Introduction to Artificial Intelligence

Theory & Practical Applications

P. Pazera

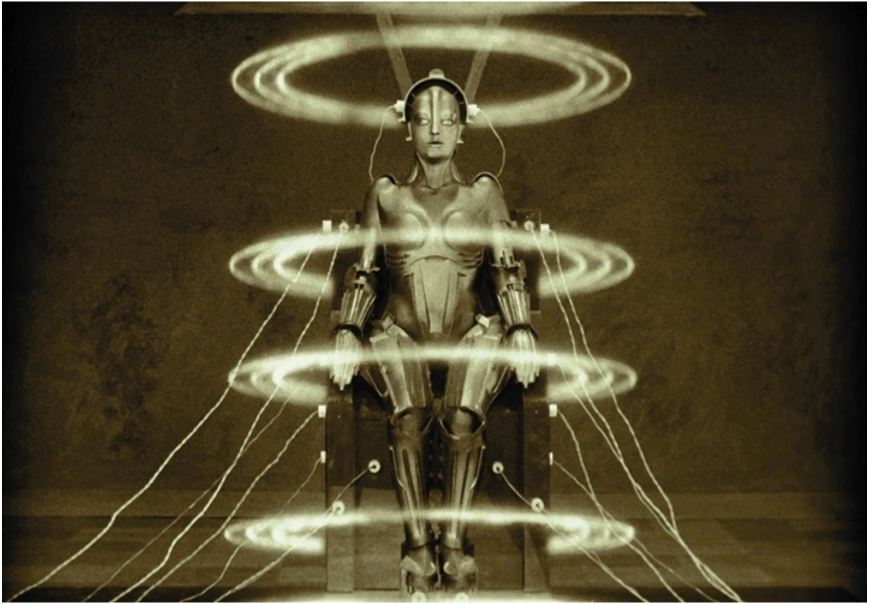
Baltic Institute of Advanced Technology
BPTI

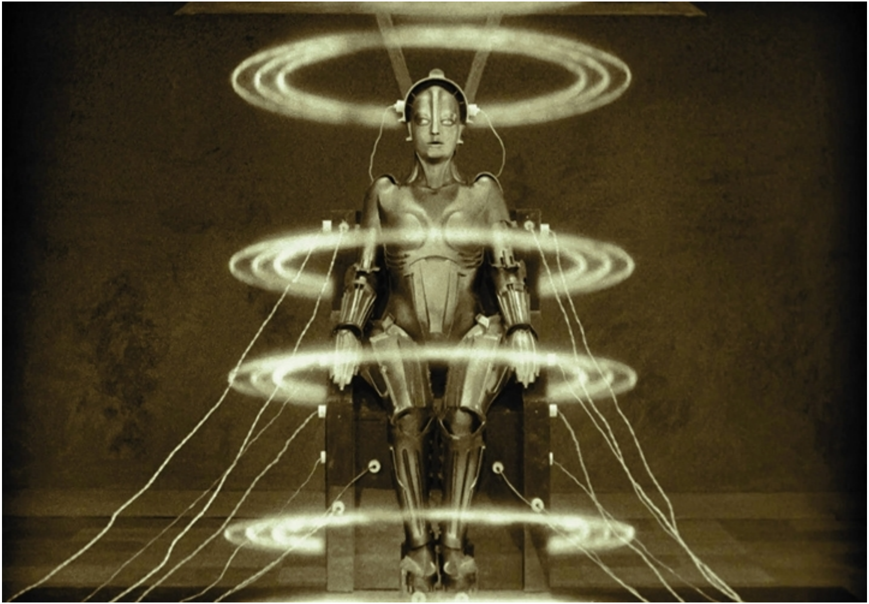
Vilnius, 2019

Outline

Outline

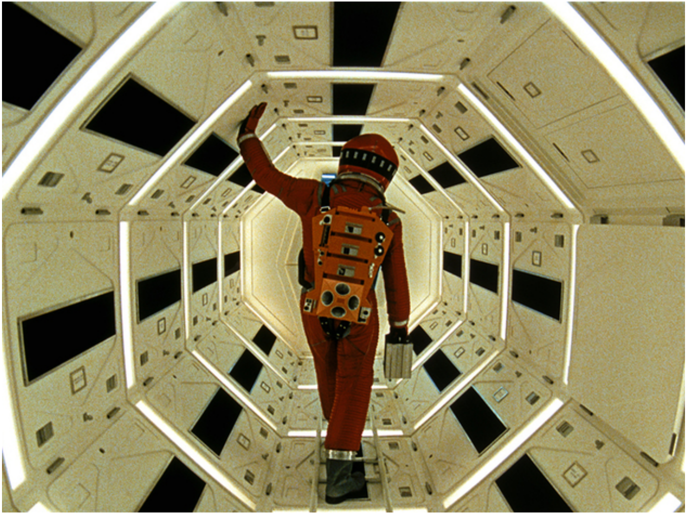
1. Introduction
   - What is artificial intelligence?
   - Development of AI systems
2. Machine Learning
   - What is Machine Learning?
   - Supervised Learning
   - Unsupervised Learning
   - Other ML Paradigms
   - Generalization & Overfitting
3. Deep Learning & Neural Networks
   - What are artificial neural networks (ANNs)?
   - What are deep neural networks (DNNs)?
4. AI Struggles & Further Discussion

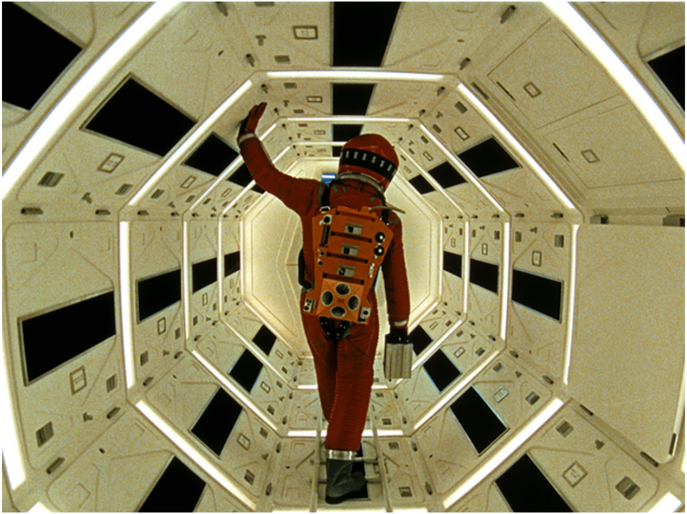




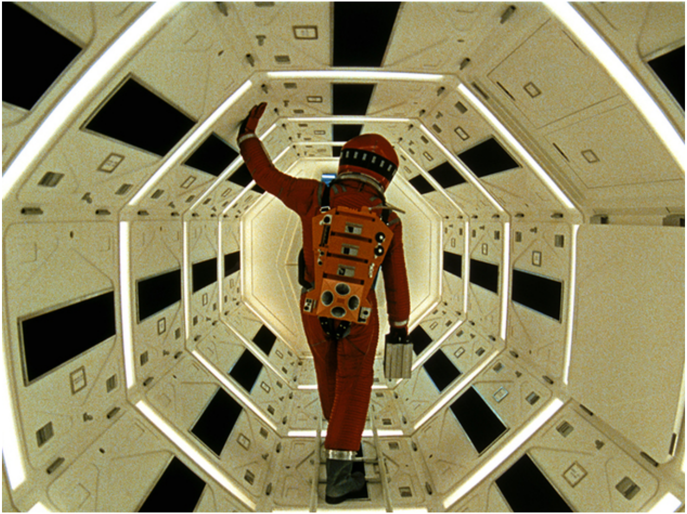
Metropolis (Fritz Lang, 1927)

The Matrix (Lana Wachovski, 1999)





2001: A Space Odyssey (Stanley Kubrick, 1968) *Video*.



2001: A Space Odyssey (Stanley Kubrick, 1968) *Video*.
For a further review of 11 popular AI science fiction movies and their scientific accuracy, see this *link*.

What is artificial intelligence?

- "*The study of the modelling of human mental functions by computer programs.*" (Collins Dictionary)

What is artificial intelligence?

- "*The study of the modelling of human mental functions by computer programs.*" (Collins Dictionary)
- "*Artificial intelligence is a type of computer technology which is concerned with making machines work in an intelligent way, similar to the way that the human mind works.*" (Collins Dictionary)

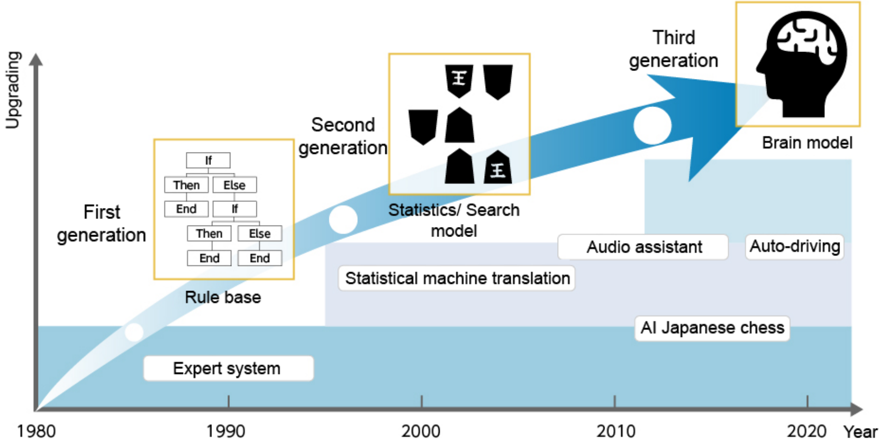
What is artificial intelligence?

