Machine learning overview

14/06/2021 - Thibault Pelletier

Overview

- Definition
- Learning from Data: Problem Setup
- Features
- Type of machine learning (non exhaustive)
- Classification / Regression
- Supervised / Unsupervised
- Classical / Deep learning
- Deep learning networks
- Metrics
- Generalisation and Bias
- Examples
- How to get started?

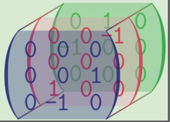
Definition

- Wikipedia :
  Automatically make predictions about new data based on information distilled from “training experiences”.

Learning from Data: Problem Setup

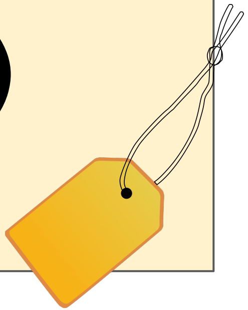
Data (x)

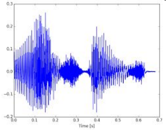
- 1/2/3D-Signals
- Sequences
- Irregular (e.g. graphs)



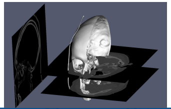
Labels (y)

- Categorical / Continuous



1D 

2D 

3D 

Supervised

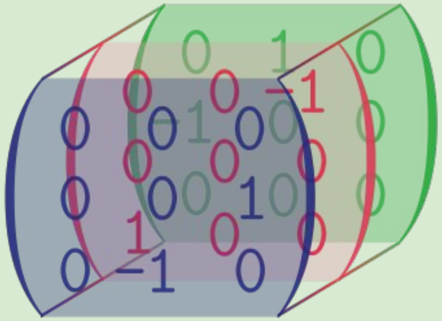
- Learn: **predict(x) = ŷ**
- Optimize: **loss(y, ŷ)**

Word
Species
Age
Tumor 0/1

Kitware

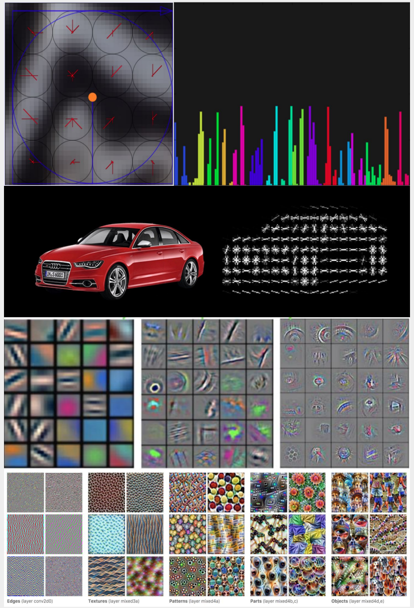
Features

Raw Data



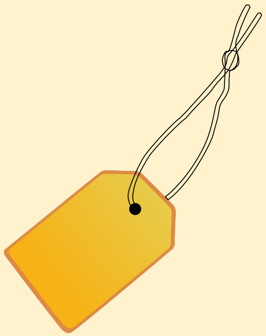
- Raw **data** is often difficult to interpret / separate.
- Transform data into a "**feature vector**".
    - Sometimes hand-crafted
    - Often learned

Feature Representations



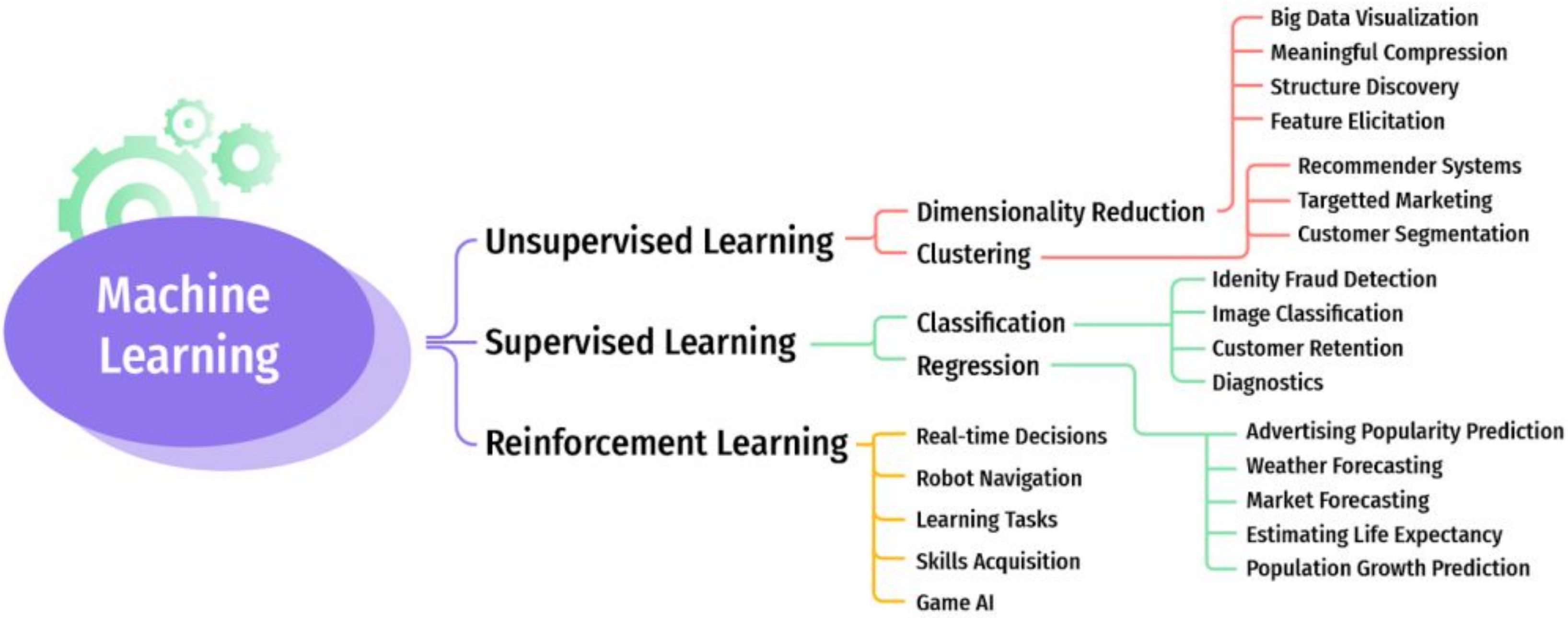
A good "**feature representation**" is the key to good results.

