SF1935 – Probability Theory and Statistics with **Application to Machine Learning**

Lecture 1: Introduction to Machine Learning

Pawel Herman

Computational Science and Technology (CST)

KTH Royal Institute of Technology

Course moment outline

1. Introduction to machine learning, a probabilistic approach
2. Linear regression with focus on Bayesian methods – towards project
3. Lab sessions – lab/project work/help sessions: May 5 & 11/16

   project demonstration: May 16 & 23

Main reasons for the AI/ML hype

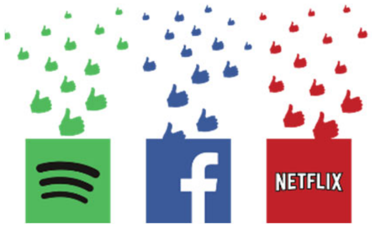
- Impressive scope of applications

Autonomous vehicles, agents, robotics

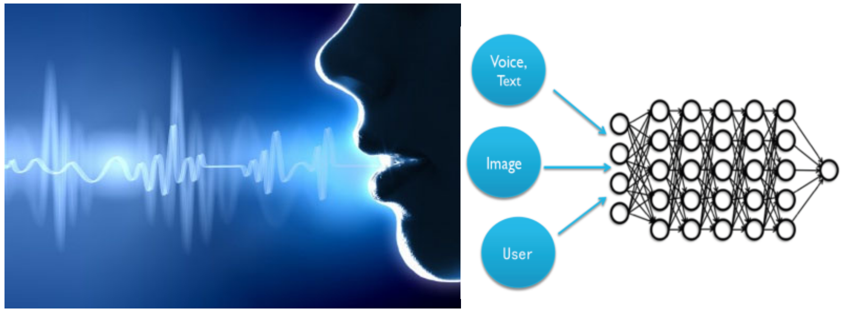
Medicine, healthcare

IBM WATSON

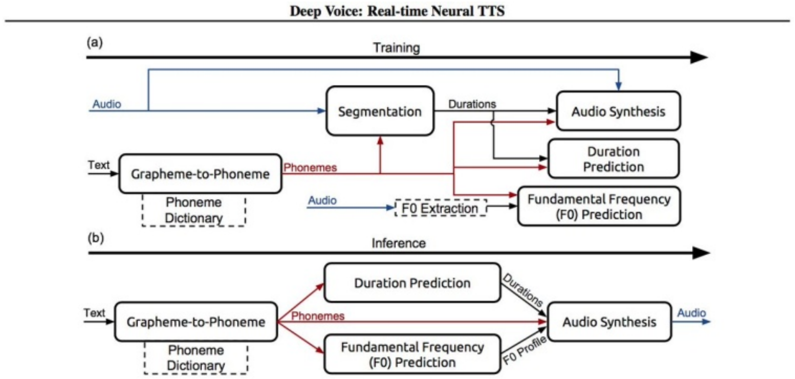
Recommendation systems



Automated speech recognition



Machine translation + text to speech transformation



Main reasons for the AI/ML hype

- Impressive scope of appli

Auton omou Tesla icles,
agents, robotics

- Image recognition
- Time series prediction
- Autonomous planning, logistics
- Decision making
- Spam filtering
- Financial applications
- Recommendation systems

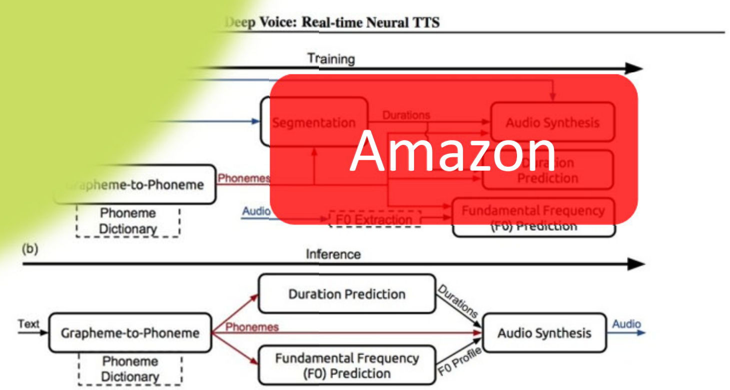
Netflix

Automated speech recognition

DeepVoice

Colorful Clouds

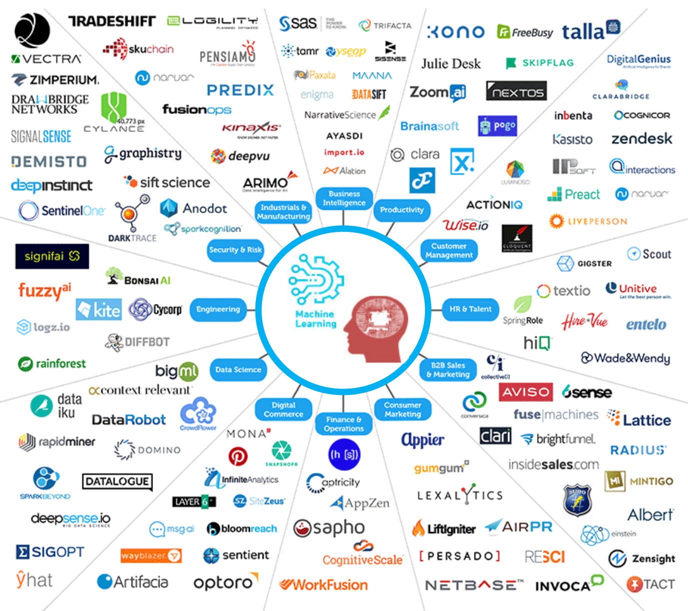
ation + text to speech
nsformation



Amazon

Main reasons for the AI/ML hype

- Impressive scope of applications and intriguingly good performance
- A huge opportunity for industry to improve products, services at a relatively low cost – large profit margin



Main reasons for the AI/ML hype

- Impressive scope of applications and intriguingly good performance
- A huge opportunity for industry to improve products, services at a relatively low cost – large profit margin
- Renewed scientific interest following Deep Learning success and implications for other fields

LIFE SCIENCES ENGINEERING PHYSICS ECONOMY

Main reasons for the AI/ML hype

- Impressive scope of applications and intriguingly good performance
- A huge opportunity for industry to improve products, services at a relatively low cost – large profit margin
- Renewed scientific interest following Deep Learning success
- Remarkable implications for other fields
- Good timing: data availability, growing compute power and the overall quest for automatization and optimisation

What is Machine Learning?

Learning comes in different forms

to gain knowledge, skill, experience.....

to modify behaviour based on the acquired experience, knowledge etc. to adapt (perform "better")....

using past experience to make future decisions or guide future actions....

What is Machine Learning?

Learning comes in different forms

to gain knowledge, skill, experience.....

to modify behaviour based on the acquired experience, knowledge etc. to adapt (perform "better")....

using past experience to make future decisions or guide future actions....

But what about Machine Learning?

"Engineering" paradigm: to automate the process of learning, to use experience (data) rather than expert knowledge to automatically build systems that perform complex tasks

Tom Mitchell: algorithm learning from experience to improve its performance in a specific task

What is Machine Learning?

Learning comes in different forms

to gain knowledge, skill, experience.....

to modify behaviour based on the acquired experience, knowledge etc. to adapt (perform "better")....

using past experience to make future decisions or guide future actions....

But what about Machine Learning?

"Engineering" paradigm: to automate the process of learning, to use experience (data) rather than expert knowledge to automatically build systems that perform complex tasks

Tom Mitchell: algorithm learning from experience to improve its performance in a specific task

Design and use algorithms that enable machines to learn from data rather than being explicitly programmed

What is Machine Learning?

Learning comes in different forms

to gain knowledge, skill, experience.....

to modify behaviour based on the acquired experience, knowledge etc. to adapt (perform "better")....

using past experience to make future decisions or guide future actions....

But what about Machine Learning?

"Engineering" paradigm: to automate the process of learning, to use experience (data) rather than expert knowledge to automatically build systems that perform complex tasks

"It is all about patterns" perspective:

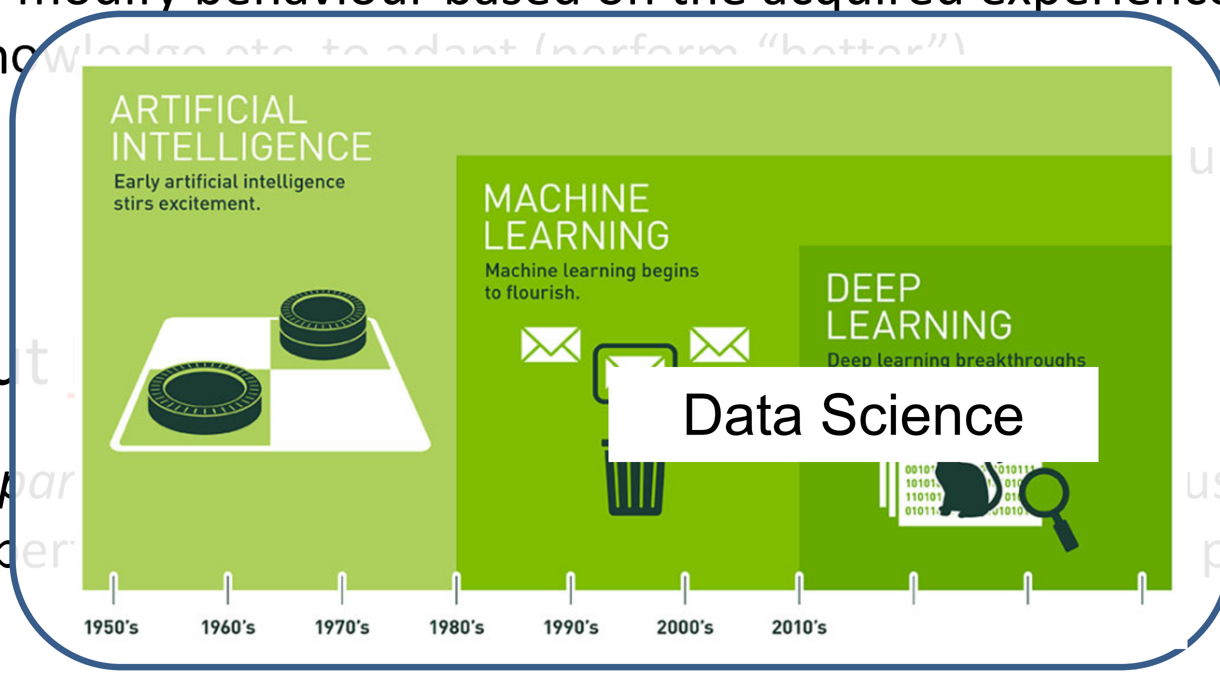
To automatically identify patterns and use them to predict future data.

What is Machine Learning?

Learning comes in different forms

to gain knowledge, skill, experience.....

to modify behaviour based on the acquired experience, kno



e decisions or

But what abou

“Engineering” use experience (data) rather than exp erform complex tasks

“It is all about patterns” perspective:

To automatically identify patterns and use them to predict future data.

What is Machine Learning?