- "*The study of the modelling of human mental functions by computer programs.*" (Collins Dictionary)
- "*Artificial intelligence is a type of computer technology which is concerned with making machines work in an intelligent way, similar to the way that the human mind works.*" (Collins Dictionary)
- "**[...] the study of agents that receive percepts from the environment and perform actions.**" *Artificial Intelligence: A Modern Approach (Russel & Norvig)*

Outline

1. Introduction
   - What is artificial intelligence?
   - Development of AI systems
2. Machine Learning
   - What is Machine Learning?
   - Supervised Learning
   - Unsupervised Learning
   - Other ML Paradigms
   - Generalization & Overfitting
3. Deep Learning & Neural Networks
   - What are artificial neural networks (ANNs)?
   - What are deep neural networks (DNNs)?
4. AI Struggles & Further Discussion

Development of AI systems



Source: *nttdata*

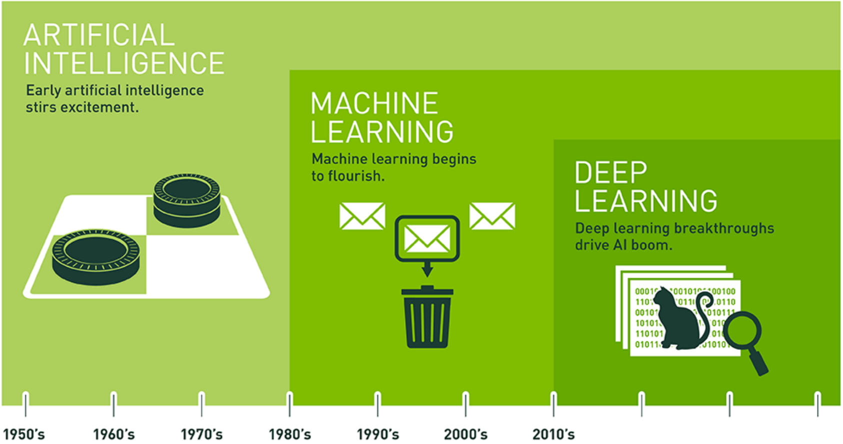
Development of AI systems

		What kind of technology	Realized function
First generation	Rule base	AI that can infer or explore by using "knowledge"	Output based on input rules as knowledge (human beings create rules)
Second generation	Statistics / Search model	AI that incorporates machine learning	Human beings give data and feature quantities as sample, learn rules and knowledge by themselves, automatically judge new input data and output
Third generation	Brain model	AI that incorporated deep learning	Even without human intervention or setting rules, autonomously learn features and rules, automatically judge and output

AI in a broad sense

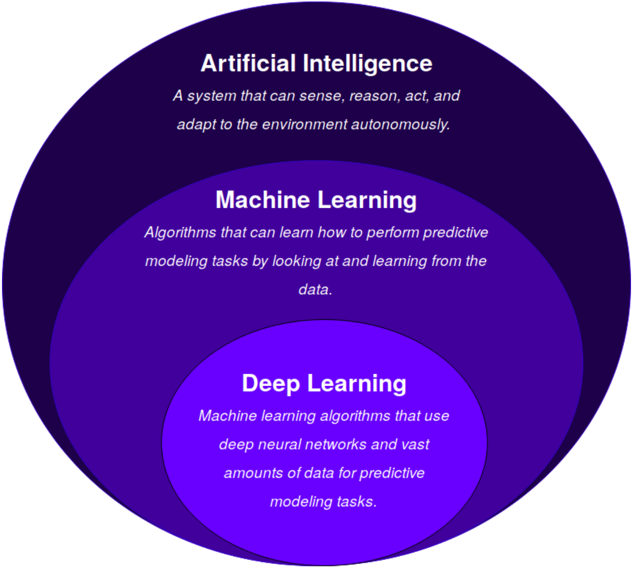
Source: *nttdata*

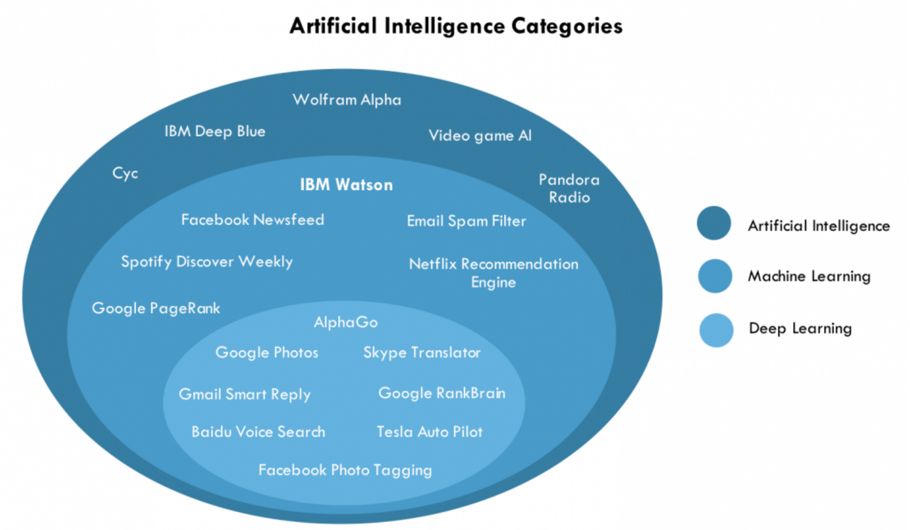
Keywords' timeline



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

AI vs. ML vs. DL


Artificial Intelligence
A system that can sense, reason, act, and adapt to the environment autonomously.
Machine Learning
Algorithms that can learn how to perform predictive modeling tasks by looking at and learning from the data.
Deep Learning
Machine learning algorithms that use deep neural networks and vast amounts of data for predictive modeling tasks.



Source: ARK Investment Management LLC

AI these days...

Autonomous Robot *Spot* (Boston Dynamics) Also see *this* presentation.

AI these days...

Autonomous Robot *Atlas* (Boston Dynamics)

Outline

What is Machine Learning (ML)?

- Algorithms that learn from the data and apply this knowledge to make informed decisions.

What is Machine Learning (ML)?

- Algorithms that learn from the data and apply this knowledge to make informed decisions.
- Numerous applications across multiple industries (e.g., spam filtering, medical diagnostics, virtual assistants, etc).

What is Machine Learning (ML)?

- Algorithms that learn from the data and apply this knowledge to make informed decisions.
- Numerous applications across multiple industries (e.g., spam filtering, medical diagnostics, virtual assistants, etc).
- Effective when the problems are simple and the tasks have clear objectives.

What is Machine Learning (ML)?

- Algorithms that learn from the data and apply this knowledge to make informed decisions.
- Numerous applications across multiple industries (e.g., spam filtering, medical diagnostics, virtual assistants, etc).
- Effective when the problems are simple and the tasks have clear objectives.

Learning of a mapping function that can correctly assign a category to some input data or predict a value related to it.

Machine Learning Paradigms

- **Supervised Learning** - the algorithm uses a labeled data to learn the key features that can be exploited to solve some classification or regression problem related to it (e.g., ham/spam classification).

Machine Learning Paradigms

- **Supervised Learning** - the algorithm uses a labeled data to learn the key features that can be exploited to solve some classification or regression problem related to it (e.g., ham/spam classification).
- **Unsupervised Learning** - the model tries to make sense of the data by extracting and analyzing its features on its own (e.g., clustering).

Machine Learning Paradigms

- **Supervised Learning** - the algorithm uses a labeled data to learn the key features that can be exploited to solve some classification or regression problem related to it (e.g., ham/spam classification).
- **Unsupervised Learning** - the model tries to make sense of the data by extracting and analyzing its features on its own (e.g., clustering).
- **Semi-supervised Learning** - uses a small amount of labeled data and a larger set of unlabeled data (e.g., sentiment classification with language models).

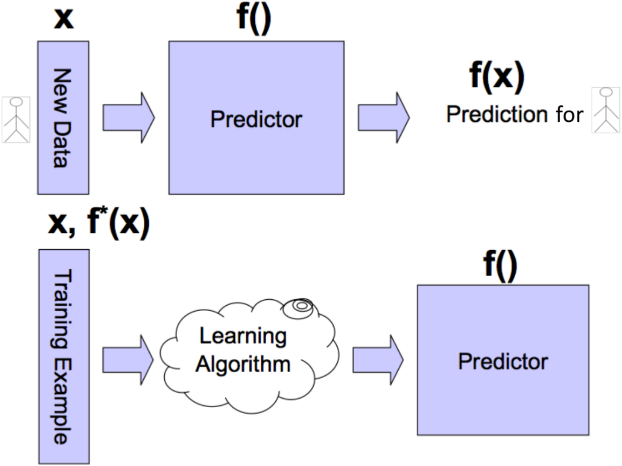
Machine Learning Paradigms

- **Supervised Learning** - the algorithm uses a labeled data to learn the key features that can be exploited to solve some classification or regression problem related to it (e.g., ham/spam classification).
- **Unsupervised Learning** - the model tries to make sense of the data by extracting and analyzing its features on its own (e.g., clustering).
- **Semi-supervised Learning** - uses a small amount of labeled data and a larger set of unlabeled data (e.g., sentiment classification with language models).
- **Reinforcement Learning** - an algorithm with a reward system that provides a feedback when it performs best in a particular situation (often used in locomotion, robotics).