Labels



- Transform **features** into **labels** to perform task:
    - Classification
    - Matching
    - Regression
    - Clustering

Type of machine learning (non exhaustive)

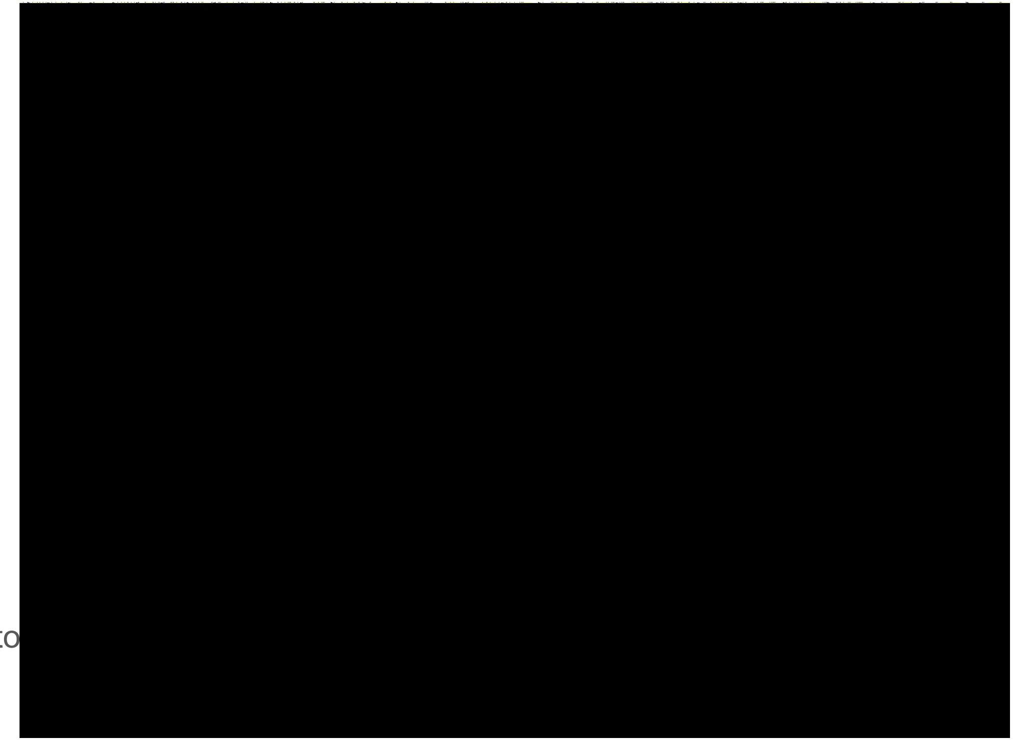


Supervised Classification

- Labels are known

- Task is to find a decision boundary between data points
    - Often this is a hyperplane in a feature space (which in 2D is a line)

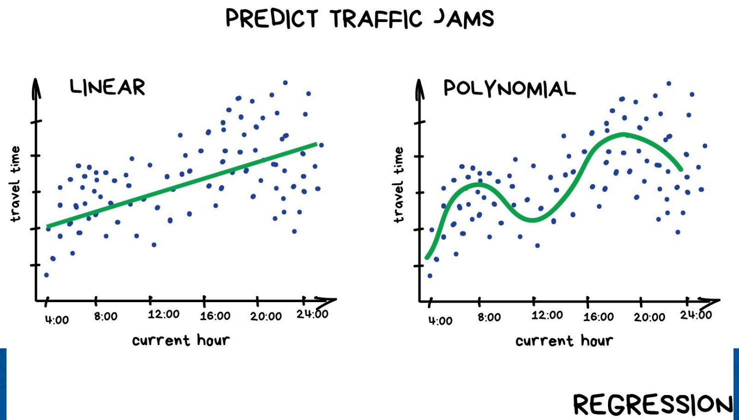
- Often formulated as binary classification.
    - Some algorithms support multiclass directly
    - Multiple binary classifiers can be converted to multi-class classifiers
        - One-vs-Rest O(N)
        - One-vs-One O(N^2)