*"The scientific discipline that explores the construction and study of algorithms that can **learn from data**."*
(*wikipedia.com*)

*"A computer's way of **learning from examples**."*
(*businessinsider.com*)

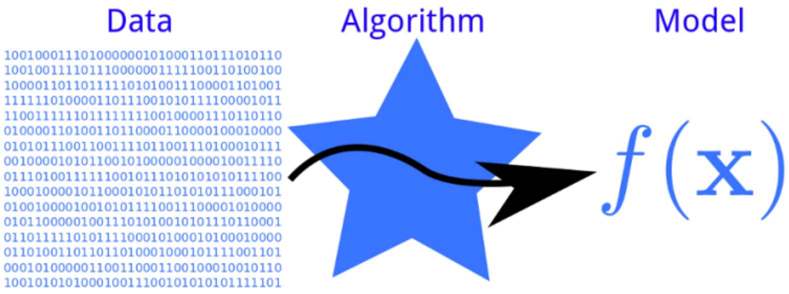
*"The science of getting computers to **act without being explicitly programmed**."*
(*Arthur Lee Samuel*)

*"The field of machine learning is concerned with the question of how to construct computer programs that **automatically improve with experience**."*
(*Tom Mitchell*)

*"A computer program is said to **learn from experience** E with respect to some class of **tasks** T and performance **measure** P, if its performance at tasks in T, as measured by P, improves with experience E."*
(*Tom Mitchell*)

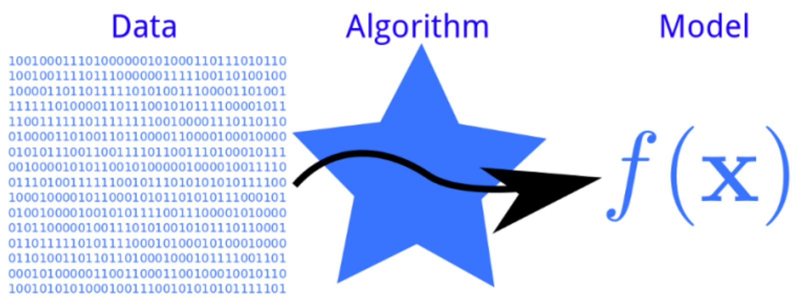
Machine Learning as a data driven approach

Learning from data

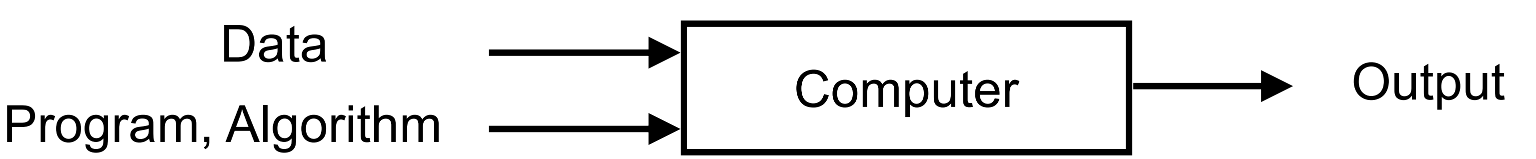


Machine Learning as a data driven approach

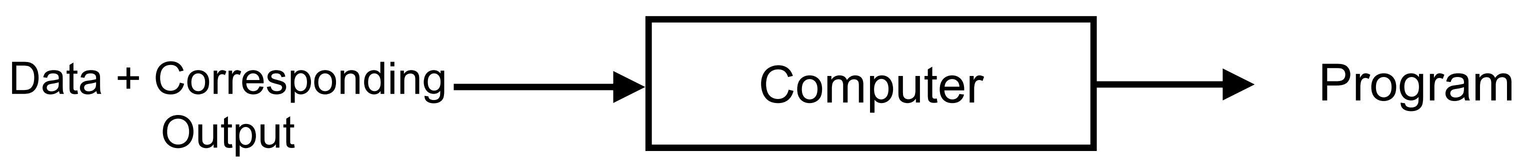
Learning from data



Classical approach to solving problems with computer algorithms

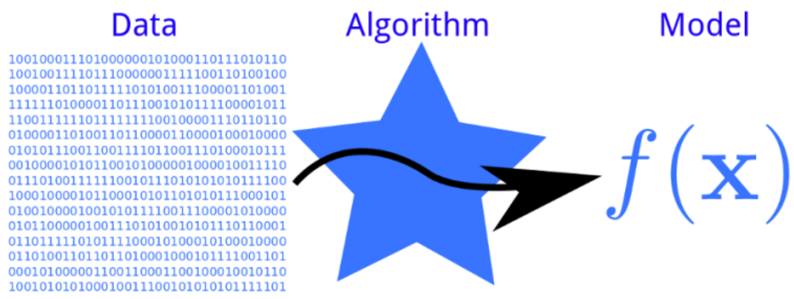


Machine Learning approach – data driven approach



Machine Learning as a data driven approach

Learning from data



Capability to *acquire* and *integrate* the *knowledge* automatically from **data** and improve performance by learning

Scope for continuous self-improvement, adaptation to changing environment

Discovery of patterns in complex data – data mining

Generating predictions based on previous experience (data driven)

Why do we need to "automate" learning from data?

- Motivation for ML (why don't we program machines explicitly?)

Some tasks cannot be defined well except by example.

The amount of knowledge available about certain tasks might be too large for explicit encoding by humans.

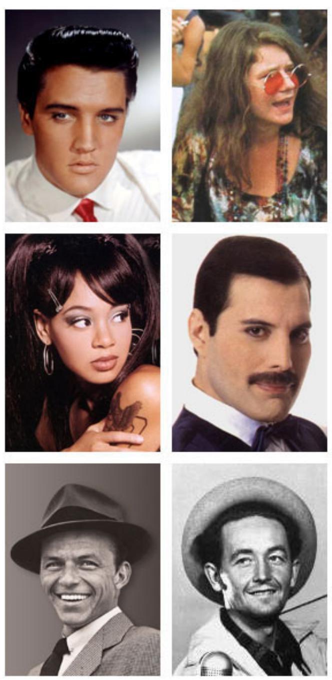
New knowledge about tasks is constantly being discovered .

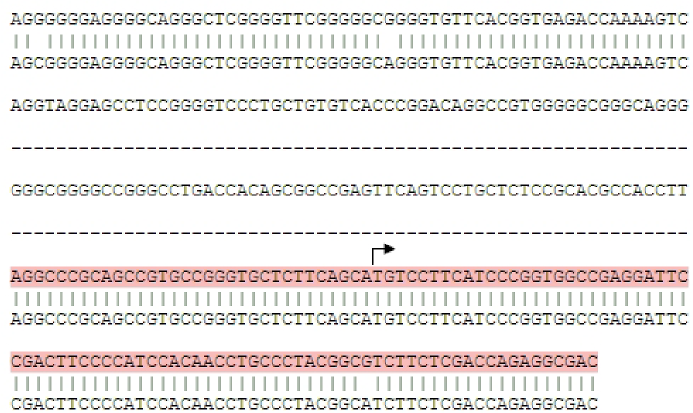
Environments change over time.

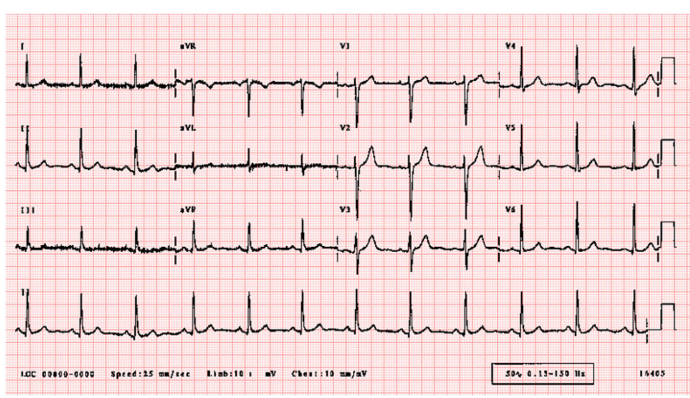
Important relationships and correlations can be hidden in large dimensionality of the data or behind nonlinearities

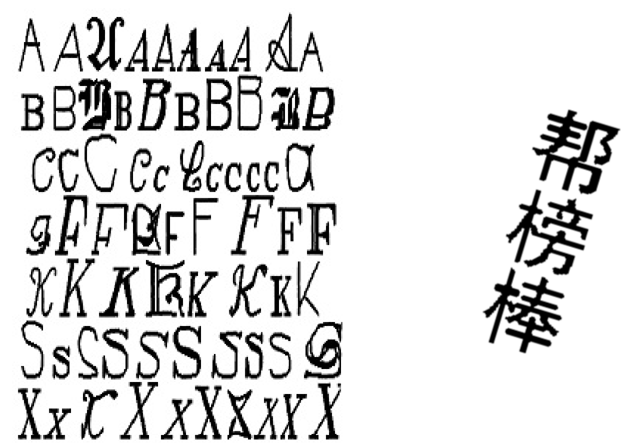
The art of searching for patterns

The problem of searching for patterns in data is a fundamental one and has a long and successful history.









The art of searching for patterns

The problem of searching for patterns in data is fundamental

- the ubiquity of patterns in the surrounding world
- it could be an object, process, event etc. and is described by attributes (features)
- the importance of finding and categorising objects

ML for pattern recognition

- the ubiquity of patterns in the surrounding world
- it could be an object, process, event etc. and is described by attributes (features)
- the importance of finding and categorising objects

Examples of real-world problems involving *Pattern Recognition*

Identify risk factors for cancer based on clinical, demographic variables

Diagnose dementia based on electrophys. data

Predict the price of a stock on the basis of company performance measures and economic data

Predict short-term electricity demand and wind power production capabilities

ML for pattern recognition

- the ubiquity of patterns in the surrounding world
- it could be an object, process, event etc. and is described by attributes (features)
- the importance of finding and categorising objects

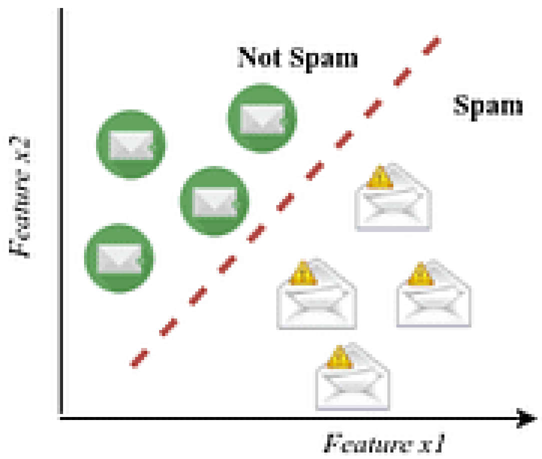
Pattern recognition amounts to grouping objects and often assigning categories (classes) to the objects.