Outline

Supervised Learning

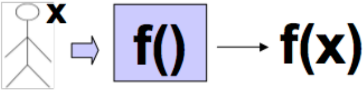
Learning from examples



Source: Victor Lavrenko, 2014 (UoE)

Learning from examples

Representing Data



- How do we represent  mathematically?
- Depends on what we're trying to do:
  - deciding to loan money?
  - predicting gender?
- Represent  as a set of attribute-value pairs
  - example: **x** = {height=*180cm*, eyes="*blue*", job="*student*"}

Source: Victor Lavrenko, 2014 (UoE)

Attribute-value pairs

- **x** = {height=*180cm*, eyes="*blue*", job="*student*"}
- un-ordered "bag-of-features"
  - if structure is essential – embed it in the attributes
- Have to convert any dataset to this form
- Generally three types of attributes:
  - categorical: *red, blue, brown, yellow*
  - ordinal: *poor, satisfactory, good, excellent*
  - numeric: *-3.14, 6E23, 0, 1*

Source: Victor Lavrenko, 2014 (UoE)

- The algorithm has to predict a **discrete value** that identifies the input data as the member of a particular class/category.

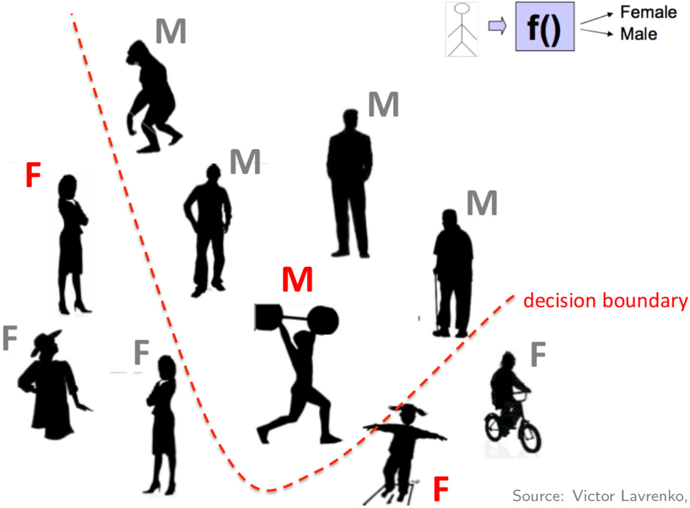
- The algorithm has to predict a **discrete value** that identifies the input data as the member of a particular class/category.
- E.g., in a training dataset of animal images, each photo would be pre-labeled as some type of animal.

- The algorithm has to predict a **discrete value** that identifies the input data as the member of a particular class/category.
- E.g., in a training dataset of animal images, each photo would be pre-labeled as some type of animal.
- The model would learn how to exploit the features present across different types of categories to make correct classifications.

Supervised Learning

Classification

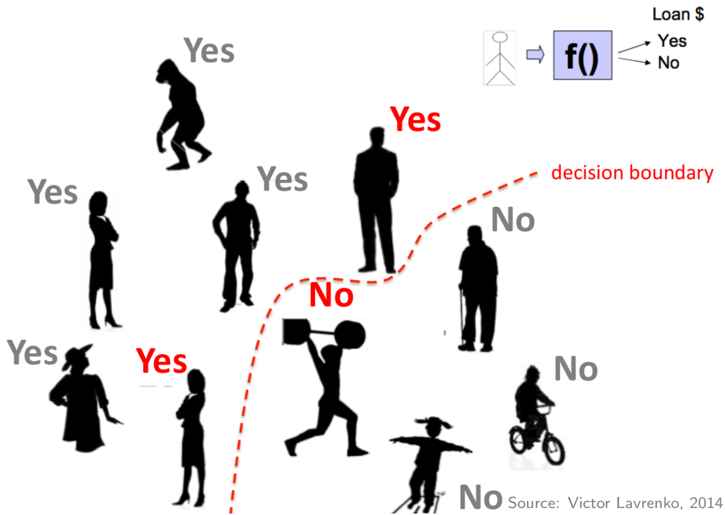
- The algorithm has to predict a **discrete value** that identifies the input data as the member of a particular class/category.
- E.g., in a training dataset of animal images, each photo would be pre-labeled as some type of animal.
- The model would learn how to exploit the features present across different types of categories to make correct classifications.
- The model would be evaluated by calculating how often it predicts the right label for some input data.
- Example algorithms: **Naive Bayes**, **Decision Trees**.



Source: Victor Lavrenko, 2014 (UoE)

Supervised Learning

Classification



Source: Victor Lavrenko, 2014 (UoE)

Bayesian classification

- Goal: learning function $f(x) \rightarrow y$
  - y ... one of k classes (e.g. spam/ham, digit 0-9)
  - $x = x_1 \ldots x_n$ – values of attributes (numeric or categorical)
- Probabilistic classification:
  - most probable class given observation: $\hat{y} = \arg\max_y P(y|x)$
- Bayesian probability of a class:

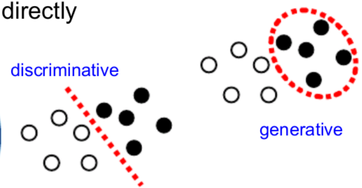
$$P(y|x) = \frac{\overbrace{P(x|y)}^{\text{class model}} \overbrace{P(y)}^{\text{prior}}}{\underbrace{\sum_{y'} P(x|y')P(y')}_{\text{normalizer } P(x)}}$$

Supervised Learning Classification Algorithms

Naïve Bayes: a generative model

- A complete probability distribution for each class
  - defines likelihood for any point *x*
  - P(class) via P(observation) $P(y|x) \propto P(x|y)P(y)$
  - can "generate" synthetic observations
    - will share many properties of the original data
- Not all probabilistic classifiers do this
  - possible to estimate P(*y*|*x*) directly
  - e.g. logistic regression:

$$P(y|x) = \frac{1}{z_y} \exp\left(\sum_i \lambda_i g_i(y,x)\right)$$



Continuous example

- Distinguish children from adults based on size
  - classes: {*a*,*c*}, attributes: height [cm], weight [kg]
  - training examples: $\{h_i, w_i, y_i\}$, 4 adults, 12 children
- Class probabilities: $P(a)=\frac{4}{4+12}=0.25\ ;P(c)=0.75$
- Model for adults:
  - height ~ Gaussian with mean, variance $\begin{cases} \mu_{h,a}=\frac{1}{4}\sum_{i:y_i=a} h_i \\ \sigma^2_{h,a}=\frac{1}{4}\sum_{i:y_i=a}(h_i-\mu_{h,a})^2 \end{cases}$
  - weight ~ Gaussian $\left(\mu_{w,a},\sigma^2_{w,a}\right)$
  - assume height and weight independent
- Model for children: same, using $(\mu_{h,c},\sigma^2_{h,c}),(\mu_{w,c},\sigma^2_{w,c})$

Continuous example

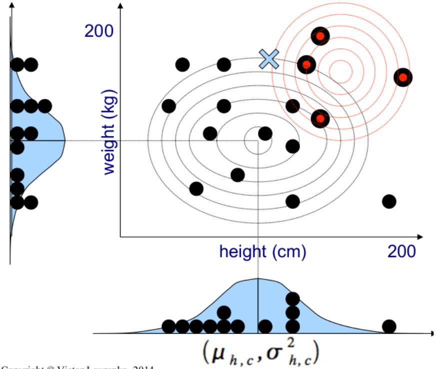
$$P(a)=\frac{4}{4+12}=0.25\ ;\ P(c)=0.75$$

$$p(h_x|c)=\frac{1}{\sqrt{2\pi\sigma^2_{h,c}}}\exp-\frac{1}{2}\left(\frac{(h_x-\mu_{h,c})^2}{\sigma^2_{h,c}}\right)$$

$$p(w_x|c)=\frac{1}{\sqrt{2\pi\sigma^2_{w,c}}}\exp-\frac{1}{2}\left(\frac{(w_x-\mu_{w,c})^2}{\sigma^2_{w,c}}\right)$$

$$p(h_x|a)=\frac{1}{\sqrt{2\pi\sigma^2_{h,a}}}\exp-\frac{1}{2}\left(\frac{(h_x-\mu_{h,a})^2}{\sigma^2_{h,a}}\right)$$

$$p(w_x|a)=\frac{1}{\sqrt{2\pi\sigma^2_{w,a}}}\exp-\frac{1}{2}\left(\frac{(w_x-\mu_{w,a})^2}{\sigma^2_{w,a}}\right)$$