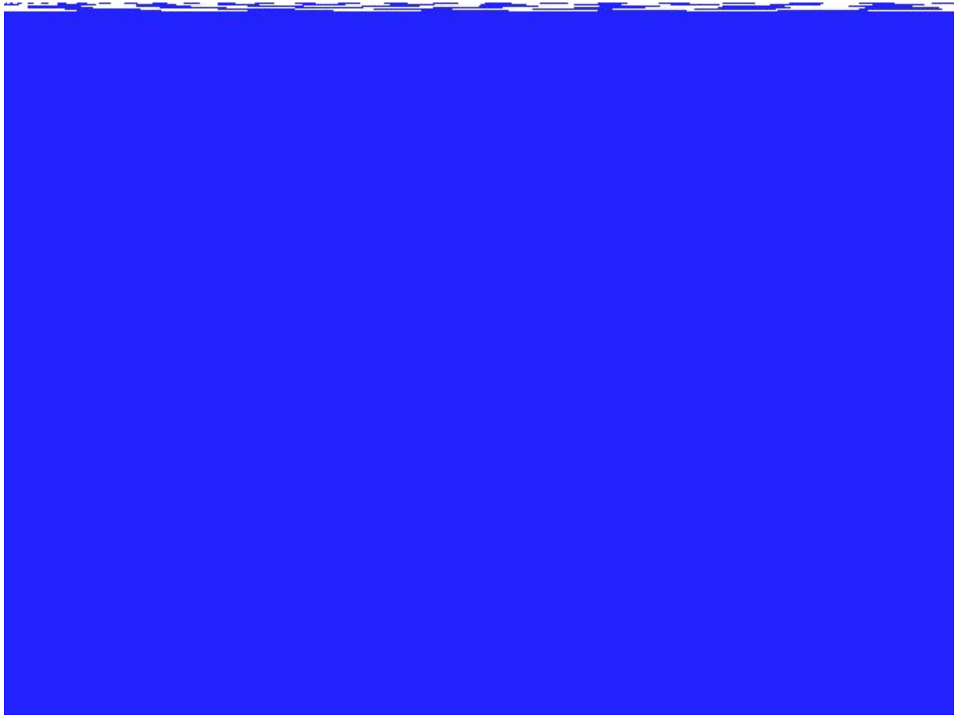


Visualization of the Perceptron Algorithm

Supervised Regression

- Labels are known
- Task is to fit a curve to predict a (usually continuous) quantity.
- Generally used when interpolation between values is important.





Visualization of the Linear Regression

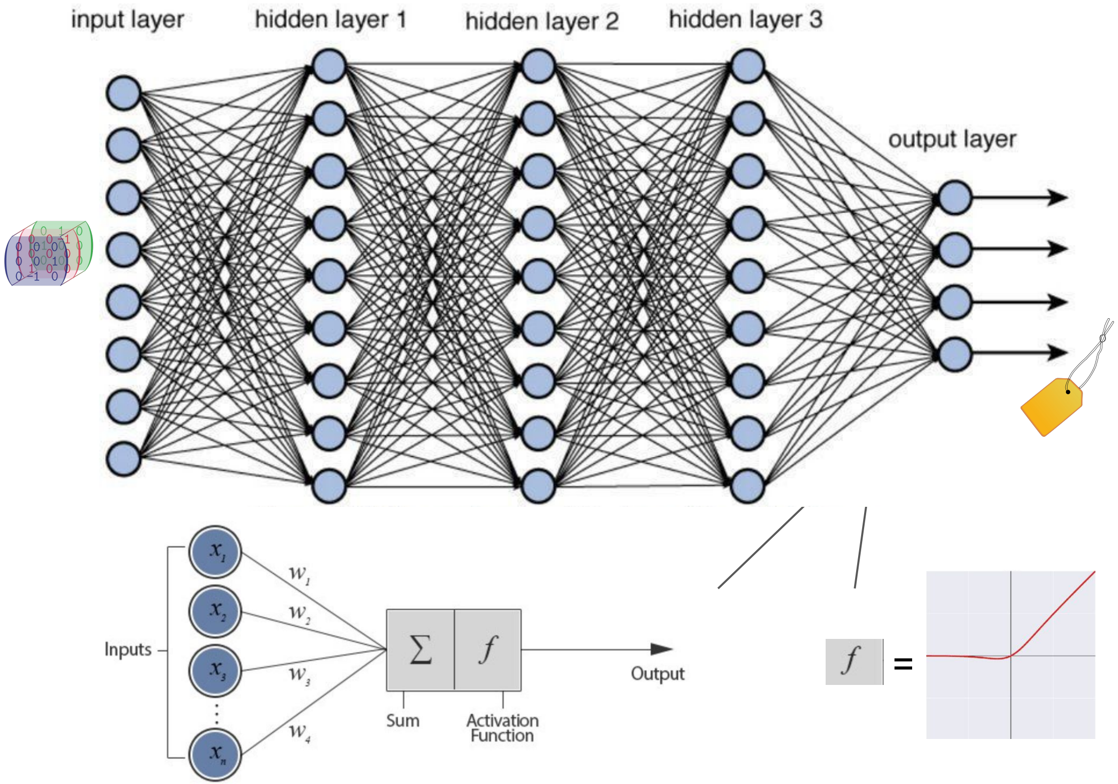
Supervised / Unsupervised

- Supervised learning
  - Use input data along “labeled” data
  - Allows you to collect or produce a data output given the information it has learned
  - Usually simple to put in place but harder to collect data
  - Examples include regression and classification
- Unsupervised learning
  - Uses input data only
  - Helps you to finds all kind of unknown patterns in data
  - Usually simple to acquire data but harder to exploit
  - Examples include clustering and association