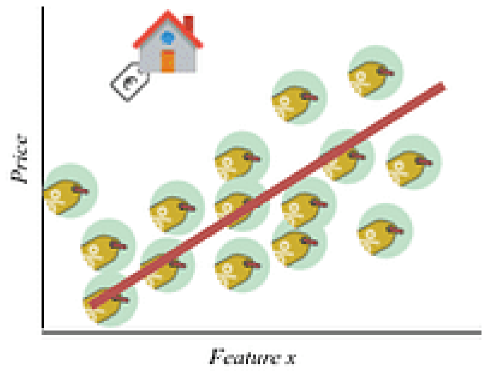
clustering

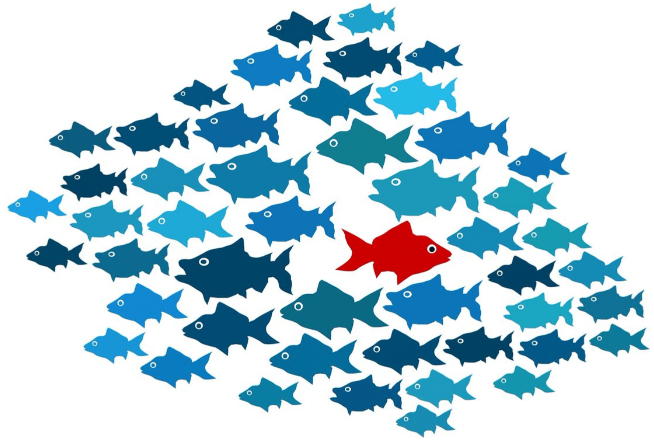
discrete classes - ***classification***

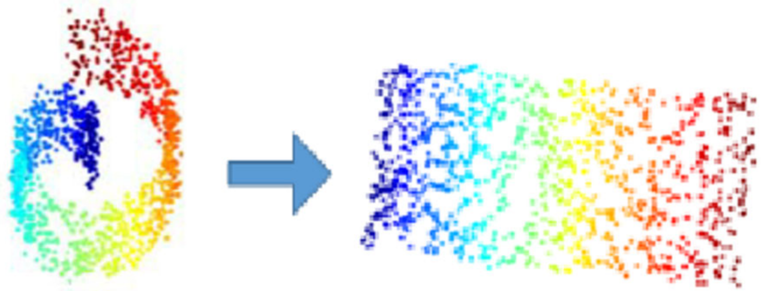
continuous labels - ***regression***

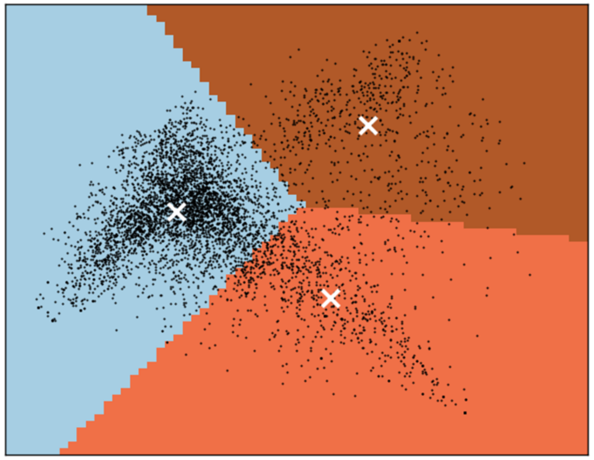
Pattern recognition tasks for ML: some toy examples

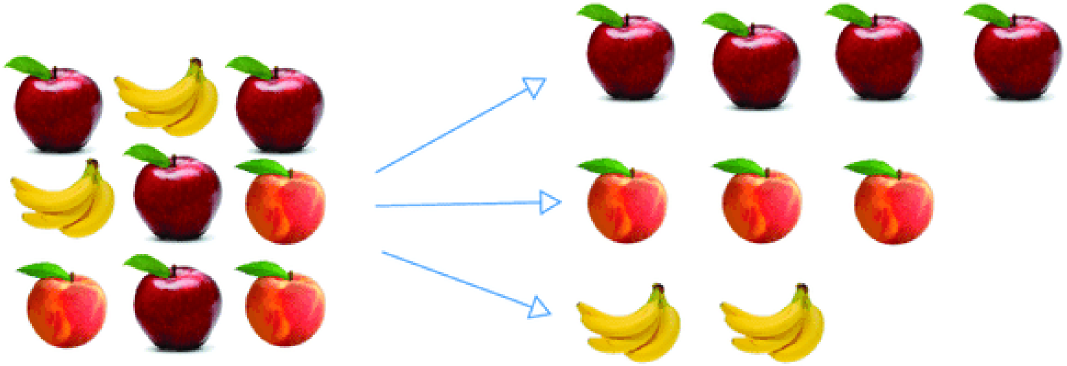












Types of inductive learning

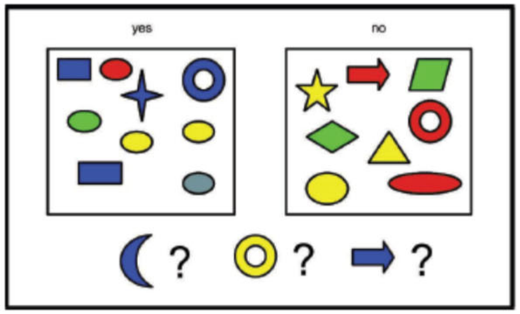
- Inductive learning: *learn by example* and *induce a general rule* from a set of examples/observations
- The *type of feedback* available for learning determines the *nature of the learning problem*

Types of inductive learning

- Inductive learning: *learn by example* and *induce a general rule* from a set of examples/observations
- The *type of feedback* available for learning determines the *nature of the learning problem*

Supervised learning

Regression vs Classification



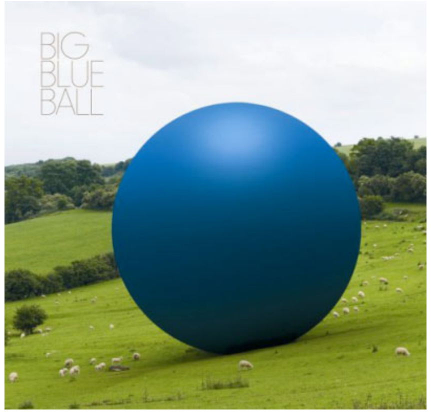
D features (attributes)

N cases

Color	Shape	Size (cm)	Label
Blue	Square	10	1
Red	Ellipse	2.4	1
Red	Ellipse	20.7	0

Murphy, 2012

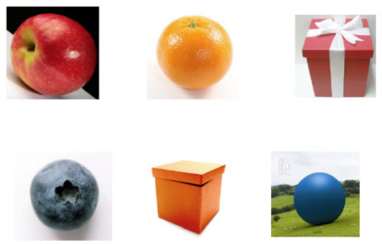
How would you group these objects and why?

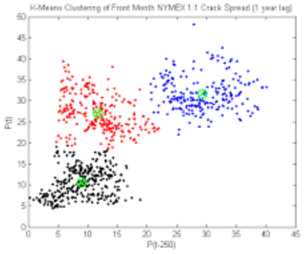


Types of inductive learning

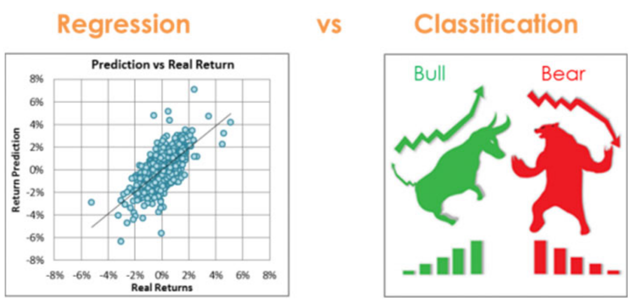
- Inductive learning: *learn by example* and *induce a general rule* from a set of examples/observations
- The *type of feedback* available for learning determines the *nature of the learning problem*

Unsupervised learning





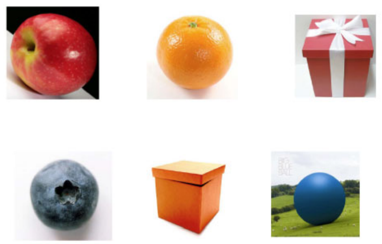
Supervised learning

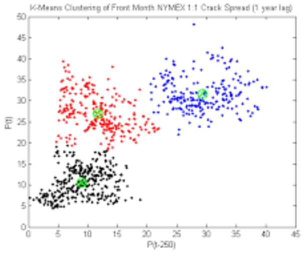


Types of inductive learning

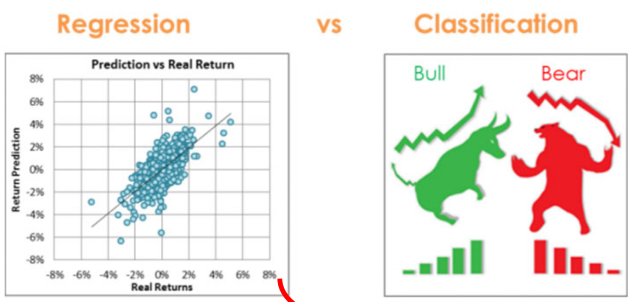
- Inductive learning: *learn by example* and *induce a general rule* from a set of examples/observations
- The *type of feedback* available for learning determines the *nature of the learning problem*

Unsupervised learning





Supervised learning



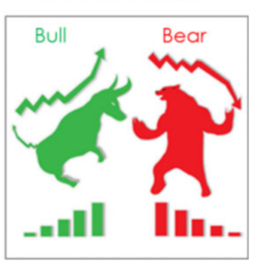
+ feature analysis

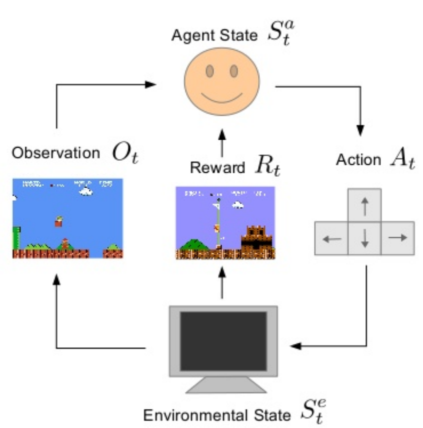
Types of inductive learning

- Inductive learning: *learn by example* and *induce a general rule* from a set of examples/observations
- The *type of feedback* available for learning determines the *nature of the learning problem*

Supervised learning

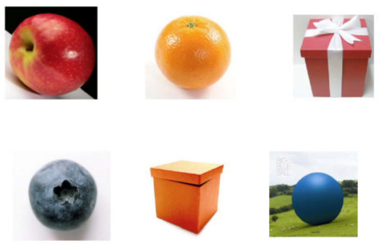
Regression vs Classification

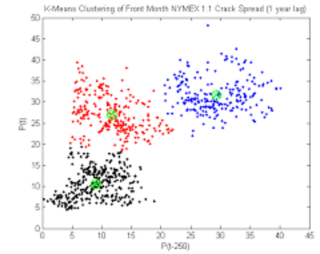




Reinforcement learning

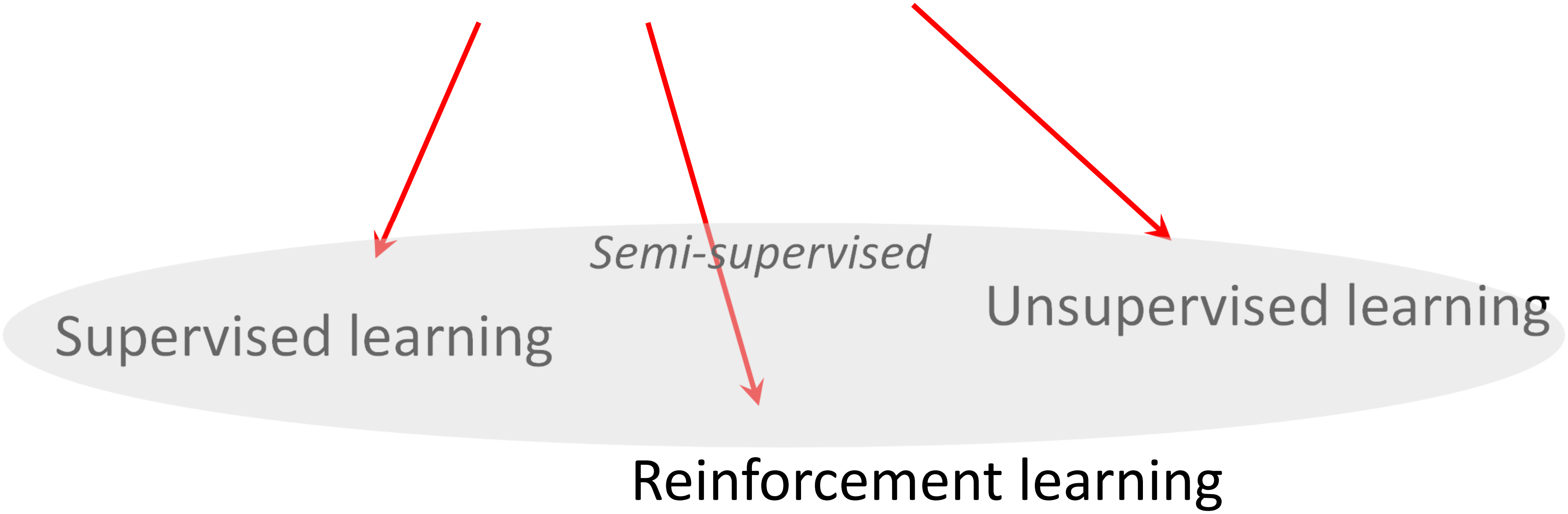
Unsupervised learning





Types of inductive learning: summary

- Inductive learning: *learn by example* and *induce a general rule* from a set of examples/observations
- The *type of feedback* available for learning determines the *nature of the learning problem*



An ML model

- What is an ML model?