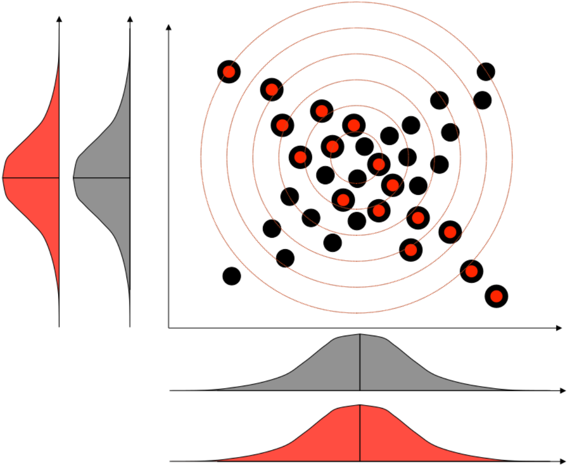


$$P(x|a)=p(h_x|a)\,p(w_x|a)$$

$$P(x|c)=p(h_x|c)\,p(w_x|c)$$

$$P(a|x)=\frac{P(x|a)P(a)}{P(x|a)P(a)+P(x|c)P(c)}$$

Problems with Naïve Bayes



Discrete example: spam

- Separate spam from valid email, attributes = words

D1: "send us your password"	spam
D2: "send us your review"	ham
D3: "review your password"	ham
D4: "review us"	spam
D5: "send your password"	spam
D6: "send us your account"	spam

new email: "review us now"

P (spam) = 4/6   P (ham) = 2/6

spam	ham	
2/4	1/2	password
1/4	2/2	review
3/4	1/2	send
3/4	1/2	us
3/4	1/2	your
1/4	0/2	account

$$P(\text{review us}\,|\text{spam}) = P(0,1,0,1,0,0\,|\text{spam}) = (1-\frac{2}{4})(\frac{1}{4})(1-\frac{3}{4})(\frac{3}{4})(1-\frac{3}{4})(1-\frac{1}{4})$$

$$P(\text{review us}\,|\text{ham}) = P(0,1,0,1,0,0\,|\text{ham}) = (1-\frac{1}{2})(\frac{2}{2})(1-\frac{1}{2})(\frac{1}{2})(1-\frac{1}{2})(1-\frac{0}{2})$$

$$P(\text{ham}\,|\text{review us}) = \frac{0.0625 \times 2/6}{0.0625 \times 2/6 + 0.0044 \times 4/6} = 0.87$$ (note identical example)

Problems with Naïve Bayes

- Zero-frequency problem
  - any mail containing "account" is spam: P(account|ham) = 0/2
  - solution: never allow zero probabilities
    - Laplace smoothing: add a small positive number to all counts: $$P(w|c)=\frac{num(w,c)+\epsilon}{num(c)+2\epsilon}$$
    - may use global statistics in place of ε: *num(w) / num*
  - very common problem (Zipf's law: 50% words occur once)
- Assumes word independence
  - every word contributes independently to P(spam|email)
  - fooling NB: add lots of "hammy" words into spam email

Predict if John will play tennis

- Hard to guess
- Try to *understand* when John plays
- Divide & conquer:
  - split into subsets
  - are they pure? (all yes or all no)
  - if yes: stop
  - if not: repeat
- See which subset new data falls into

Training examples: **9 yes / 5 no**

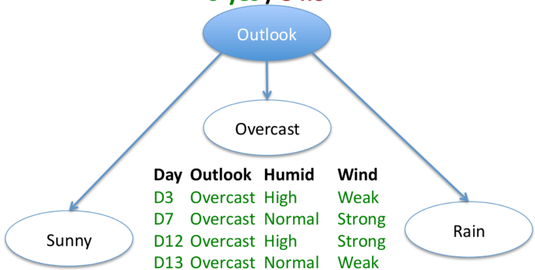
Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

New data:

D15	Rain	High	Weak	?

Decision Trees

Supervised Learning Classification Algorithms



Overcast:

Day	Outlook	Humid	Wind
D3	Overcast	High	Weak
D7	Overcast	Normal	Strong
D12	Overcast	High	Strong
D13	Overcast	Normal	Weak

4 yes / 0 no
pure subset

Sunny:

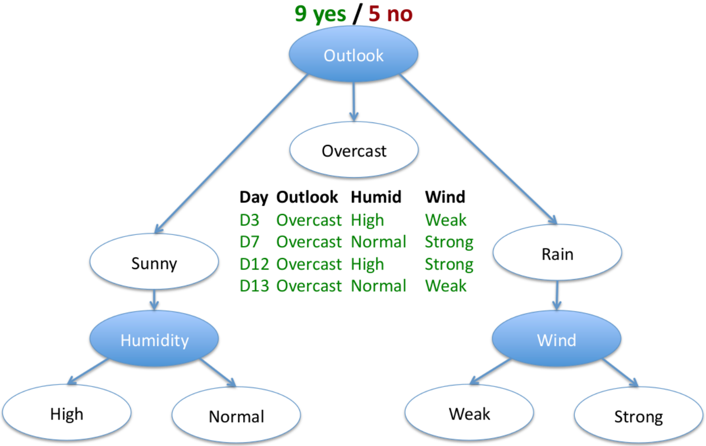
Day	Outlook	Humid	Wind
D1	Sunny	High	Weak
D2	Sunny	High	Strong
D8	Sunny	High	Weak
D9	Sunny	Normal	Weak
D11	Sunny	Normal	Strong

2 yes / 3 no
split further

Rain:

Day	Outlook	Humid	Wind
D4	Rain	High	Weak
D5	Rain	Normal	Weak
D6	Rain	Normal	Strong
D10	Rain	Normal	Weak
D14	Rain	High	Strong

3 yes / 2 no
split further



Overcast:

Day	Outlook	Humid	Wind
D3	Overcast	High	Weak
D7	Overcast	Normal	Strong
D12	Overcast	High	Strong
D13	Overcast	Normal	Weak

Sunny → Humidity = High:

Day	Humid	Wind
D1	High	Weak
D2	High	Strong
D8	High	Weak

Sunny → Humidity = Normal:

Day	Humid	Wind
D9	Normal	Weak
D11	Normal	Strong

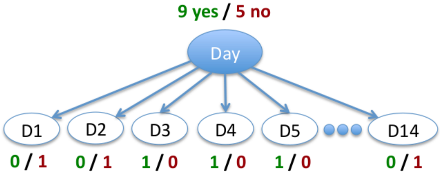
Rain → Wind = Weak:

Day	Humid	Wind
D4	High	Weak
D5	Normal	Weak
D10	Normal	Weak

Rain → Wind = Strong:

Day	Humid	Wind
D6	Normal	Strong
D14	High	Strong

Problems with Information Gain



all subsets perfectly pure => optimal split

- Biased towards attributes with many values
- Won't work for new data: D15 Rain High Weak
- Use GainRatio:

$$SplitEntropy(S,A) = -\sum_{V \in Values(A)} \frac{|S_V|}{|S|} \log \frac{|S_V|}{|S|}$$

A … candidate attribute
V … possible values of A
S … set of examples {X}
S_V … subset where $X_A = V$

$$GainRatio(S,A) = \frac{Gain(S,A)}{SplitEntropy(S,A)}$$

penalizes attributes with many values

Continuous Attributes

- Dealing with continuous-valued attributes:
  - create a split: (Temperature > 72.3) = True,False
- Threshold can be optimized (WF 6.1)

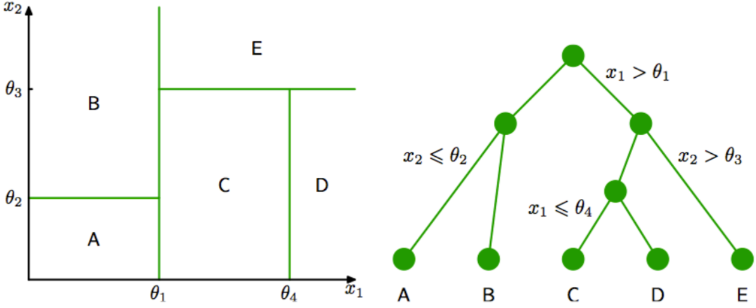
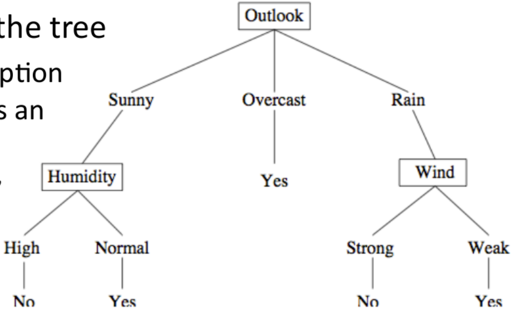


Figure credit: Chris Bishop, PRML

Trees are interpretable

- Read rules off the tree
  - concise description of what makes an item positive
- No "black box"
  - important for users