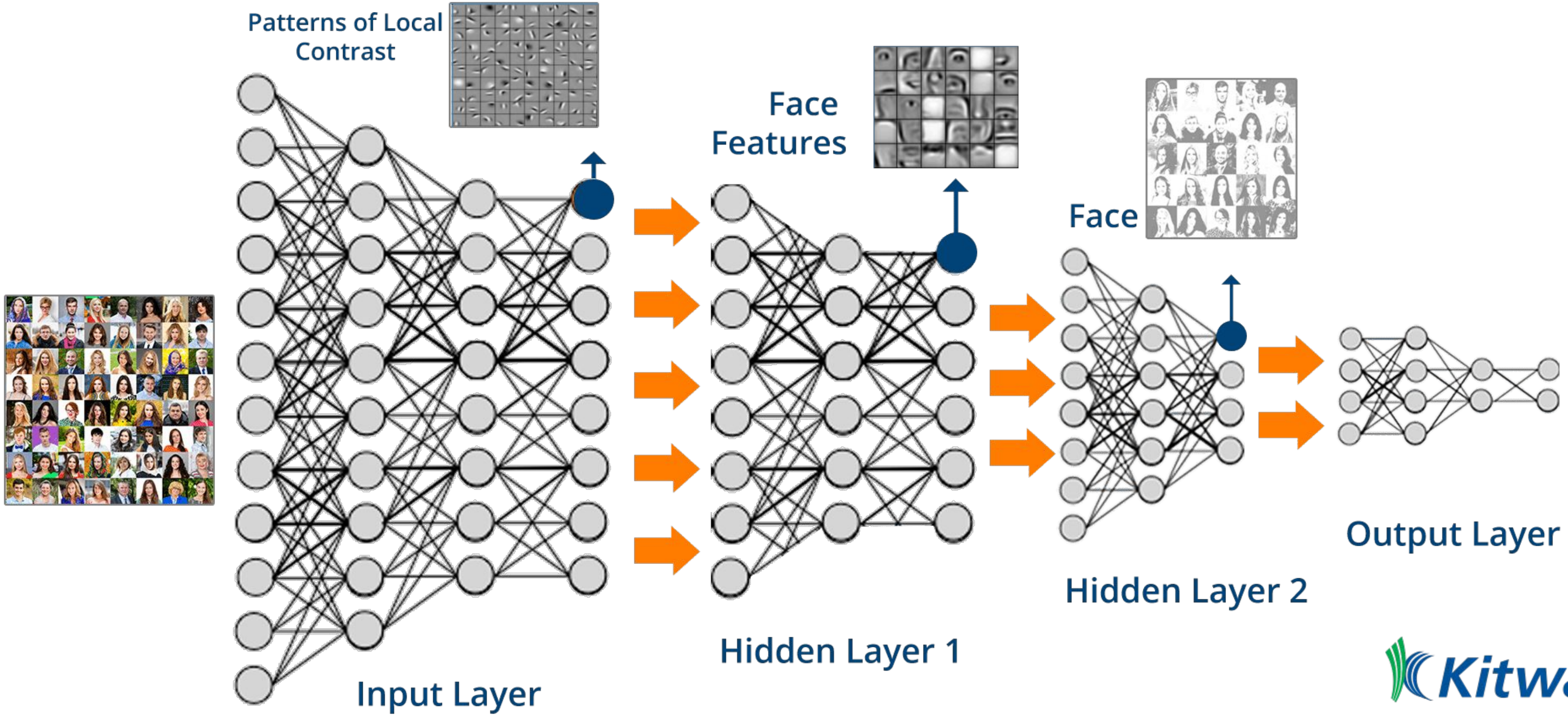
Deep learning / Classical machine learning

- Deep learning
  - Best-in-class performance
  - Scales effectively with data
  - No need for feature engineering
  - Adaptable and transferable
- Classical machine learning
  - Works better on small data
  - Financially and computationally cheap
  - Easier to interpret

