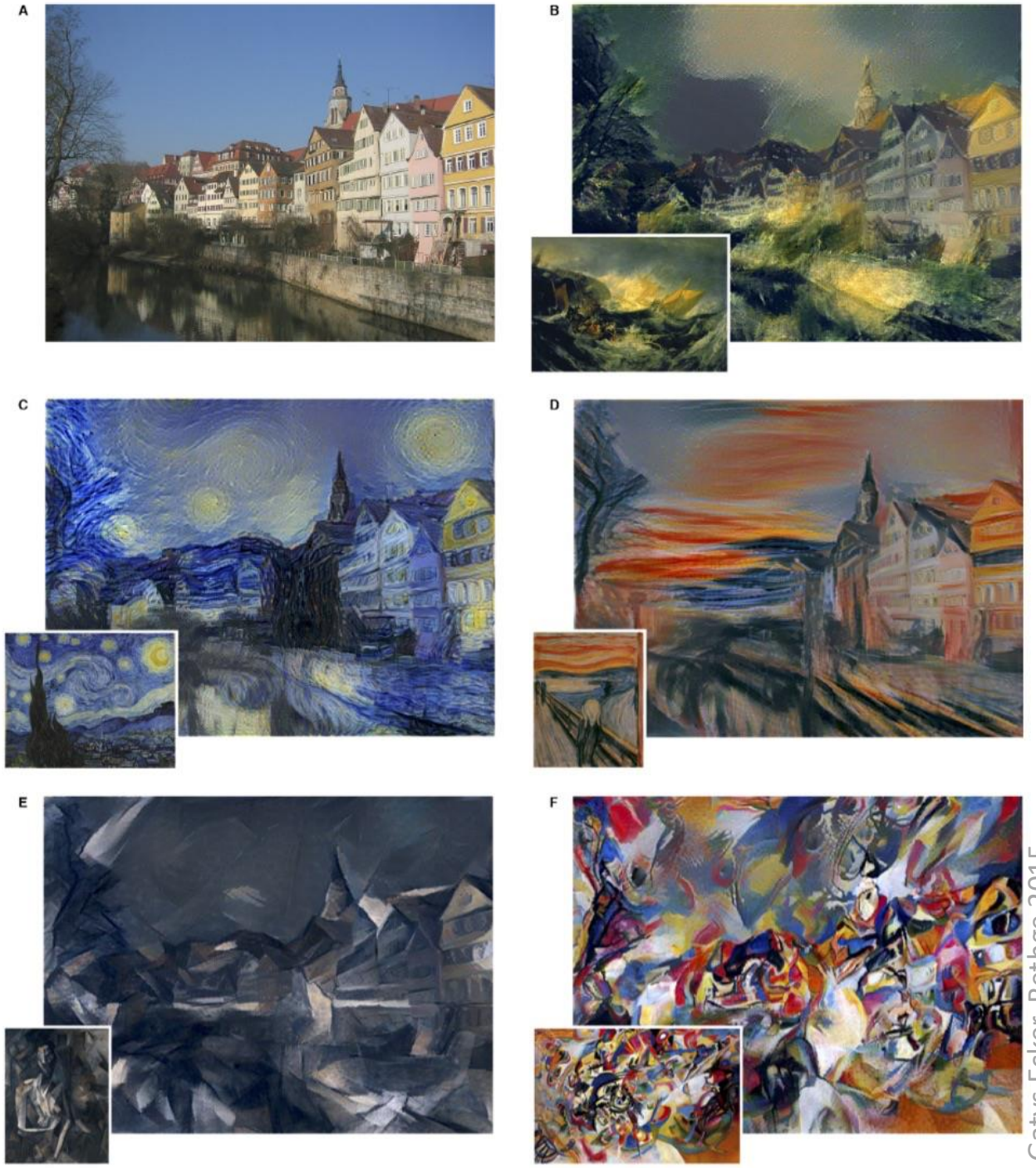
Bias-variance tradeoff



https://vitalflux.com/overfitting-underfitting-concepts-interview-questions/

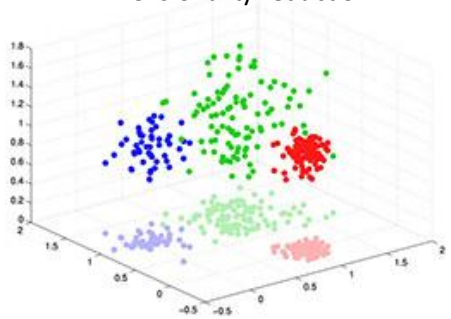
Take home messages

- Each context is different - many moving parts
- Experimentation is key
- Finding a method that works on your dataset is a science
  - (or perhaps it is an art)

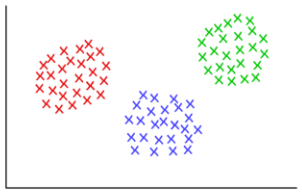


Gatys Ecker, Bethge 2015

Dimensionality reduction



Clustering



Machine Learning

Unsupervised Learning

- Dimensionality Reduction
  - Meaningful Compression
  - Structure Discovery
  - Big data alistaion
  - Feature Elicitation
- Clustering
  - Recommender Systems
  - Targetted Marketing
  - Customer Segmentation

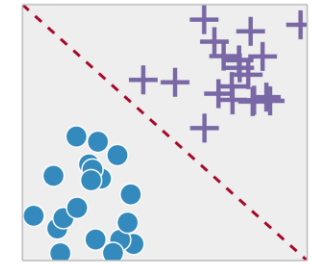
Supervised Learning

- Classification
  - Image Classification
  - Customer Retention
  - Idenity Fraud Detection
  - Diagnostics
- Regression
  - Advertising Popularity Prediction
  - Weather Forecasting
  - Population Growth Prediction
  - Market Forecasting
  - Estimating life expectancy

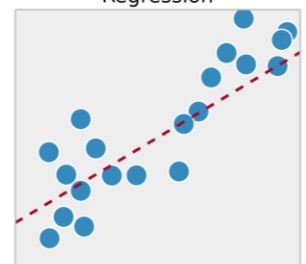
Reinforcement Learning

- Real-time decisions
- Game AI
- Robot Navigation
- Skill Acquisition
- Learning Tasks

Classification



Regression



towardsdatascience.com

SINTEF

Technology for a better society