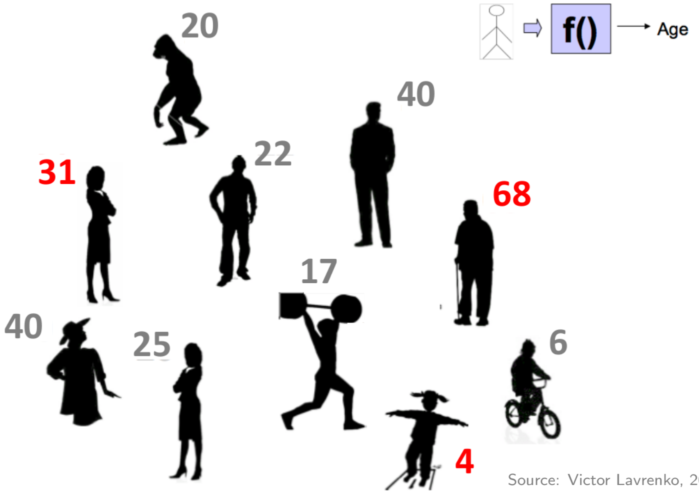


Rule:
(Outlook = Overcast) ∨
(Outlook = Rain ∧ Wind = Weak) ∨
(Outlook = Sunny ∧ Humidity = Normal)

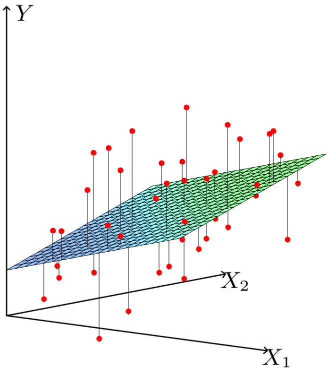
logical formula in DNF (disjunctive normal form)

Figure credit: Tom Mitchell, 1997

- The algorithm has to predict a **continuous value** that corresponds to some independent variable (e.g., price of an item).
- E.g., in a training dataset of car prices, each car could have labels identifying its features (e.g., N of doors, power, etc.) and the corresponding prices.
- The model would be evaluated by checking how close to the real prices its predictions are (e.g., RMSE).



Source: Victor Lavrenko, 2014 (UoE)



- A common choice: *squared error* (makes the maths easy)

$$O(\mathbf{w}) = \sum_{i=1}^{n} (y_i - \mathbf{w}^T \mathbf{x}_i)^2$$

- In the picture: this is sum of squared length of black sticks.
- (Each one is called a *residual*, i.e., each $y_i - \mathbf{w}^T \mathbf{x}_i$)

Source: Victor Lavrenko, 2014 (UoE)

Review: Supervised Learning Pipeline

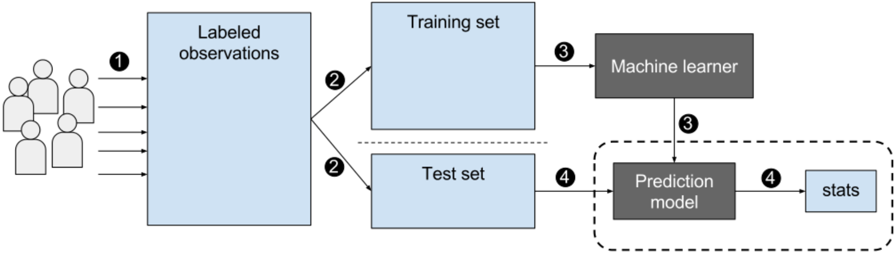


Figure: Pipeline of procedure carried out in supervised learning paradigm.

Source: *Nvidia Blog*

Outline

Unsupervised Learning

The main incentives are related to the problems/limits of the supervised learning models:

Unsupervised Learning

The main incentives are related to the problems/limits of the supervised learning models:

- Large datasets with good quality annotations are hard to find.

Unsupervised Learning

The main incentives are related to the problems/limits of the supervised learning models:

- Large datasets with good quality annotations are hard to find.
- Some problems might not have any solutions available yet.

The main incentives are related to the problems/limits of the supervised learning models:

- Large datasets with good quality annotations are hard to find.
- Some problems might not have any solutions available yet.
- The number of required labels might not be constant.

Unsupervised Learning

The main incentives are related to the problems/limits of the supervised learning models:

- Large datasets with good quality annotations are hard to find.
- Some problems might not have any solutions available yet.
- The number of required labels might not be constant.

Solution: **get these labels by analyzing the data and its inter-correlations.**

Unsupervised Learning

- A computational model is given a bunch of **data without explicit instructions** about what to do with it - the dataset is a **collection of samples with no labels**

- A computational model is given a bunch of **data without explicit instructions** about what to do with it - the dataset is a **collection of samples with no labels**
- The model attempts to **automatically find patterns in the data by extracting features and analyzing its structure**.

Unsupervised Learning

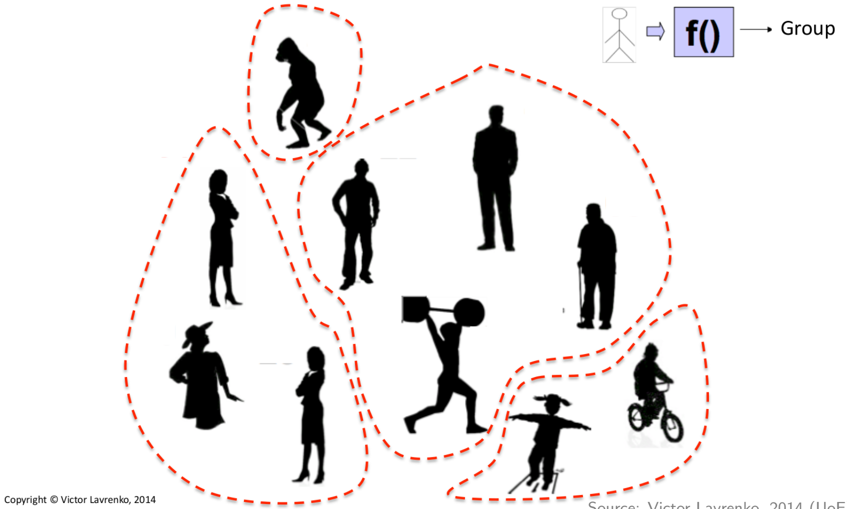
- A computational model is given a bunch of **data without explicit instructions** about what to do with it - the dataset is a **collection of samples with no labels**
- The model attempts to **automatically find patterns in the data by extracting features and analyzing its structure**.
- Depending on the problem, the unsupervised learning model can organize this data in various ways.

Unsupervised Learning

- A computational model is given a bunch of **data without explicit instructions** about what to do with it - the dataset is a **collection of samples with no labels**
- The model attempts to **automatically find patterns in the data by extracting features and analyzing its structure**.
- Depending on the problem, the unsupervised learning model can organize this data in various ways.
- The **evaluation** procedure might be **complicated**, as there are no ground-truths to compare the model results to.

Grouping of multi-dimensional data based on its features.

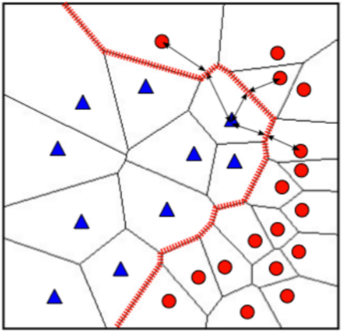
- Essentially, the model looks at the training data features and attempts to **group the points into inter-similar clusters**.
- E.g., looking at a collection of birds and separating them into groups based on their wing or beak shape.
- Example algorithms: **k-Nearest Neighbour** (kNN), **k-Means**.
  - **kNN Classification**: most common category of N neighbours
  - **kNN Regression**: average of values of N neighbours



Source: Victor Lavrenko, 2014 (UoE)

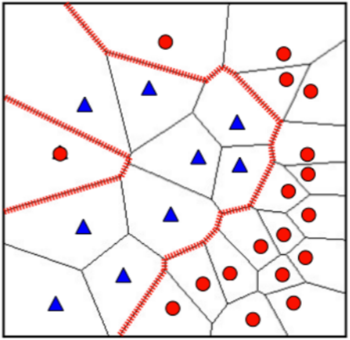
Nearest-neighbor classification

- Use the intuition to classify a new point x:
  - find the most similar training example x’
  - predict its class y’
- Voronoi tesselation
  - partitions space into regions
  - boundary: points at same distance from two different training examples
- classification boundary
  - non-linear, reflects classes well
  - compare to NB, DT, logistic
  - impressive for simple method



Nearest neighbour: outliers

- Algorithm is sensitive to outliers
  - single mislabeled example dramatically changes boundary
- No confidence P(y|x)
- Insensitive to class prior
- Idea:
  - use more than one nearest neighbor to make decision
  - count class labels in k most similar training examples
    - many "triangles" will outweigh single "circle" outlier

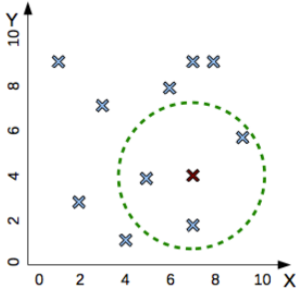


kNN: Practical Issues

- Resolving Ties
  - equal number of positive/negative neighbours
  - use odd k (doesn't solve multiclass)
  - breaking ties:
    - random: flip a coin to decide positive/negative
    - prior: pick a class with greater prior
    - nearest: use 1-nn classifier to decide
- Missing Values
  - have to "fill in", otherwise can't compute distance
  - key concern: should affect distance as little as possible
  - reasonable choice: average value across entire dataset

Why is kNN slow?