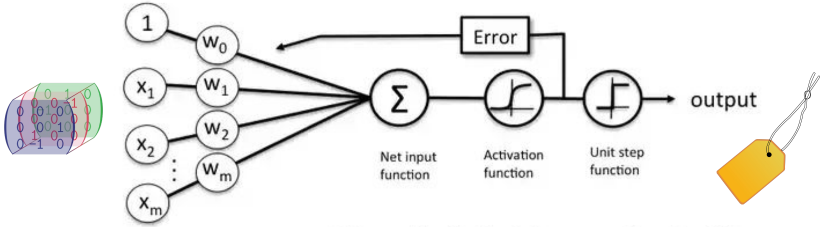
Anatomy of a Deep Network



Deep Networks Learn Good Feature Representations

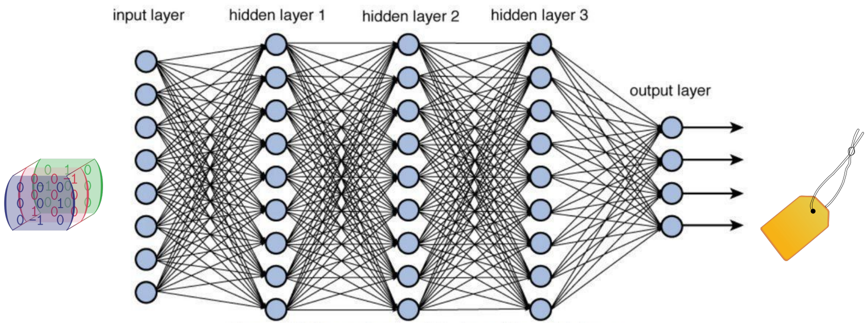


Deep network vs traditional network



Traditional Logistic Regression

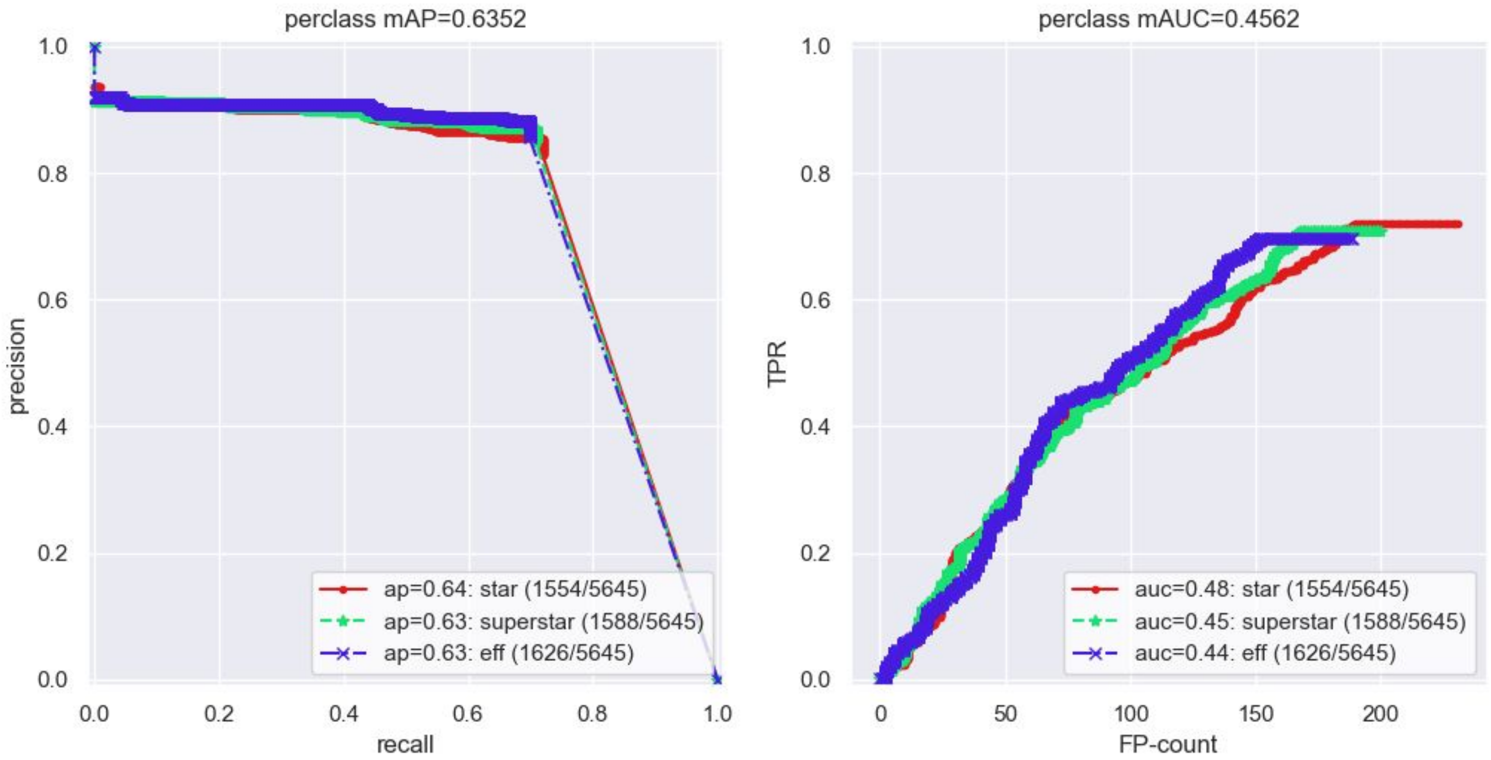
(Requires good features, learns from few examples)



Deep Network

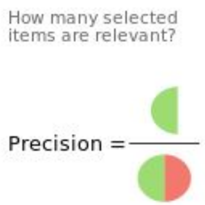
(Learns good features, requires many examples)