$D = \{\mathbf{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots, \boldsymbol{X}^{(m)}]$

$\boldsymbol{X} = [x_1, x_2, \ldots, x_N] \longrightarrow$ Hypothesis, h $\longrightarrow h(\boldsymbol{X})$

$h \in H$

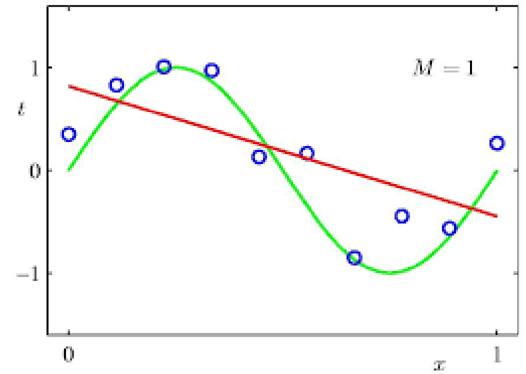
- Model as an instance of a hypothesis (inductive learning perspective)
  - hypothesis about the underlying relationship, function behind the data (observations)
  - hypothesis -> a model family

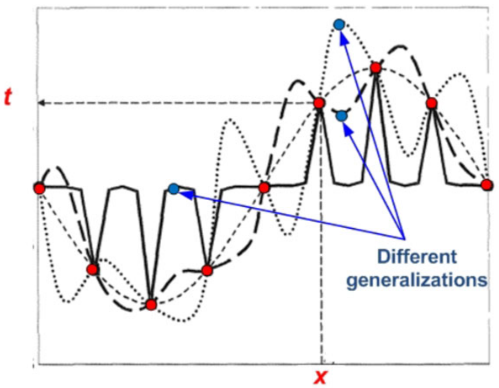
A general notion of a supervised learning problem

$$f : \mathbf{X} \to T$$

$$T = f(\mathbf{X}) + \varepsilon \qquad (\mathbf{X}, T) : \{(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), \ldots, (\mathbf{x}_N, t_N),$$

$$\mathcal{H} = \{h : \mathbf{X} \to T\}$$



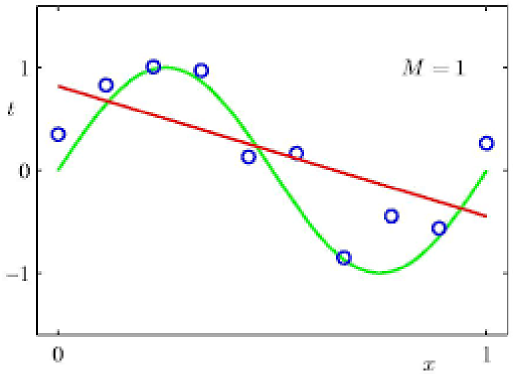


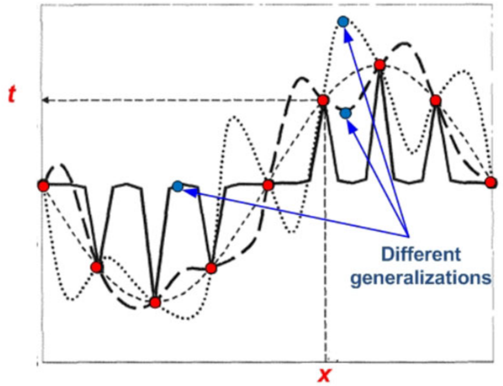
A general notion of a supervised learning problem

$$f : \mathbf{X} \rightarrow T$$

$$T = f(\mathbf{X}) + \varepsilon \qquad (\mathbf{X}, T) : \{(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), \ldots, (\mathbf{x}_N, t_N),$$

$$\mathcal{H} = \{h : \mathbf{X} \rightarrow T\}$$





Fundamental assumptions

- Pattern exists.
- We have no underlying mathematical model / explicit problem formulation.
- There is data.

An ML model

- What makes models different?
  - Underlying assumptions behind different model families
    - parameterisation of a problem
    - parametric vs non-parametric models
    - computational capacity ("size" of the model, hyperparameters)

An ML model

- What makes models different?
  - Underlying assumptions behind different model families
    - parameterisation of a problem
    - parametric vs non-parametric models
    - computational capacity (“size” of the model, hyperparameters)
  - Loss function, metric of a particular model fit
  - Optimisation: how we tune parameters to minimise the loss

ML models make predictions

- It is not enough to describe the data
  - probabilistic description
  - black-box description (transparency issues)
  - parametric vs non-parametric
- Generalisation as a holy grail of ML's predictive power
  - capability to make good predictions on unseen data
  - fundamental assumption: training (current) and unseen (future) data come from the same distribution

ML models make predictions

- How to make models that produce ”good predictions”?
  - Inductive learning cycle:

    i) observe a phenomenon, collect data (*observations, samples*)

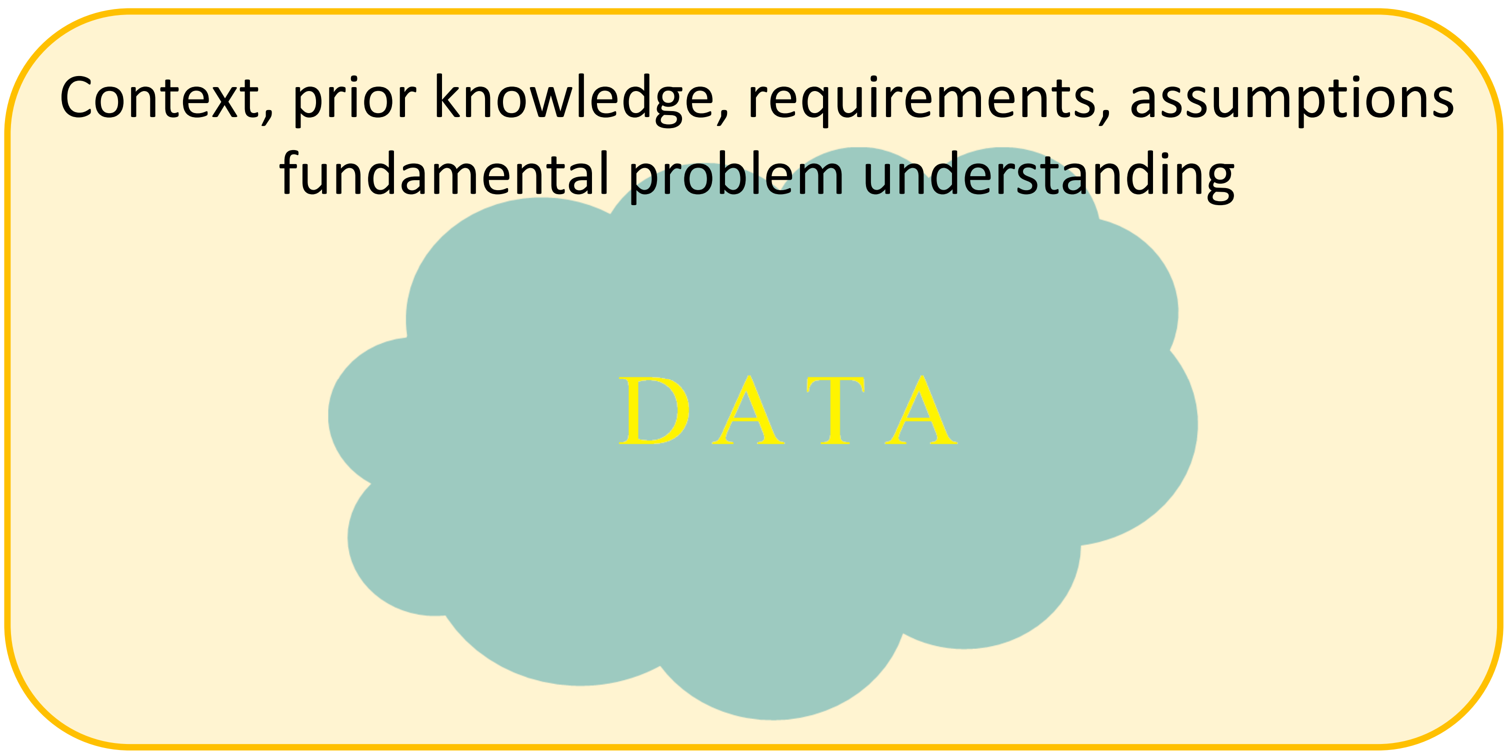
    ii) create a model of that phenomenon (*fit* and *validate* the model)

    iii) deploy the model and make *predictions* (ultimate test of the quality)

ML models make predictions

- How to make models that produce ”good predictions”?
  - Inductive learning cycle:

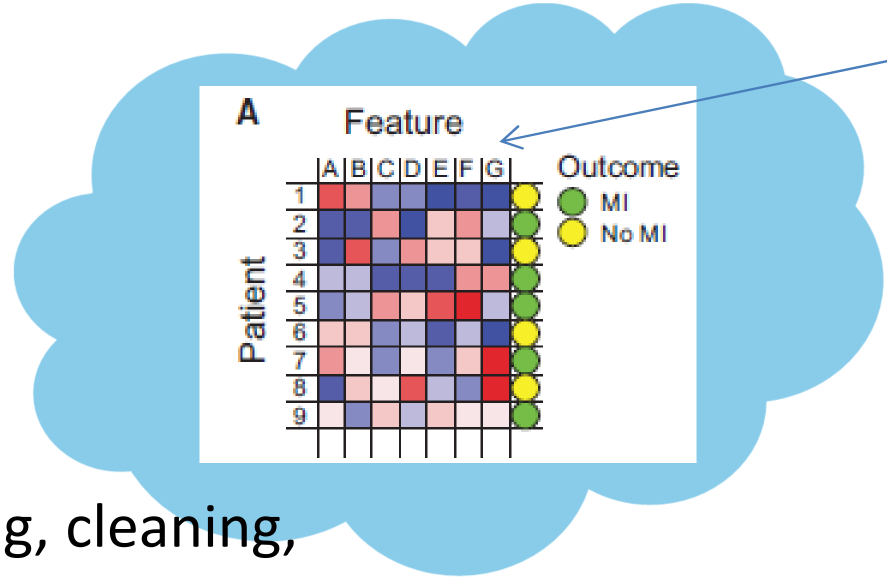
    i) observe a phenomenon, collect data (observations)



