Fundaments of Machine learning for and with engineering applications

Enrico Riccardi[1]

Department of Mathematics and Physics, University of Stavanger (UiS).[1]

Aug 28, 2025

1. Personal Intro
2. Intro to MOD550
3. Modeling
4. Recaps: Data concepts, Data types and Models
5. Descriptive and Predictive statistics
6. Simulations
7. Statistics, Machine learning or Artificial intelligence?
8. Metadata

Let me first introduce US

Enrico Riccardi
email: enrico.riccardi@uis.no

Pål Østebø Andersen
email: pal.andersen@uis.no

Let me first introduce myself

Enrico Riccardi email: enrico.riccardi@uis.no office: KE-E-542

Before to be here...

- Chemical engineer graduated from the Politecnico di Torino
- PhD at M&ST university of Rolla, Missouri, USA
- Post Doc at TUD Darmstadt, Germany
- Researcher at NTNU, Trondheim, Norway
- Post Doc UiO, Oslo, Norway
- Assoc. Prof. at Uis, Stavanger, Norway
- Prof. at Uis, Stavanger, Norway from 28.8.2025!!

Let me first introduce myself

Enrico Riccardi email: enrico.riccardi@uis.no office: KE-E-542

Before to be here...

- Chemical engineer graduated from the Politecnico di Torino
- PhD at M&ST university of Rolla, Missouri, USA
- Post Doc at TUD Darmstadt, Germany
- Researcher at NTNU, Trondheim, Norway
- Post Doc UiO, Oslo, Norway
- Assoc. Prof. at Uis, Stavanger, Norway
- Prof. at Uis, Stavanger, Norway from 28.8.2025!!

Course objectives

Understanding of

- data sources and consequent data properties.
- data analysis/machine learning approaches outcomes.
- sensitivity analysis
- predictive modeling
- multivariate data analysis
- machine learning techniques application
- ensemble methods
- Bayes approach
- visualization and reporting

Course objectives

Understanding of

- data sources and consequent data properties.
- data analysis/machine learning approaches outcomes.
- sensitivity analysis
- predictive modeling
- multivariate data analysis
- machine learning techniques application
- ensemble methods
- Bayes approach
- visualization and reporting

Metalearning

- Pedagogic empiricism teaches us that good learning happens in social contexts where the students are actively participating in the classroom dialogue
- In general, the number of questions, and the dialogue in general, drops significantly when lectures are recorded
- Streaming tends to lead to empty classrooms and higher failure rate.

Metalearning

- Pedagogic empiricism teaches us that good learning happens in social contexts where the students are actively participating in the classroom dialogue
- In general, the number of questions, and the dialogue in general, drops significantly when lectures are recorded
- Streaming tends to lead to empty classrooms and higher failure rate.

Metalearning

- Pedagogic empiricism teaches us that good learning happens in social contexts where the students are actively participating in the classroom dialogue
- In general, the number of questions, and the dialogue in general, drops significantly when lectures are recorded
- Streaming tends to lead to empty classrooms and higher failure rate.

Metalearning

- Pedagogic empiricism teaches us that good learning happens in social contexts where the students are actively participating in the classroom dialogue

- In general, the number of questions, and the dialogue in general, drops significantly when lectures are recorded

- Streaming tends to lead to empty classrooms and higher failure rate.

Teaching method

!Active Learning!

- There will be a combination of lectures and tutorials.
- Tutorial and hands-on will be presented during the course.
- Study groups are strongly encouraged.
- In class discussions are encouraged at any stage.
- Flip classroom approach: problem first (when possible).
- Feedback expected from you!
- Note: each teacher, even if coordinated, will have a different approach/expectations.

Teaching method

!Active Learning!

- There will be a combination of lectures and tutorials.
- Tutorial and hands-on will be presented during the course.
- Study groups are strongly encouraged.
- In class discussions are encouraged at any stage.
- Flip classroom approach: problem first (when possible).
- Feedback expected from you!
- Note: each teacher, even if coordinated, will have a different approach/expectations.

Teaching method

!Active Learning!

- There will be a combination of lectures and tutorials.
- Tutorial and hands-on will be presented during the course.
- Study groups are strongly encouraged.
- In class discussions are encouraged at any stage.
- Flip classroom approach: problem first (when possible).
- Feedback expected from you!
- Note: each teacher, even if coordinated, will have a different approach/expectations.

Teaching method

!Active Learning!

- There will be a combination of lectures and tutorials.
- Tutorial and hands-on will be presented during the course.
- Study groups are strongly encouraged.
- In class discussions are encouraged at any stage.
- Flip classroom approach: problem first (when possible).
- Feedback expected from you!
- Note: each teacher, even if coordinated, will have a different approach/expectations.