What you see



Find nearest neighbors
of the testing point (red)

What algorithm sees

- Training set:

  {(1,9), (2,3), (4,1),
  (3,7), (5,4), (6,8),
  (7,2), (8,8), (7,9), (9,6)}

- Testing instance:

  (7,4)

- Nearest neighbors?

  compare one-by-one
  to each training instance

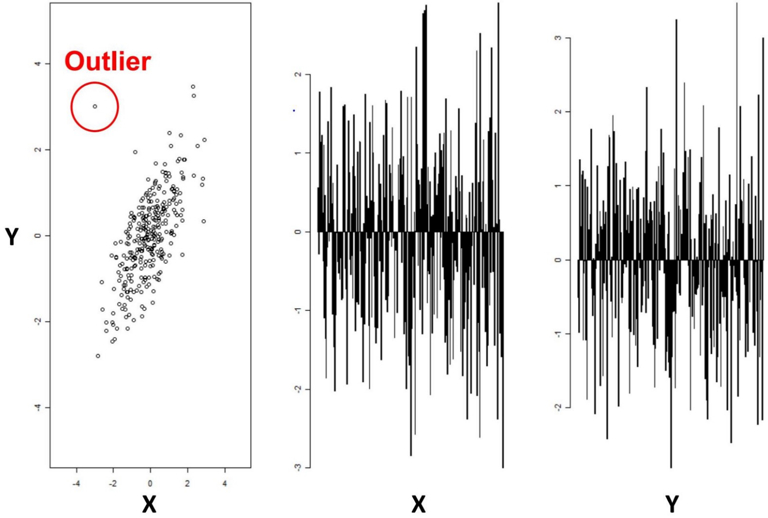
- n comparisons
- each takes d operations

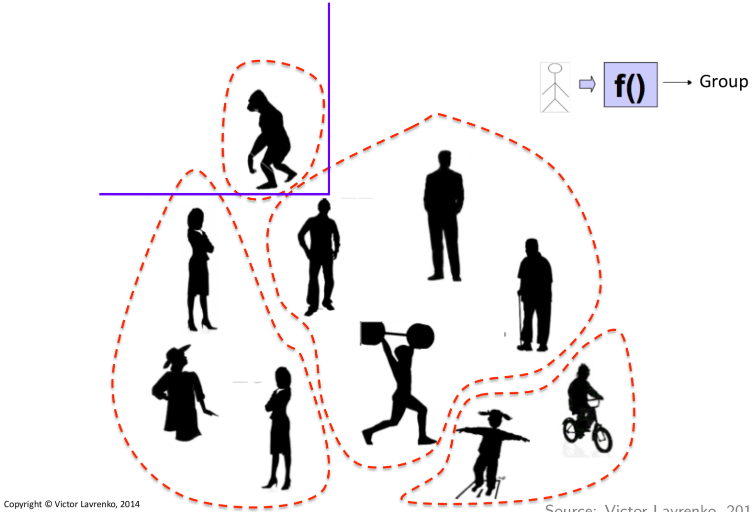
Anomaly Detection (Outlier Detection)

Identification of items, events or observations that differ significantly from the majority of the data and, thus, raise suspicion.

- E.g.: detection of fraudulent transactions by looking for weird patterns in customer's purchasing behaviour.
- The effectiveness of applied method heavily depends on the data domain and its specificity.
- There hardly exists a standard algorithm to perform this (ranges from simple statistical solutions, to complex deep learning approaches).
- Can be also used to flag outliers in a supervised learning problem.

Anomaly Detection (Outlier Detection)
Unsupervised Learning

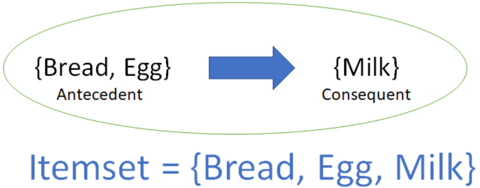




Source: Victor Lavrenko, 2014 (UoE)

Predicting some associated features based on their inter-correlations.

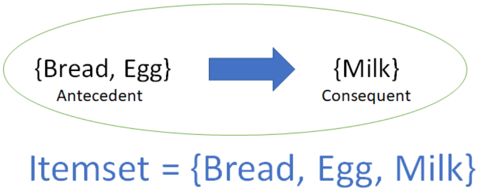
- The model looks at a couple of key attributes of a data point and predicts the other attributes with which these are commonly associated.
- E.g.: recommendation of related products when forming a shopping cart in an on-line store.



Source: Anisha Garg, *Towards Data Science*

The strength of the association between items may be defined using various metrics:

- **Support** - the frequency of an itemset among all the transactions.
- **Confidence** - the likeliness of occurrence of consequent on the cart given that the cart already has the antecedents.



Source: Anisha Garg, *Towards Data Science*

The strength of the association between items may be defined using various metrics:

- **Support** - the frequency of an itemset among all the transactions.
- **Confidence** - the likeliness of occurrence of consequent on the cart given that the cart already has the antecedents.

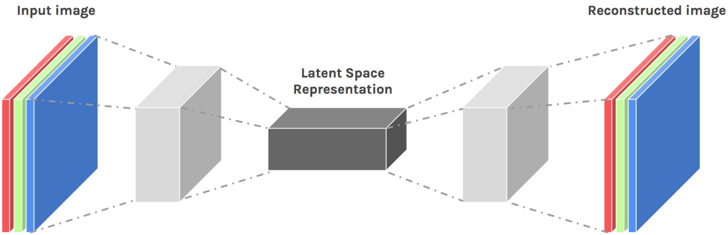
$$Confidence(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Transactions\ containing\ X}$$

$$Support(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Total\ number\ of\ transactions}$$

Source: Anisha Garg, *Towards Data Science*

Compress input data into a set of features, then try to recreate the original.

- Hardly useful on its own, but has numerous application when combined with other methods.
- E.g., by training an autoencoder on noisy/clean images, we will develop a noise-removal model, which could be then used to reduce noise in medical scans.
- More common to deep learning.



Outline

Semi-Supervised Learning

A middle ground between supervised and unsupervised machine learning methods:

- Uses a training dataset with both labeled and unlabeled data.
- Useful when extracting relevant features from the data is difficult, and labeling examples is a time-intensive task for experts.
- E.g., think of an autoencoder that generates additional synthetic data for a supervised learning algorithm, also GANs).

Reinforcement Learning

AI agents that learn how to accomplish a particular goal or improve their performance on a specific task based on the rewards they get.

- Predict such next step that earns the biggest final reward.
- E.g., game AIs.
- Iterative process: the more rounds of feedback, the better the agents strategy becomes.
- Very effective in robotics (e.g., automated steering)
- Main issues: how to define the reward function?

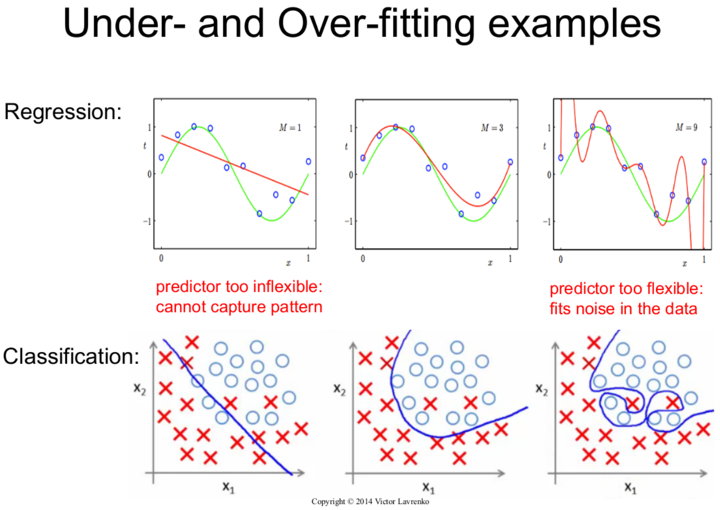
Source: *Video Link*

Outline

Generalization

- Training data: $\{x_i, y_i\}$
  - examples that we used to train our predictor
  - e.g. all emails that our users labelled ham / spam
- Future data: $\{x_i, ?\}$
  - examples that our classifier has never seen before
  - e.g. emails that will arrive tomorrow
- Want to do well on future data, not training
  - not very useful: we already know y_i
  - easy to be perfect on training data (DT, kNN, kernels)
  - does not mean you will do well on future data
    - can over-fit to idiosyncrasies of our training data

Source: Victor Lavrenko, 2014 (UoE)

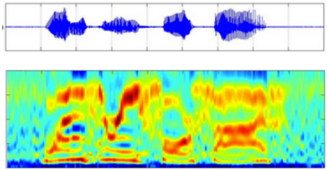


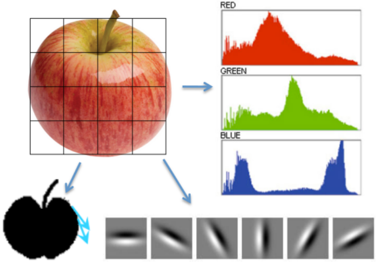
Source: Victor Lavrenko, 2014 (UoE)