ML models make predictions

- How to make models that produce ”good predictions”?
    - Inductive learning cycle:

      i) observe a phenomenon, collect data (observations)



attributes
e.g. blood pressure, glucose level, age

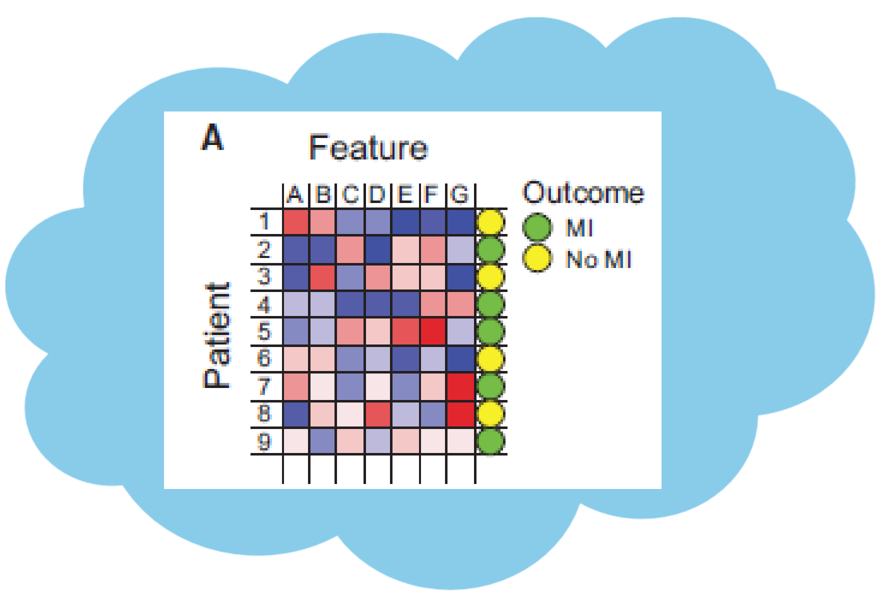
Data preprocessing, cleaning, feature / attribute analysis etc.

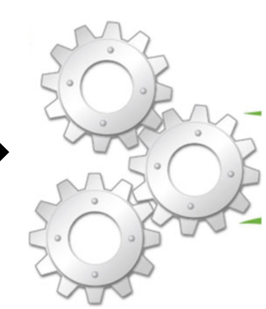
ML models make predictions

- How to make models that produce ”good predictions”?
  - Inductive learning cycle:

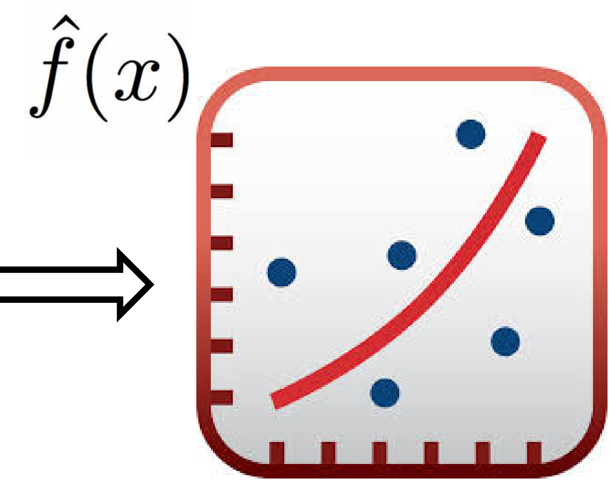
    i) observe a phenomenon, collect data (observations)

    ii) create a model of that phenomenon (fit and validate the model)





ML algorithm



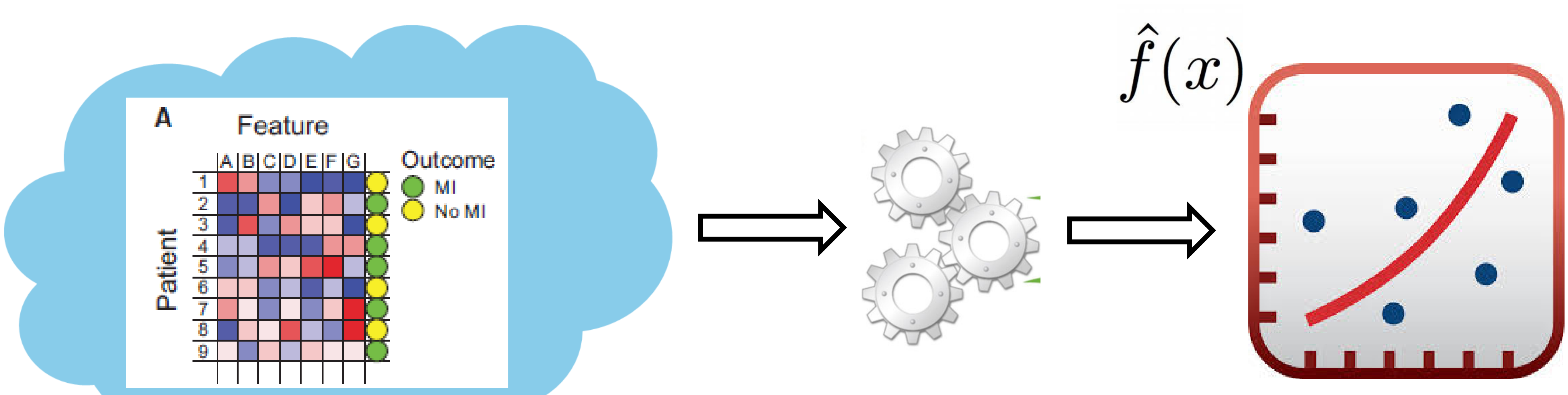
Model derived from data using the **ML algorithm**

ML models make predictions

- How to make models that produce ”good predictions”?
  - Inductive learning cycle:

    i) observe a phenomenon, collect data (observations)

    ii) create a model of that phenomenon (fit and validate the model)



attributes (e.g. in EHR record) → outcomes / labels (e.g., disease state)

ML models make predictions

- How to make models that produce ”good predictions”?
  - Inductive learning cycle:

    i) observe a phenomenon, collect data (observations)

    ii) create a model of that phenomenon (fit and validate the model)

    iii) deploy the model and make predictions (ultimate test of the quality)

new (unseen) data
(e.g. new patient’s EHR record) $\xrightarrow[\hat{f}(x)]{}$ predicted outcome
(e.g., disease state)

supervised learning context

ML models make predictions

- How to make models that produce ”good predictions”?
  - Inductive learning cycle:

    i) observe a phenomenon, collect data (observations)

    ii) create a model of that phenomenon (fit and validate the model)

    iii) deploy the model and make predictions (ultimate test of the quality)

new (unseen) data
(e.g. new patient’s EHR record)

$$\xrightarrow[\hat{f}(x)]{}$$

finding similar patients / data points

detecting anomalies

unsupervised learning context

ML models make predictions

- How to make models that produce ”good predictions”?
  - Inductive learning cycle:

    i) observe a phenomenon, collect data (observations)

    ii) create a model of that phenomenon (fit and validate the model)

    iii) deploy the model and make predictions (ultimate test of the quality)
  - *“Good” models* – what does it mean?

ML models make predictions

- How to make models that produce ”good predictions”?
  - Inductive learning cycle:

    i) observe a phenomenon, collect data (observations)

    ii) create a model of that phenomenon (fit and validate the model)

    iii) deploy the model and make predictions (ultimate test of the quality)

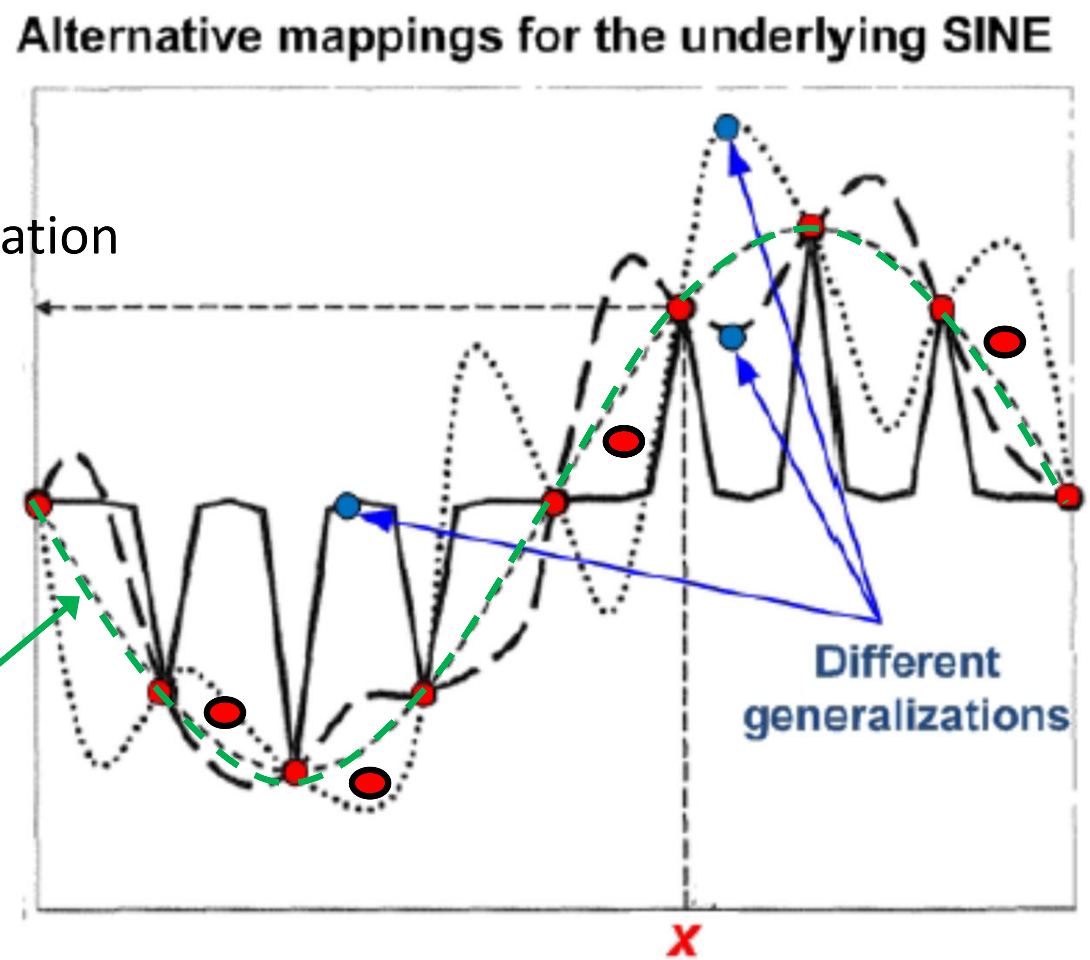
  - *“Good” models* with desirable predictive power (*generalising* well)
    - task-relevant performance criteria and the “operational” loss function
    - simplicity (Occam’s razor, “A model should be as simple as it can be but no simpler”)
    - computational cost (typically associated with model fitting)

Generalisation as the central concept

Curve fitting analogy

- inferring hypothesis about the underlying model
- data and model determine the generalisation power
- there are more and less generalisable models

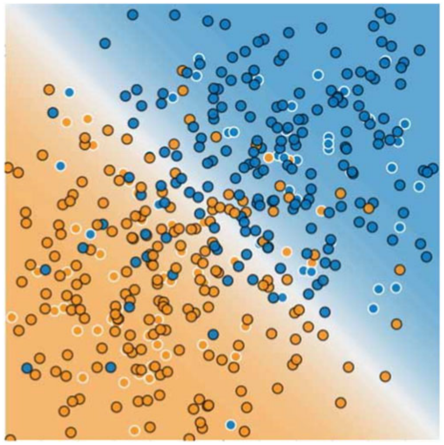
observation



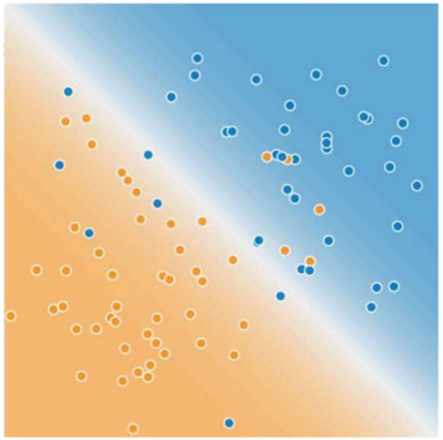
$f(x) = \sin(x)$
underlying mapping
(unknown to the observer)