Teaching method

!Active Learning!

- There will be a combination of lectures and tutorials.
- Tutorial and hands-on will be presented during the course.
- Study groups are strongly encouraged.
- In class discussions are encouraged at any stage.
- Flip classroom approach: problem first (when possible).
- Feedback expected from you!
- Note: each teacher, even if coordinated, will have a different approach/expectations.

Teaching method

!Active Learning!

- There will be a combination of lectures and tutorials.
- Tutorial and hands-on will be presented during the course.
- Study groups are strongly encouraged.
- In class discussions are encouraged at any stage.
- Flip classroom approach: problem first (when possible).
- Feedback expected from you!
- Note: each teacher, even if coordinated, will have a different approach/expectations.

Teaching method

!Active Learning!

- There will be a combination of lectures and tutorials.
- Tutorial and hands-on will be presented during the course.
- Study groups are strongly encouraged.
- In class discussions are encouraged at any stage.
- Flip classroom approach: problem first (when possible).
- Feedback expected from you!
- Note: each teacher, even if coordinated, will have a different approach/expectations.

Teaching method

!Active Learning!

- There will be a combination of lectures and tutorials.
- Tutorial and hands-on will be presented during the course.
- Study groups are strongly encouraged.
- In class discussions are encouraged at any stage.
- Flip classroom approach: problem first (when possible).
- Feedback expected from you!
- Note: each teacher, even if coordinated, will have a different approach/expectations.

Python

The tutorials and hands-on will be mostly based on python.

- An introduction to python will be provided during the course. Yet, this is not a Python programming course!
- You are **strogly advised** to construct your own Git repository (and to keep it in order).
- Group projects will have different roles.

Python

The tutorials and hands-on will be mostly based on python.

- An introduction to python will be provided during the course. Yet, this is not a Python programming course!
- You are **strogly advised** to construct your own Git repository (and to keep it in order).
- Group projects will have different roles.

Python

The tutorials and hands-on will be mostly based on python.

- An introduction to python will be provided during the course. Yet, this is not a Python programming course!
- You are **strogly advised** to construct your own Git repository (and to keep it in order).
- Group projects will have different roles.

Python

The tutorials and hands-on will be mostly based on python.

- An introduction to python will be provided during the course. Yet, this is not a Python programming course!
- You are **strogly advised** to construct your own Git repository (and to keep it in order).
- Group projects will have different roles.

A personal considerations on generative AI

With great power comes great responsibility (Spider-Man)

LLMs can and shall be used. As their development is surging in the last years, their help in writing code and reports is undeniable. Students shall learn how to master these tools. Yet, while their usage is encouraged, the risk of excessively rely on them **DO** hinder learning.

If you want to be a student (i.e. whom is learning), you are encouraged to take the responsibility in delivering original output.

A personal considerations on generative AI

> With great power comes great responsibility (Spider-Man)

LLMs can and shall be used. As their development is surging in the last years, their help in writing code and reports is undeniable. Students shall learn how to master these tools. Yet, while their usage is encouraged, the risk of excessively rely on them **DO** hinder learning.

If you want to be a student (i.e. whom is learning), you are encouraged to take the responsibility in delivering original output.

1. Personal Intro
2. Intro to MOD550
3. Modeling
4. Recaps: Data concepts, Data types and Models
5. Descriptive and Predictive statistics
6. Simulations
7. Statistics, Machine learning or Artificial intelligence?
8. Metadata

Why are we here? Why this course?

- Decisions are the hinge. What influences your decisions has value.

- Better results come from better decisions.

STATISTICAL argument

- Machine learning can be a useful ONLY when it helps making better decisions.

Why are we here? Why this course?

- Decisions are the hinge. What influences your decisions has value.

- Better results come from better decisions.

STATISTICAL argument

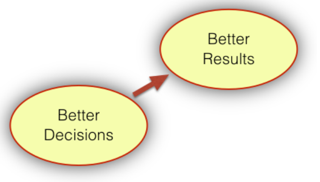
- Machine learning can be a useful ONLY when it helps making better decisions.

Decisions, decisions, decisions...

Life lesson?!

Life is a sum of all your choices (Albert Camus)

The only way you can purposefully influence your life, your family, your organization, your country or your world is through the decisions you make.

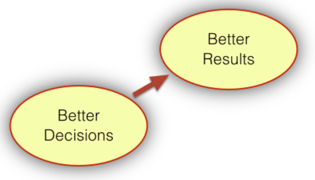


Decisions, decisions, decisions...

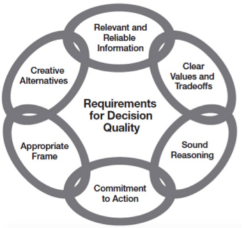
Life lesson?!