Metrics : Precision and Recall (PR) and ROC Curves

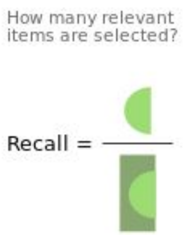


Metrics : Binary Classification Confusion

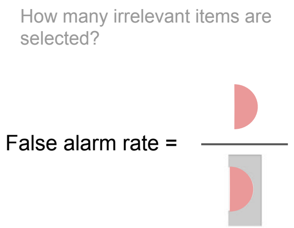
		True condition				
	Total population	Condition positive	Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) $= \frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
Predicted condition	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$	F_1 score $= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$		



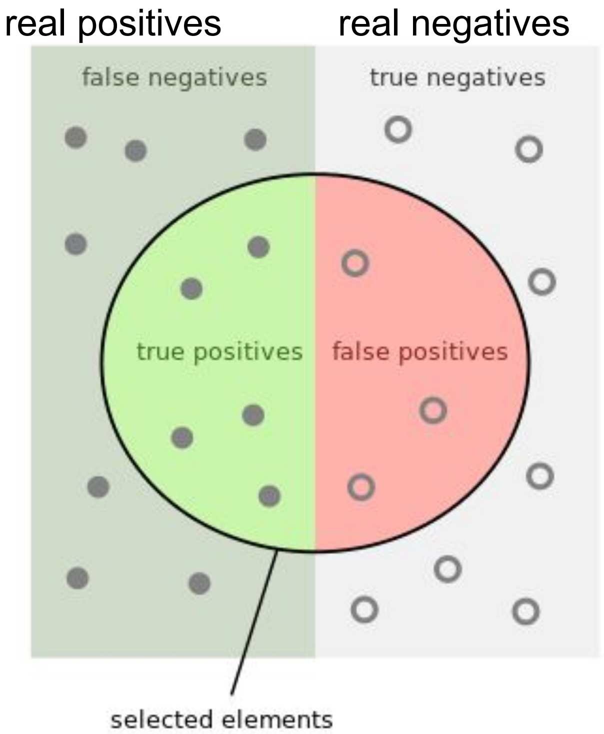
PPV



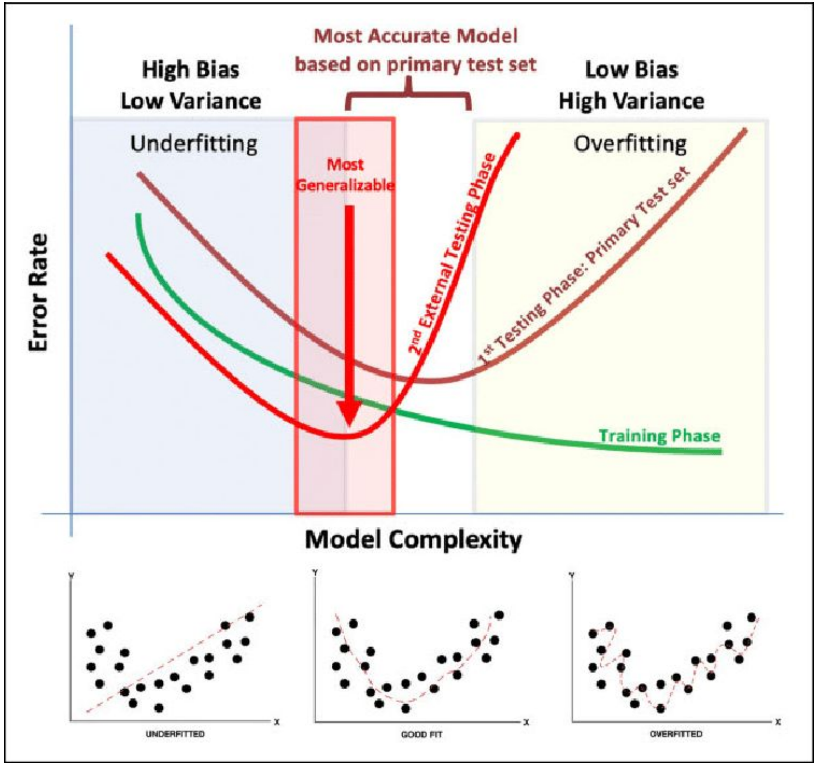
TPR



FPR

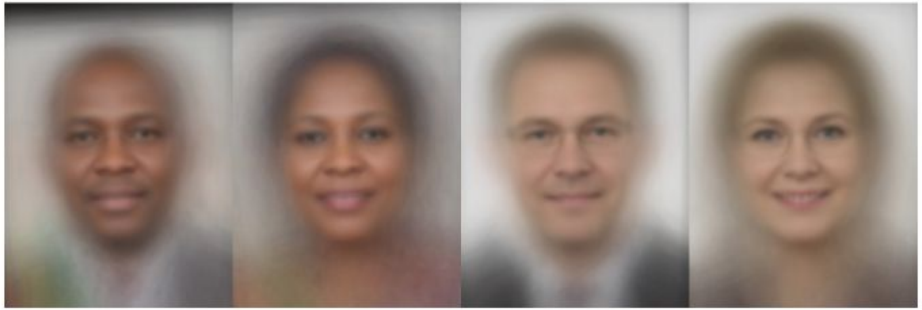


Evaluation : Bias / Variance tradeoff



Data bias

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%



Kitware

https://broutonlab.com/blog/ai-bias-solved-with-synthetic-data-generation

Example of classical machine learning in cancer detection

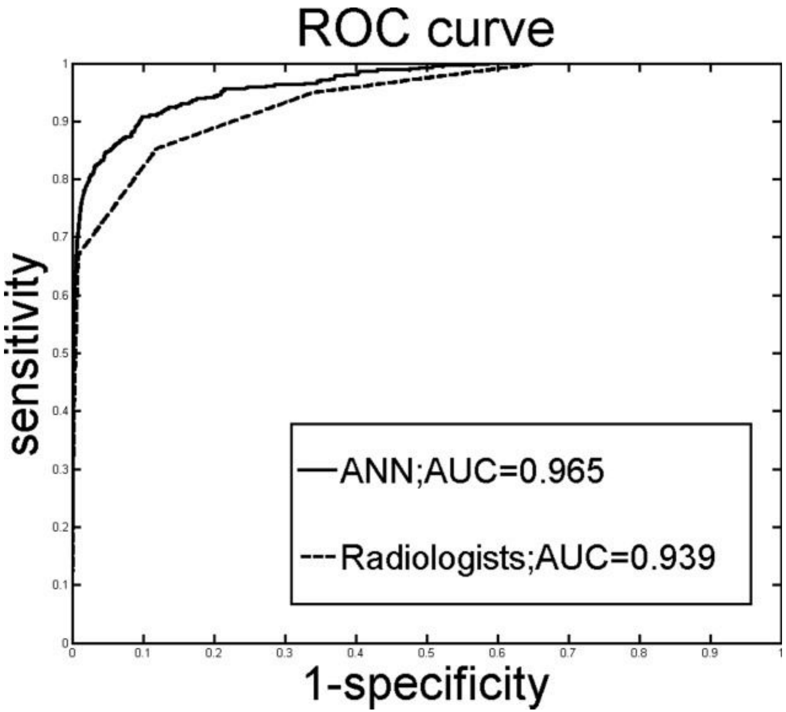
Publication	Method	Cancer type	No of patients	Type of data	Accuracy	Validation method	Important features
Ayer T et al. [19]	ANN	Breast cancer	62,219	Mammographic, demographic	AUC = 0.965	10-fold cross validation	Age, mammography findings
Waddell M et al. [44]	SVM	Multiple myeloma	80	SNPs	71%	Leave-one-out cross validation	snp739514, snp521522, snp994532
Listgarten J et al. [45]	SVM	Breast cancer	174	SNPs	69%	20-fold cross validation	snpCY11B2 (+) 4536 T/C snpCYP1B1 (+) 4328 C/G
Stajadinovic et al. [46]	BN	Colon carcinomatosis	53	Clinical, pathologic	AUC = 0.71	Cross-validatio n	Primary tumor histology, nodal staging, extent of peritoneal cancer

Example : Breast cancer detection

- 36 discrete features
  - Age group, hormone therapy, personal history of breast cancer, family history of breast cancer, breast density, etc.

- Dedicated model
  - 3 layer feed-forward neural network
  - 36 inputs (features), 1000 hidden layer, 1 output (breast cancer probability)