ML model and generalisation

- What can we do to maximise chances for good generalisation power of our model in the given problem (data)?
  - assumptions about data



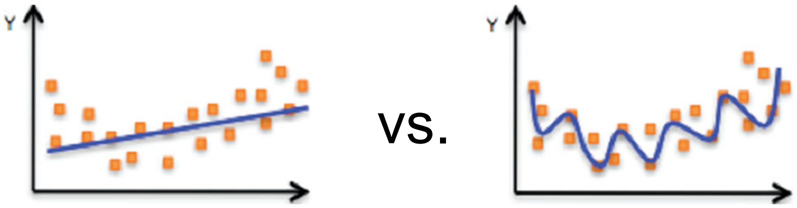
Training Data



Test Data

ML model and generalisation

- What can we do to maximise chances for good generalisation power of our model in the given problem (data)?
  - assumptions about data
  - tendency towards simplicity (Occam's razor)



ML model and generalisation

- What can we do to maximise chances for good generalisation power of our model in the given problem (data)?
  - assumptions about data
  - tendency towards simplicity (Occam's razor)
  - model selection as a good balance between the model's capacity (aka complexity) and simplicity of the hypothesis explaining the data

ML model and generalisation

- What can we do to maximise chances for good generalisation power of our model in the given problem (data)?
  - assumptions about data
  - tendency towards simplicity (Occam's razor)
  - model selection as a good balance between the model's capacity (aka complexity) and simplicity of the hypothesis explaining the data
  - how can we "measure" generalisation to strike the balance?

ML model and generalisation

- What can we do to maximise chances for good generalisation power of our model in the given problem (data)?
  - assumptions about data
  - tendency towards simplicity (Occam's razor)
  - model selection as a good balance between the model's capacity (aka complexity) and simplicity of the hypothesis explaining the data
  - how can we "measure" generalisation to strike the balance?

$$\mathbf{X} \rightarrow T: \; T = f(\mathbf{X}) + \varepsilon, \qquad \mathbb{E}\left[\varepsilon f(\mathbf{X})\right] = 0$$

ML model and generalisation

- What can we do to maximise chances for good generalisation power of our model in the given problem (data)?
  - assumptions about data
  - tendency towards simplicity (Occam's razor)
  - model selection as a good balance between the model's capacity (aka complexity) and simplicity of the hypothesis explaining the data
  - how can we "measure" generalisation to strike the balance?

$$\mathbf{X} \to T: \; T = f(\mathbf{X}) + \varepsilon, \qquad \mathbb{E}\left[\varepsilon f(\mathbf{X})\right] = 0$$

We approach the problem of finding *f* with an ML model, $F(\boldsymbol{x},\boldsymbol{w})$, and define the L2-norm risk, $R[F]$:

our hypothesis about f

$$R\left[F\right] = \mathbb{E}_D\left[\left(t - F\left(\boldsymbol{x},\boldsymbol{w}\right)\right)^2\right]$$

ML model and generalisation

- What can we do to maximise chances for good generalisation power of our model in the given problem (data)?
  - assumptions about data
  - tendency towards simplicity (Occam's razor)
  - model selection as a good balance between the model's capacity (aka complexity) and simplicity of the hypothesis explaining the data
  - how can we "measure" generalisation to strike the balance?

$$\mathbf{X} \to T: \; T = f(\mathbf{X}) + \varepsilon, \qquad \mathbb{E}\left[\varepsilon f(\mathbf{X})\right] = 0$$

$$\underbrace{Err_{Dtrain}[F] = \frac{1}{N}\sum_{i=1}^{N}\left[\left(t_i - F(\boldsymbol{x}_i, \boldsymbol{w})\right)^2\right]}_{\textit{training error}} \neq \mathbb{E}_D\left[\left(t - F(\boldsymbol{x}, \boldsymbol{w})\right)^2\right]$$

does not measure *generalisation*

Generalisation: bias-variance dilemma

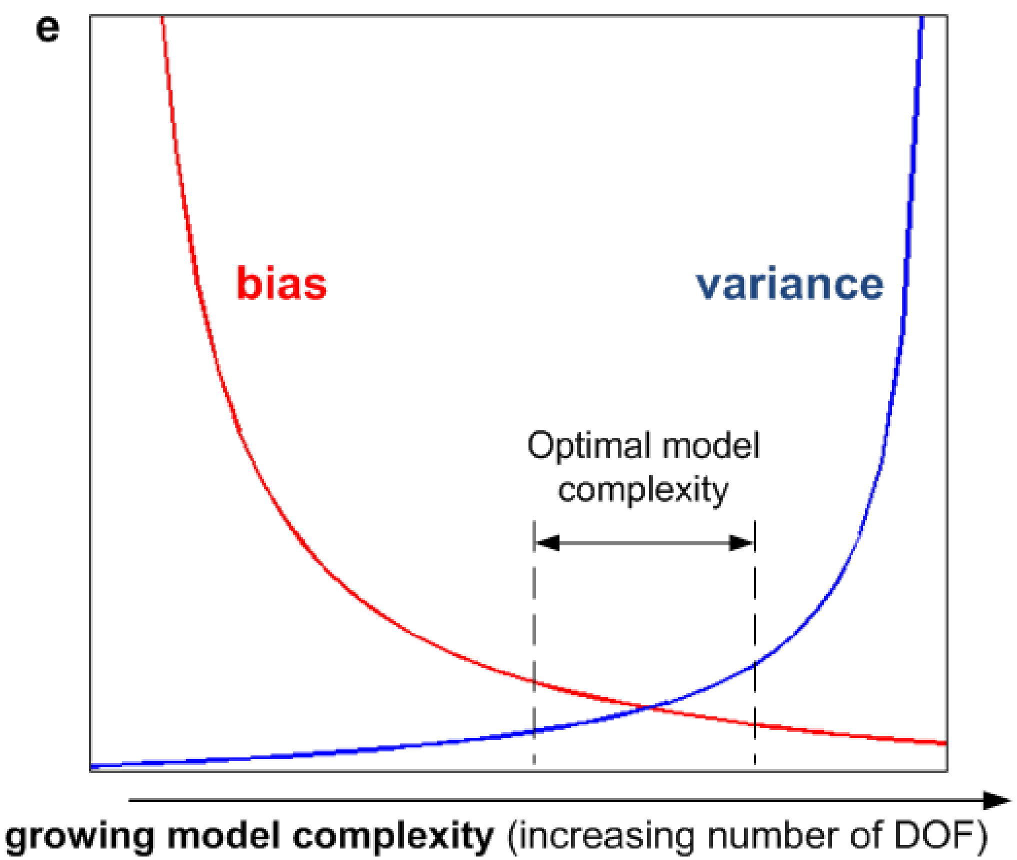
Generalisation error

$$\mathbf{E}_D\left[\left(t-F(\boldsymbol{x},\boldsymbol{w})\right)^2\right]=\mathbf{E}_D\left[\left(t-f(\boldsymbol{x})\right)^2\right]+\mathbf{E}_D\left[\left(F(\boldsymbol{x},\boldsymbol{w})-f(\boldsymbol{x})\right)^2\right]=\dots$$

$$\dots=\underbrace{\mathbf{E}_D\left[\left(t-f(\boldsymbol{x})\right)^2\right]}_{\text{irreducible error}}+\underbrace{\left(\mathbf{E}_D\left[F(\boldsymbol{x},\boldsymbol{w})\right]-f(\boldsymbol{x})\right)^2}_{\text{bias}}+\underbrace{\mathbf{E}_D\left[\left(F(\boldsymbol{x},\boldsymbol{w})-\mathbf{E}_D\left[F(\boldsymbol{x},\boldsymbol{w})\right]\right)^2\right]}_{\text{variance}}$$

The need to balance bias (*approximation* error) and variance (*estimation* error) on a limited data sample

The need to balance bias (*approximation* error) and variance (*estimation* error) on a limited data sample

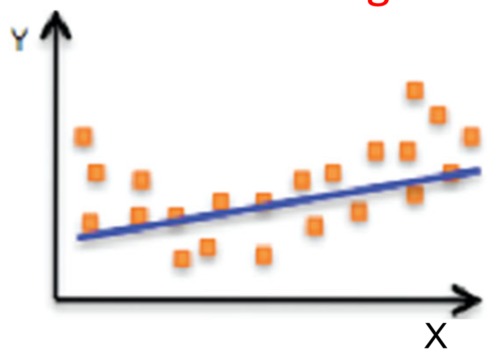


- The problem can be alleviated by large data amount
- Balance boils down to selecting optimal network *capacity*
- Representational vs *effective* capacity
- The concept of Occam's razor