Outline

Features in Machine Learning

- 1-line summary of ML: $y = \mathrm{sgn}\,(\mathbf{w}^T\mathbf{x} + b)$
  - SVM can learn very effective weights **w**
  - … if you use the right representation **x**

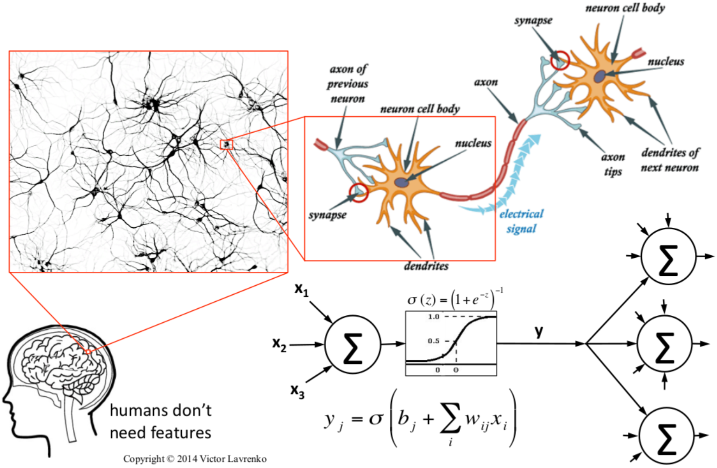




Would be nice to "learn" these representations automatically…

Neural Networks: Biological Inspiration

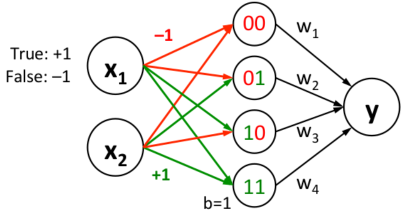
Neurons and the brain



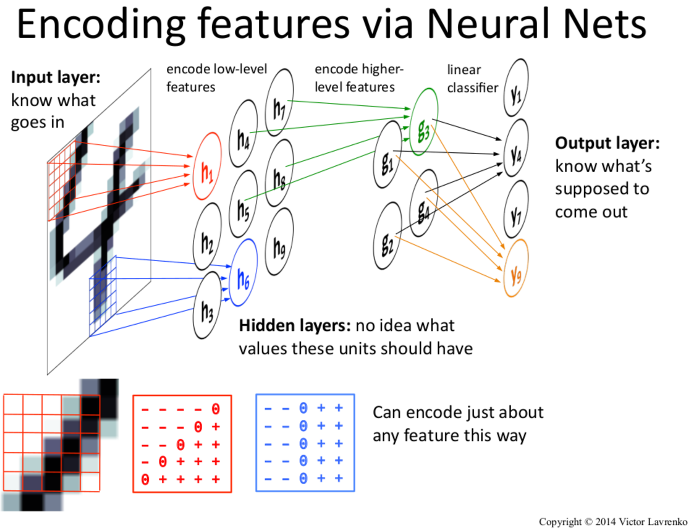
Representation Power of NNs

- NN with 1 hidden layer can represent:
  - any bounded continuous function (to arbitrary ε)
    - Universal Approximation Theorem [Cybenko 1989]
  - any Boolean function (exactly)

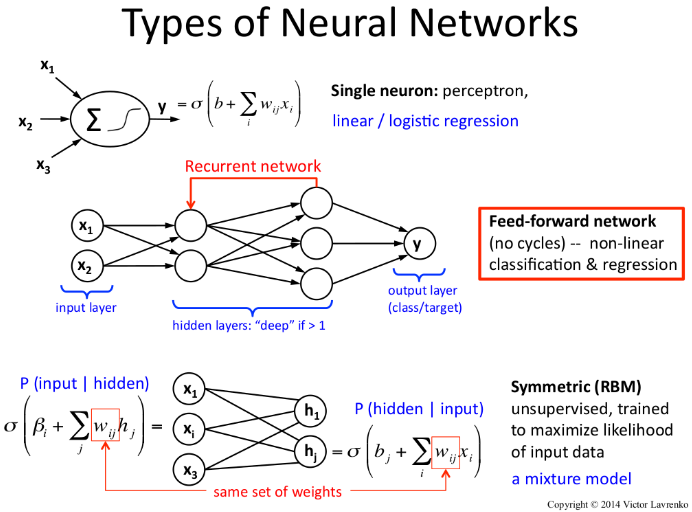
    may require 2^d hidden units for input $x_1 \ldots x_d$



Neural Networks: Feature Extraction



Types of Neural Networks

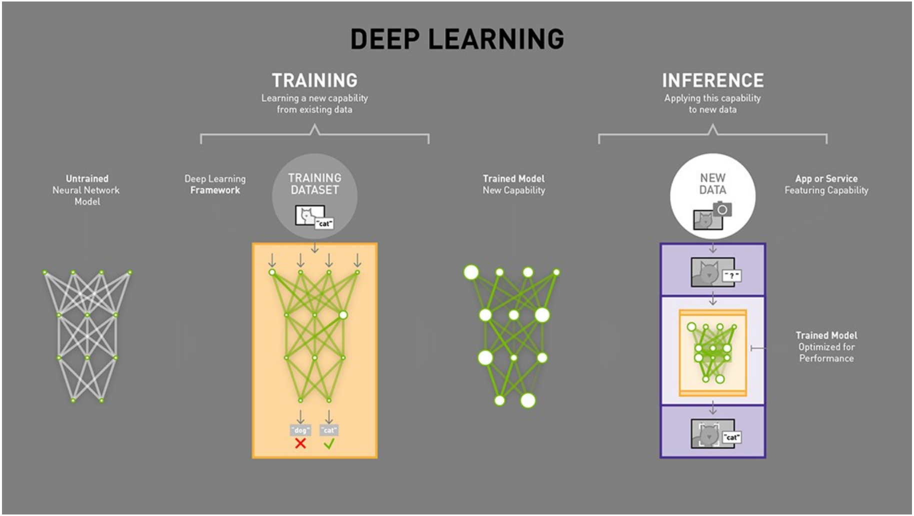


Neural Networks: Summary

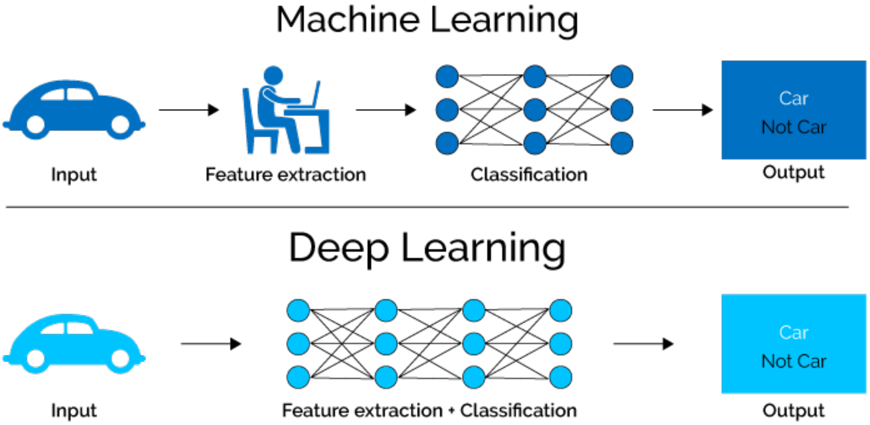
"Third generation" AI method, capable of modeling non-linear data effectively.

- Algorithms (somewhat) based on the human brain neural circuitry.
- Don't require manual feature engineering - determine the best features by looking at the data and choosing the best features for the task.
- Variations of NNs are the current state-of-art predictive models for numerous problems and domains (machine translation, language modeling, speech recognition, image classification, image segmentation, etc.)

Outline


DEEP LEARNING
TRAINING
Learning a new capability from existing data
INFERENCE
Applying this capability to new data
Untrained Neural Network Model
Deep Learning Framework
TRAINING DATASET
"cat"
"dog"
"cat"
Trained Model New Capability
NEW DATA
"?"
"cat"
App or Service Featuring Capability
Trained Model Optimized for Performance

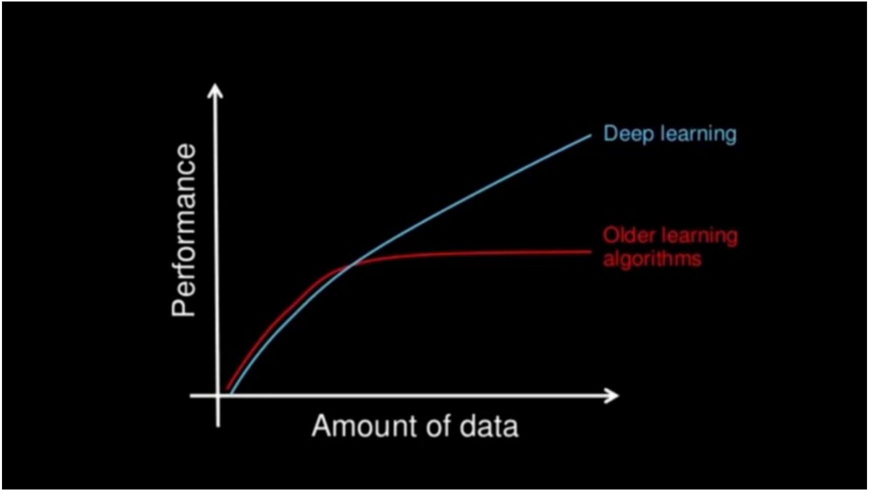
ML vs. DL.