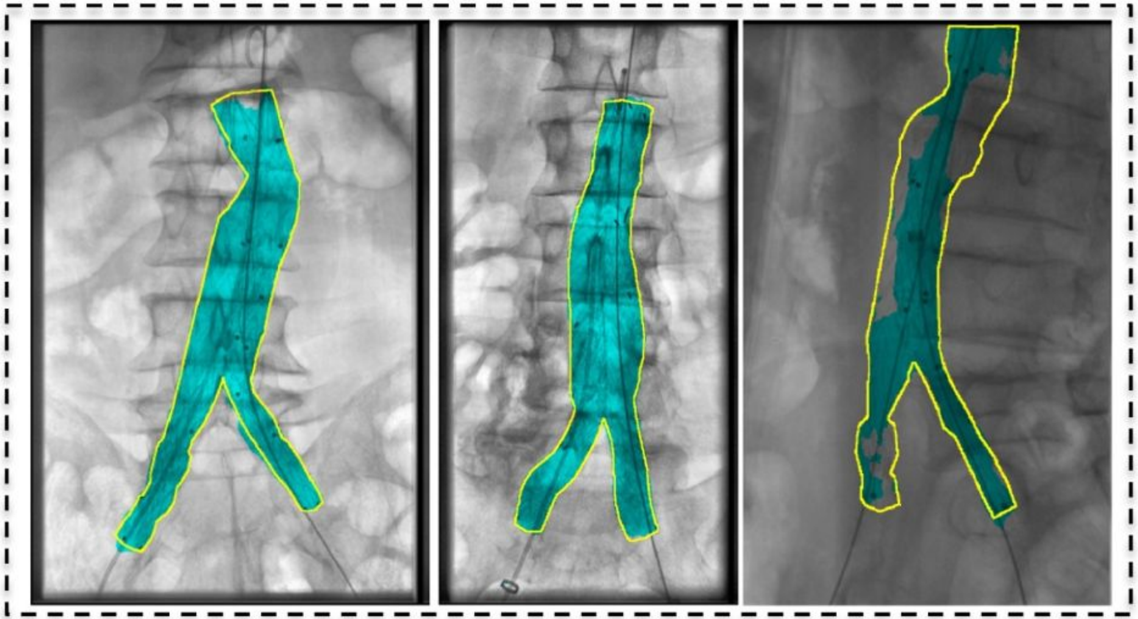
- Radiologist level of results



https://acsjournals.onlinelibrary.wiley.com/doi/full/10.1002/cncr.25081

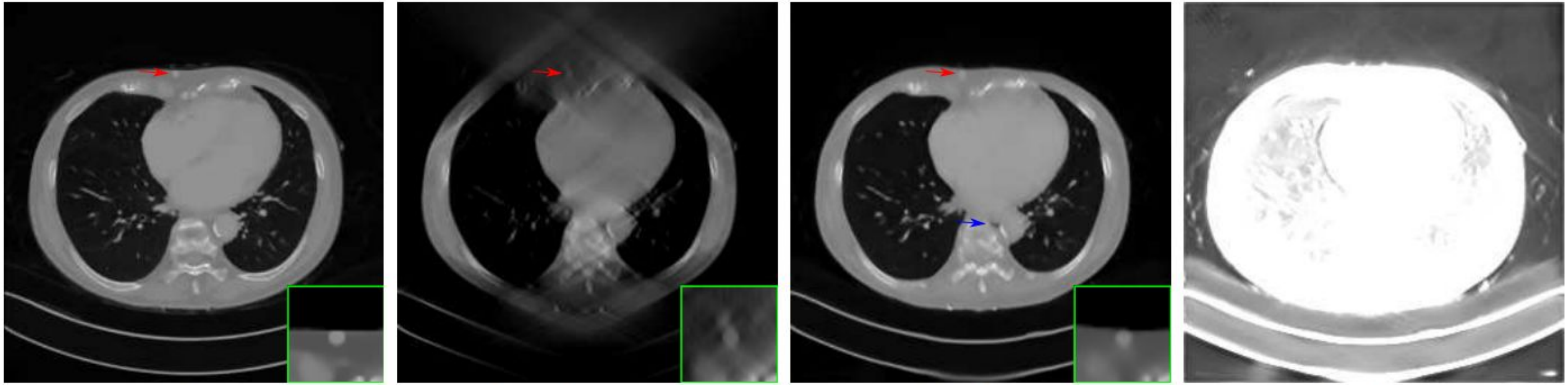
Deep learning algorithm example

- UNet based aortic stents in single uncontrasted X-ray image



Deep learning algorithm example

- Investigations on Robustness of Deep Learning in Limited Angle Tomography

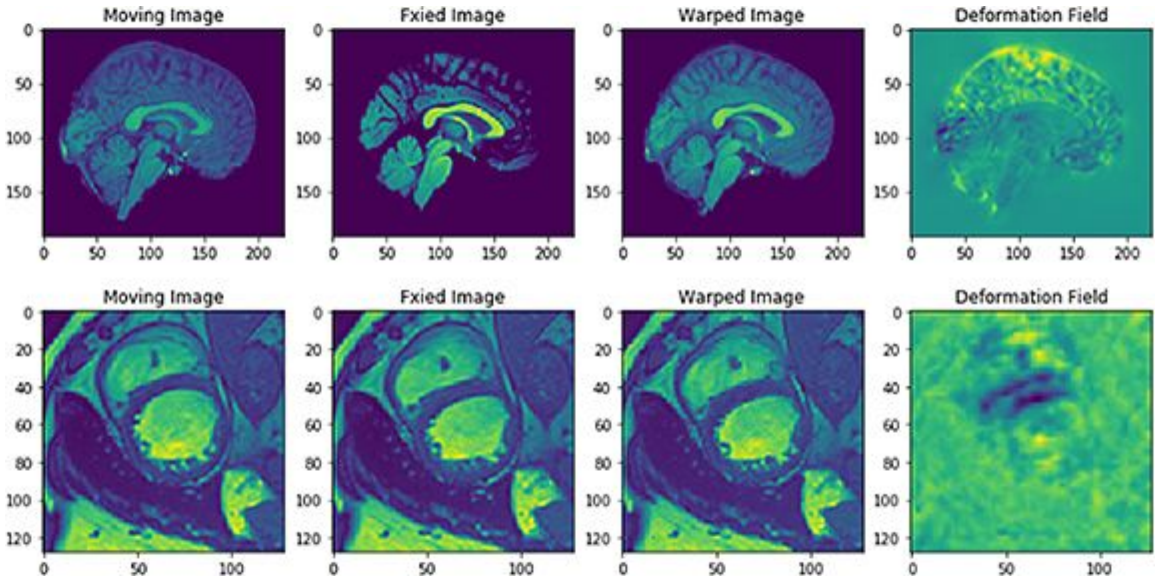


(a) $\boldsymbol{f}_{\mathrm{ref,L}}$ (b) $\boldsymbol{f}_{\mathrm{limited,L}}$ (c) $\boldsymbol{f}_{\mathrm{est,L}}$ (d) $\boldsymbol{f}_{\mathrm{est,L}}$

https://www5.informatik.uni-erlangen.de/Forschung/Publikationen/2018/Huang18-SIO.pdf

Deep learning algorithm example

- Example of brain and cardiac MRI image registration with Voxelmorph



Kitware

https://iopscience.iop.org/article/10.1088/2516-1091/abd37c

How to get started ?

Questions ?