Generalisation: overfitting vs. underfitting

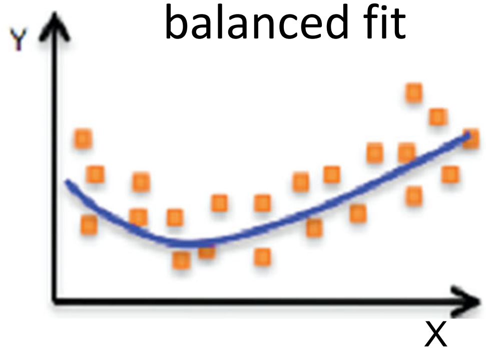
Underfitting vs *overfitting*

underfitting

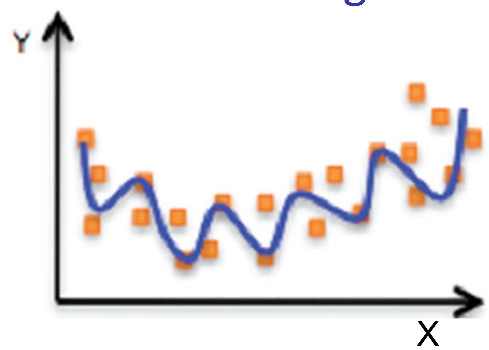


model is too simple,
low variance, high bias

suitable,
balanced fit



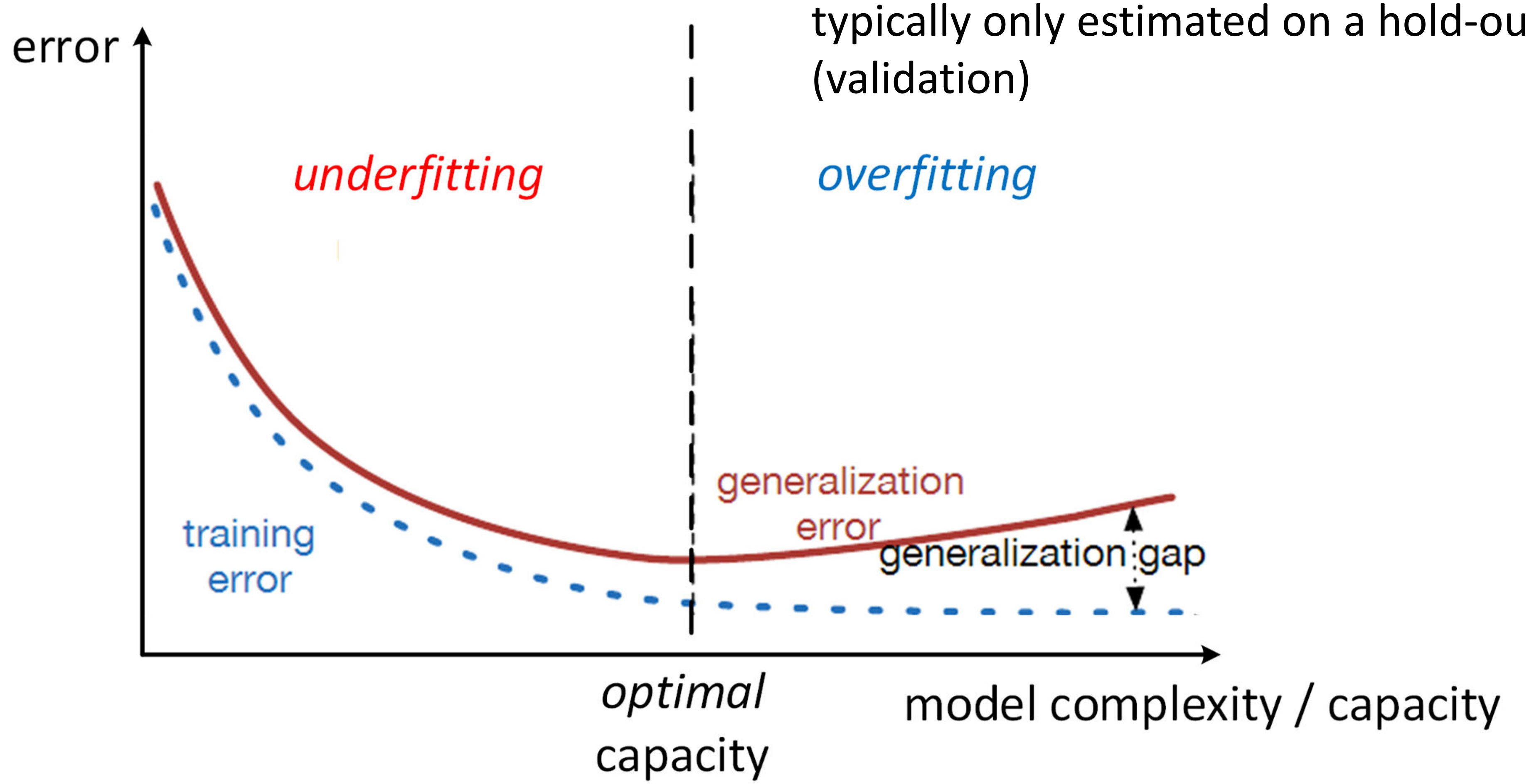
overfitting



model is too complex,
high variance, low bias

Generalisation gap

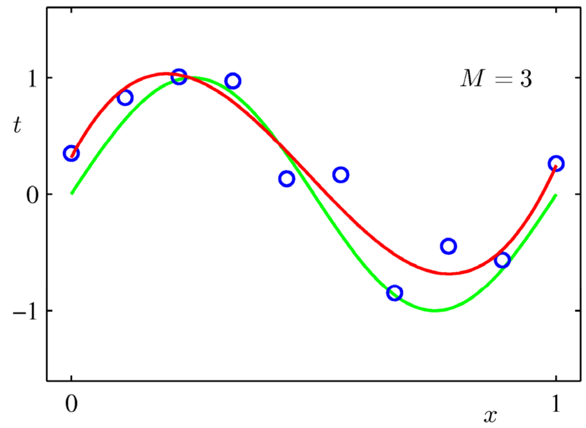
***generalisation** error* – defined in statistical sense, typically only estimated on a hold-out set (validation)

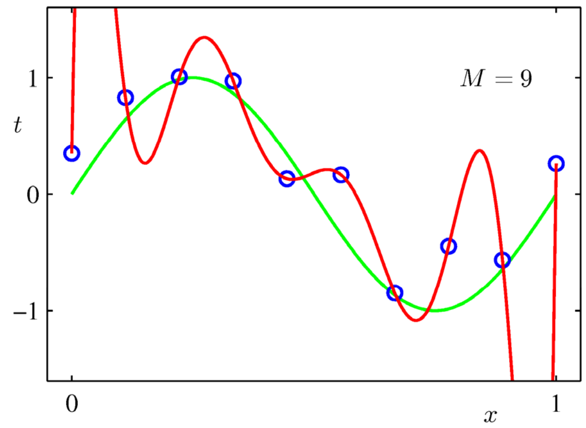


***training** error* = empirical error reported on a training set

Model fitting

- It is not the same as *model selection* (model family)



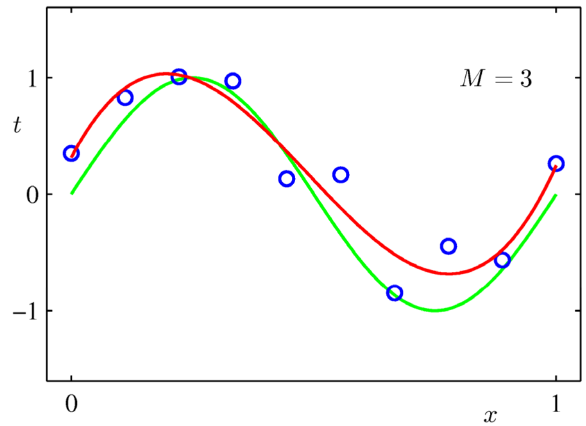


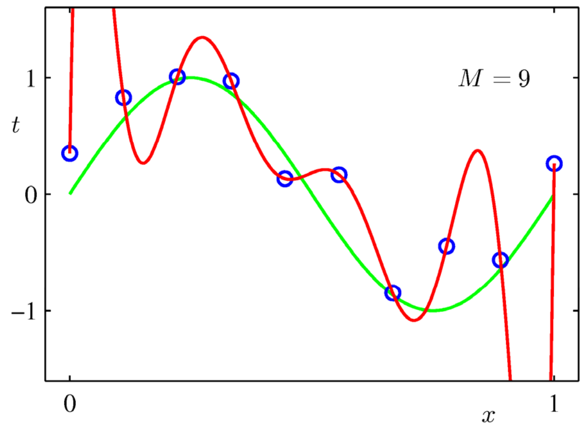
Linear basis function model

$$F(x, \mathbf{w}, M) = \sum_{j=0}^{M} w_j \phi_j(x) = \mathbf{w}^{\mathrm{T}} \phi(x)$$

Model fitting

- It is not the same as *model selection* (model family)





model selection

Linear basis function model

$$F(x,\mathbf{w},M) = \sum_{j=0}^{M} w_j \phi_j(x) = \mathbf{w}^{\mathrm{T}} \phi(x)$$

Model fitting

- It is not the same as *model selection* (model family)

model selection

Linear basis function model

$$F(x, \mathbf{w}, M) = \sum_{j=0}^{M} w_j \phi_j(x) = \mathbf{w}^{\mathrm{T}} \phi(x)$$

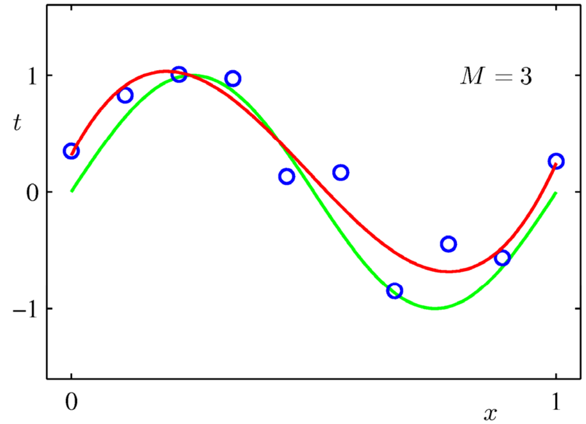
Polynomial model $\phi_j(x) = x^j$

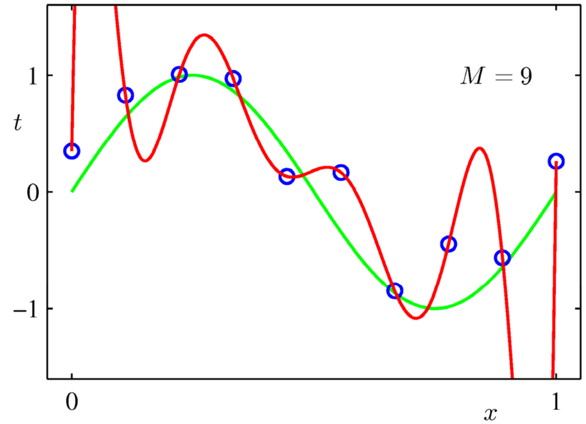
$$F(x, \mathbf{w}, M) = \sum_{j=0}^{M} w_j x^j = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M$$

hyperparameter

Model fitting

- It is not the same as *model selection* (model family)





model selection

Linear basis function model

$$F(x,\mathbf{w},M) = \sum_{j=0}^{M} w_j \phi_j(x) = \mathbf{w}^{\mathrm{T}}\phi(x)$$

Polynomial model $\phi_j(x) = x^j$

tunable, trainable parameters

$$F(x,\mathbf{w},M) = \sum_{j=0}^{M} w_j x^j = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M$$

hyperparameter

Fitting the model

1. Define model fitting criterion (loss function, goodness of fit)
2. Select a model (model family, hyperparameters) – No Free Lunch
3. Choose and apply **(training)** an optimisation method to minimise the loss by changing/adapting/tuning adjustable/trainable parameters:

$$F(x, \mathbf{w}, M) = \sum_{j=0}^{M} w_j x^j$$

Fitting the model

1. Define model fitting criterion (loss function, goodness of fit)
2. Select a model (model family, hyperparameters) – No Free Lunch
3. Choose and apply **(training)** an optimisation method to minimise the loss by changing/adapting/tuning adjustable/trainable parameters:

$$F(x, \mathbf{w}, M) = \sum_{j=0}^{M} w_j x^j$$

OPTIMISATION is an important tool in ML

Fitting the model

1. Define model fitting criterion (loss function, goodness of fit)
2. Select a model (model family, hyperparameters) – No Free Lunch
3. Choose and apply **(training)** an optimisation method to minimise the loss by changing/adapting/tuning adjustable/trainable parameters:

$$F(x, \mathbf{w}, M) = \sum_{j=0}^{M} w_j x^j$$

4. Validate the model (perhaps go back to p.2 and 3)

Fitting the model

1. Define model fitting criterion (loss function, goodness of fit)
2. Select a model (model family, hyperparameters) – No Free Lunch
3. Choose and apply **(training)** an optimisation method to minimise the loss by changing/adapting/tuning adjustable/trainable parameters:

$$F(x, \mathbf{w}, M) = \sum_{j=0}^{M} w_j x^j$$

4. Validate the model (perhaps go back to p.2 and 3)

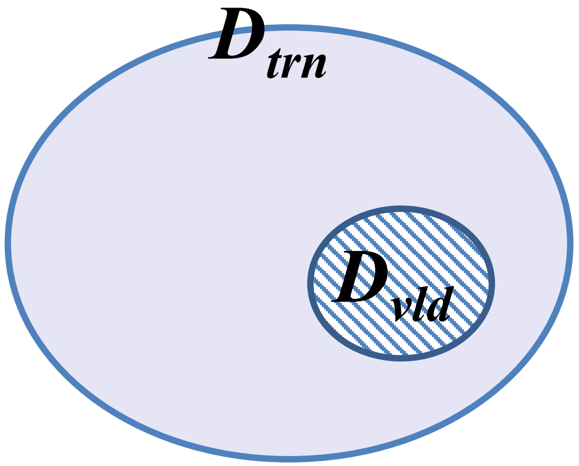
BUT: how do we decide if we are happy enough with the model fit?