Machine Learning
Input
Feature extraction
Classification
Car
Not Car
Output
Deep Learning
Input
Feature extraction + Classification
Car
Not Car
Output

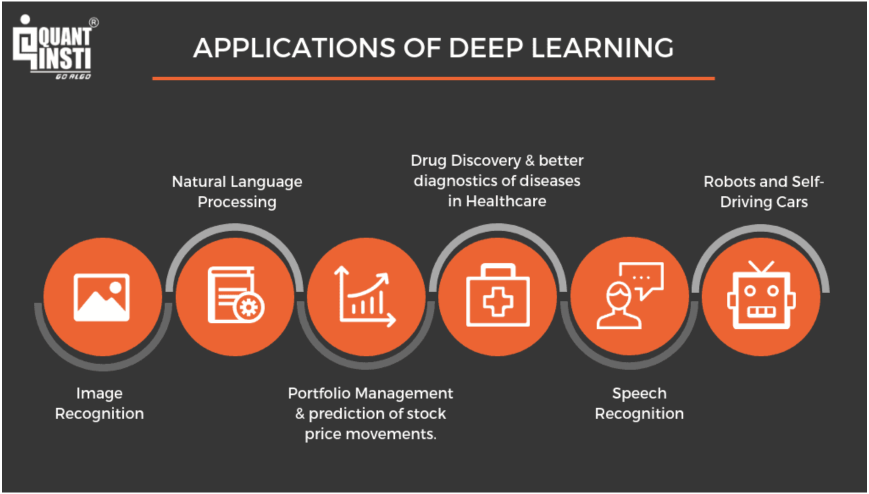
- Large amounts of data (e.g., Google, Amazon)
- Advanced algorithms (e.g., variants of DNNs)
- Computing Hardware (cheap & powerful GPUs/TPUs)

ML vs. DL.


Performance
Deep learning
Older learning
algorithms
Amount of data

What are DL models good at?

Basically... Everything


APPLICATIONS OF DEEP LEARNING
Natural Language Processing
Drug Discovery & better diagnostics of diseases in Healthcare
Robots and Self-Driving Cars
Image Recognition
Portfolio Management & prediction of stock price movements.
Speech Recognition

What are DL models good at?

Basically... Everything

What are DL models good at?

Basically... Everything specific


WHAT AI STRUGGLES WITH
Understand nuances
Detecting and decoding emotional subtleties based on social cues
Create original content
Producing work autonomously without large sets of data and defined parameters
Filter biases
Identifying biases based on social or ethical consciousness

AI Struggles: Ambiguity

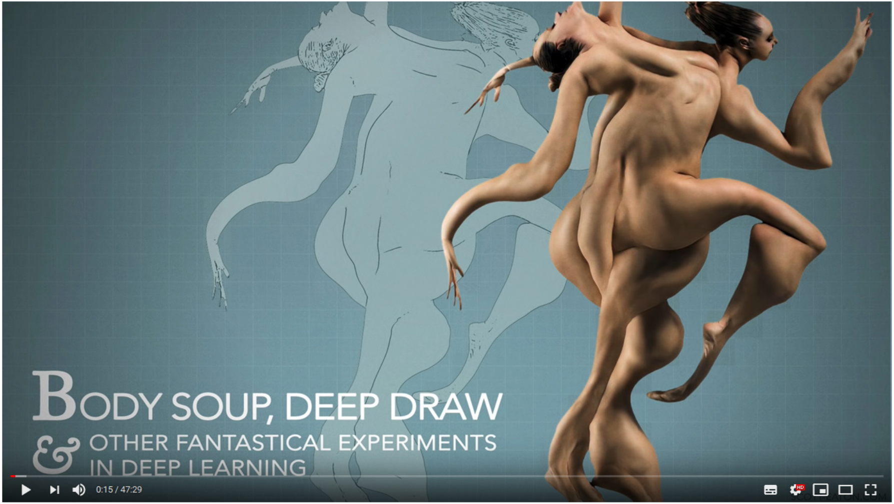
Are you a candle? Because youre so hot of the looks with you.

I want to get my heart with you.

You look like a thing and I love you.

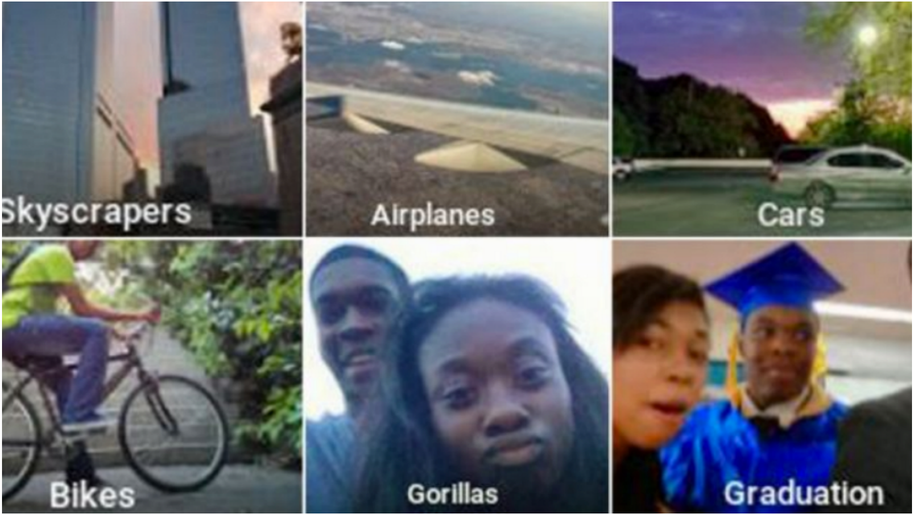
AI Struggles: Originality (?)

AI Struggles: Originality (?)

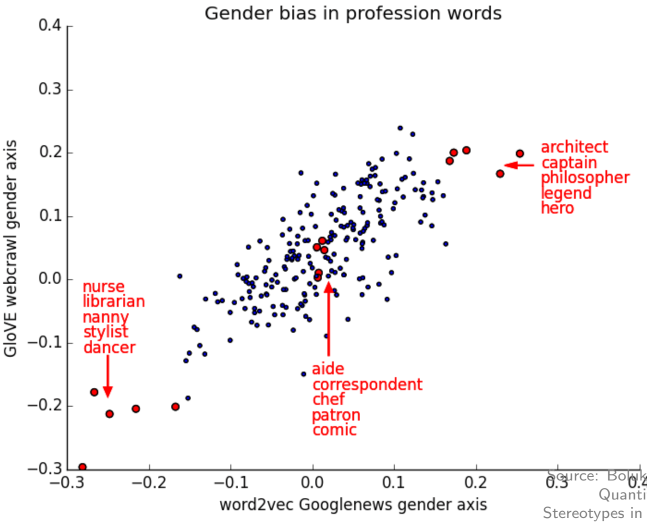


Source: *Video Link*

AI Struggles: Bias & Morality

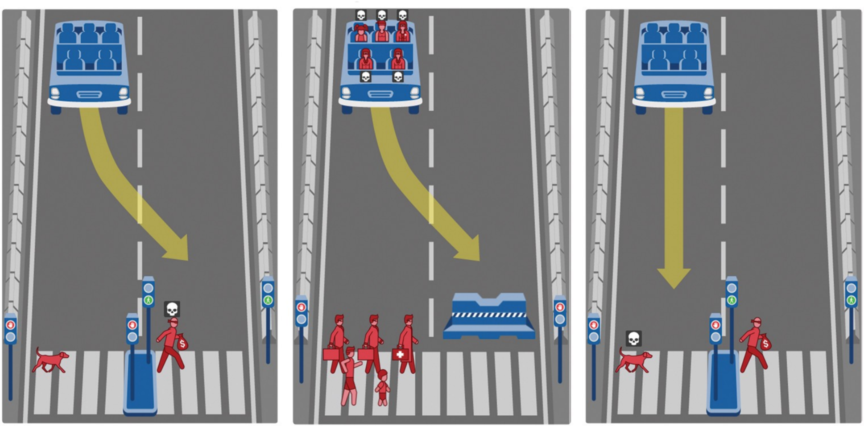


AI Struggles: Bias & Morality



Source: Bolukbasi et. al. (2016). Quantifying and Reducing Stereotypes in Word Embeddings.

AI Struggles: Bias & Morality



Source: *Moral Machine*

Summary

- **Artificial General Intelligence (AGI) does not exist yet!** The current AI cannot understand things like common sense, morality, or emotions; But...
- For some specific tasks, **AI can perform superior to humans**, e.g.:
  - In financial field, ultra-high-speed and frequency algorithmic transactions correspond to more than a half of all market transactions.
  - In medical diagnostics, some image recognition models achieve diagnostic accuracy higher than an average doctor.

Random Quote

Success in creating effective AI could be the biggest event in the history of our civilization. Or the worst. We just dont know. So we cant know for sure if well be infinitely helped by AI, or ignored by it and side-lined, or conceivably destroyed by it. (S. Hawking)

Thank you!