Empirical approach to "measuring" generalisation

Purpose of empirical assessment of generalization capabilities

i. To validate the model's "quality" (generalisation capability)

ii. As a criterion for model selection (comparison)



Estimate based on a **hold-out set**

- Separating *training* from ***validation*** data (different from *unseen test* data)
- Different ways of *resampling* data to form a hold-out set
- Costly to sacrifice data for validation if there is little data for training

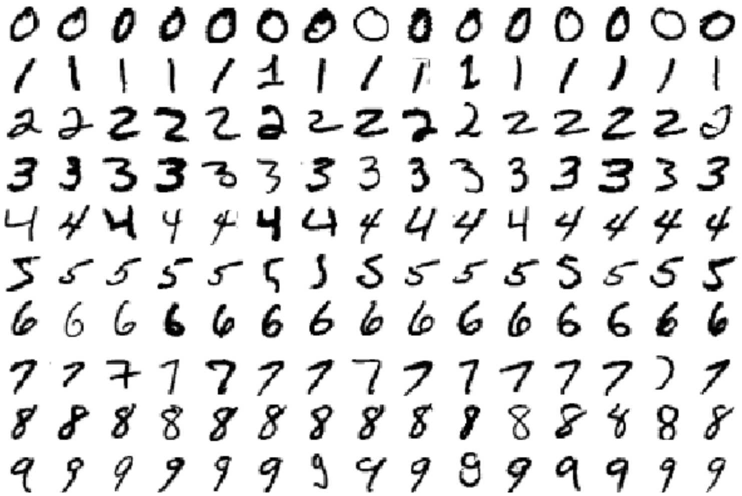
Fitting the model

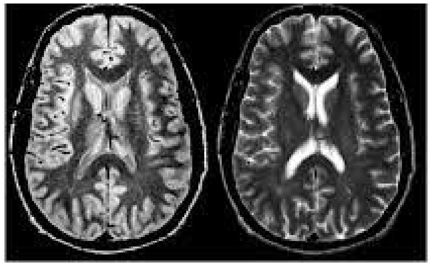
1. Define model fitting criterion (loss function, goodness of fit)
2. Select a model (model family, hyperparameters) – No Free Lunch
3. Choose and apply **(training)** an optimisation method to minimise the loss by changing/adapting/tuning adjustable/trainable parameters:

$$F(x, \mathbf{w}, M) = \sum_{j=0}^{M} w_j x^j$$

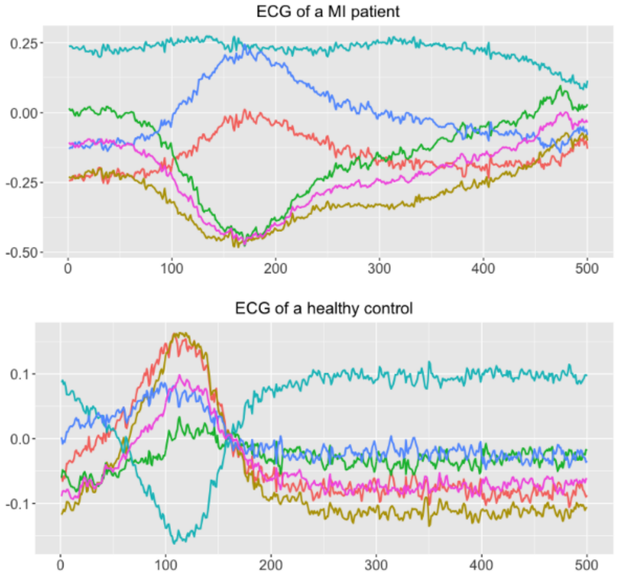
4. Validate the model (perhaps go back to p.2 and 3)
5. Deploy the model – use for testing and making novel predictions

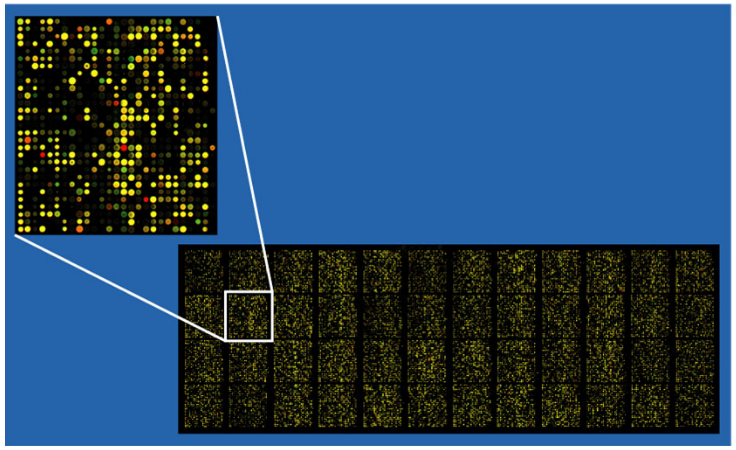
Different data types






```
<REUTERS TOPICS=''YES'' LEWISSPLIT=''TRAIN''
CGISPLIT=''TRAINING-SET'' OLDID=''12981'' NEWID=''798''>
<DATE> 2-MAR-1987 16:51:43.42</DATE>
<TOPICS><D>livestock</D><D>hog</D></TOPICS>
<TITLE>AMERICAN PORK CONGRESS KICKS OFF TOMORROW</TITLE>
<DATELINE> CHICAGO, March 2 - </DATELINE><BODY>The American Pork
Congress kicks off tomorrow, March 3, in Indianapolis with 160
of the nations pork producers from 44 member states determining
industry positions on a number of issues, according to the
National Pork Producers Council, NPPC.
Delegates to the three day Congress will be considering 26
resolutions concerning various issues, including the future
direction of farm policy and the tax law as it applies to the
agriculture sector. The delegates will also debate whether to
endorse concepts of a national PRV (pseudorabies virus) control
and eradication program, the NPPC said. A large
trade show, in conjunction with the congress, will feature
the latest in technology in all areas of the industry, the NPPC
added. Reuter
\&\#3;</BODY></TEXT></REUTERS>
```

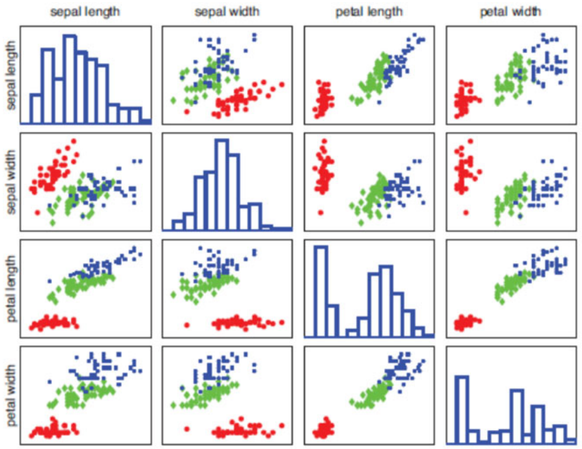




Data formats, types etc.

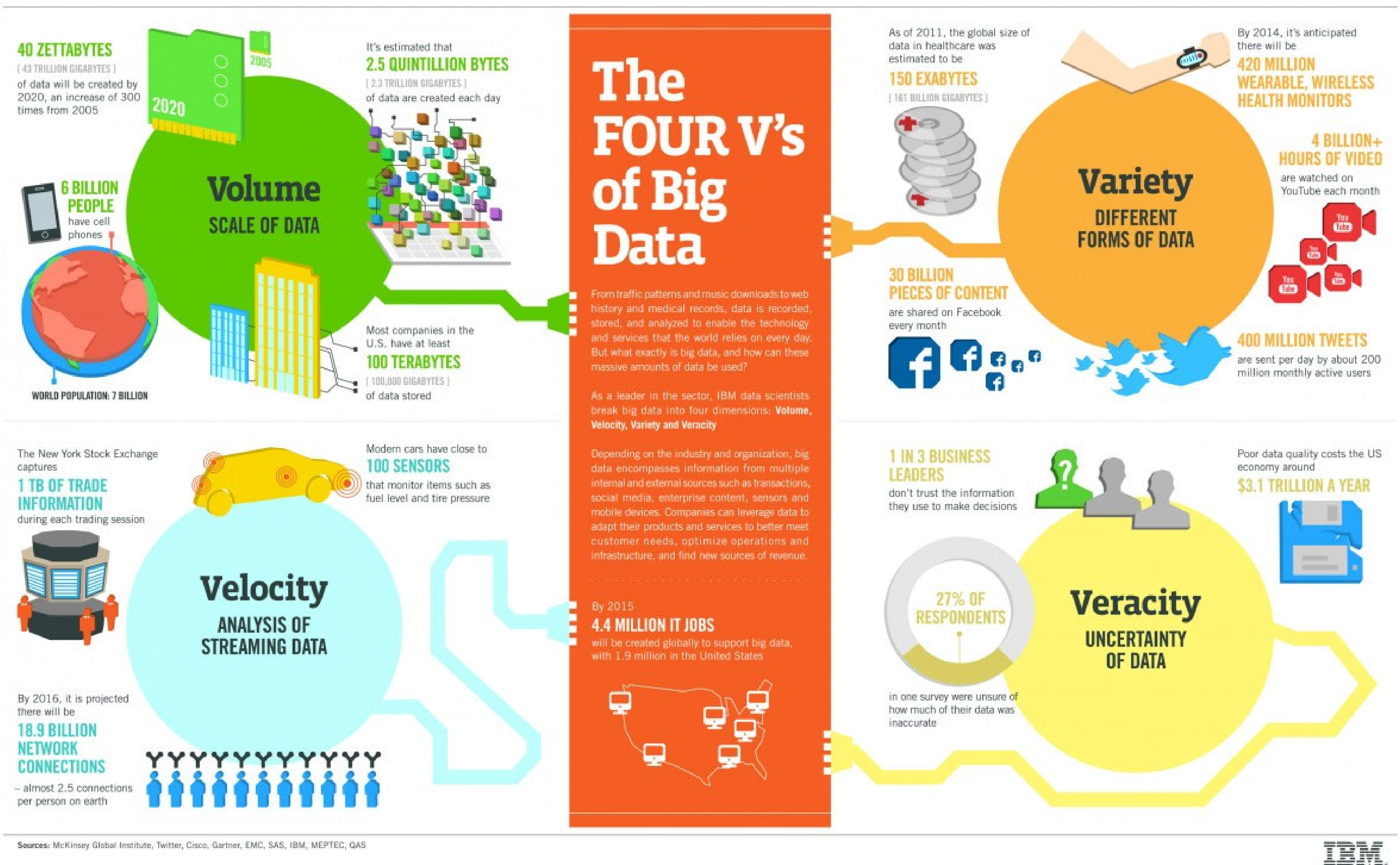
- Vectors (the most common): set of features, coordinates, sensor readouts (incl. temporal data)



- Matrices (also very common): images, spectral data
- Strings: documents, gene sequences etc.
- Other structured data types: graphs, trees, XML documents

Four Vs of **Big Data** (IBM perspective)



https://www.ibm.com/blog/ https://opensistemas.com/en/the-four-vs-of-big-data/

Other important general considerations

- Ethical and sustainability aspects
- Privacy issues
- Transparency, explainability
- Trustworthiness, reliability
- Narrow AI vs General AI (or AGI)

On the subject of explainability, interpretable ML