Life is a sum of all your choices (Albert Camus)

The only way you can purposefully influence your life, your family, your organization, your country or your world is through the decisions you make.



Uncertainty is fed to decisions

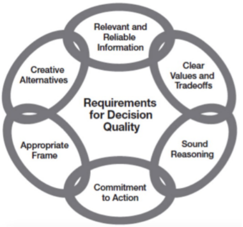


- In this course we will learn to use data and models (ML) to make better decisions.
- Each application will be a mere exercise of the concepts we will introduce here.

Do not try this at home!

- Please Do not use a technical approach for love matters.

Uncertainty is fed to decisions



- In this course we will learn to use data and models (ML) to make better decisions.
- Each application will be a mere exercise of the concepts we will introduce here.

Do not try this at home!

- Please Do not use a technical approach for love matters.

Uncertainty

It is present in every stage of life (at each decision).

You would thus need to:

1. Rationalize uncertainty (qualify)
2. Quantify uncertainty
3. Make decisions under uncertainty
4. Operate under uncertainty

Uncertainty

It is present in every stage of life (at each decision).

You would thus need to:

1. Rationalize uncertainty (qualify)
2. Quantify uncertainty
3. Make decisions under uncertainty
4. Operate under uncertainty

Uncertainty

It is present in every stage of life (at each decision).

You would thus need to:

1. Rationalize uncertainty (qualify)
2. Quantify uncertainty
3. Make decisions under uncertainty
4. Operate under uncertainty

Uncertainty

It is present in every stage of life (at each decision).

You would thus need to:

1. Rationalize uncertainty (qualify)
2. Quantify uncertainty
3. Make decisions under uncertainty
4. Operate under uncertainty

Uncertainty

It is present in every stage of life (at each decision).

You would thus need to:

1. Rationalize uncertainty (qualify)
2. Quantify uncertainty
3. Make decisions under uncertainty
4. Operate under uncertainty

Uncertainty or Probability?

Our aim here is to provide good guidance on how to link data, models and output to value creation.

- First we need to understand uncertainty and probability, and the difference between the two.

- Spoiler: Probability is the language of uncertainty!

- We will analyse, quantify and structure models as a function of uncertainty.
- Clarity of language in probability is one of the hallmarks of decision analysis and of modeling.

Uncertainty or Probability?

Our aim here is to provide good guidance on how to link data, models and output to value creation.

- First we need to understand uncertainty and probability, and the difference between the two.

- Spoiler: Probability is the language of uncertainty!

- We will analyse, quantify and structure models as a function of uncertainty.
- Clarity of language in probability is one of the hallmarks of decision analysis and of modeling.

Uncertainty or Probability?

Our aim here is to provide good guidance on how to link data, models and output to value creation.

- First we need to understand uncertainty and probability, and the difference between the two.

- Spoiler: Probability is the language of uncertainty!

- We will analyse, quantify and structure models as a function of uncertainty.
- Clarity of language in probability is one of the hallmarks of decision analysis and of modeling.

Uncertainty or Probability?

Our aim here is to provide good guidance on how to link data, models and output to value creation.

- First we need to understand uncertainty and probability, and the difference between the two.

- Spoiler: Probability is the language of uncertainty!

- We will analyse, quantify and structure models as a function of uncertainty.
- Clarity of language in probability is one of the hallmarks of decision analysis and of modeling.

Uncertainty or Probability?

Our aim here is to provide good guidance on how to link data, models and output to value creation.

- First we need to understand uncertainty and probability, and the difference between the two.

- Spoiler: Probability is the language of uncertainty!

- We will analyse, quantify and structure models as a function of uncertainty.
- Clarity of language in probability is one of the hallmarks of decision analysis and of modeling.

Data properties

- All starts from data: what are data-properties?

- Are there such things as good data and bad data?

Life lesson (or exam question, same thing ;))

- Data DO NOT always have value.

- TRASH in TRASH out

Data properties

- All starts from data: what are data-properties?

- Are there such things as good data and bad data?

Life lesson (or exam question, same thing ;))

- Data DO NOT always have value.

- TRASH in TRASH out

Data properties

- All starts from data: what are data-properties?

- Are there such things as good data and bad data?

Life lesson (or exam question, same thing ;))

- Data DO NOT always have value.

- TRASH in TRASH out

Data properties

- All starts from data: what are data-properties?

- Are there such things as good data and bad data?

Life lesson (or exam question, same thing ;))

- Data DO NOT always have value.

- TRASH in TRASH out

Spatial and Temporal Data

Statistics is collecting, organizing, and interpreting data

//

Spatial and temporal statistics is a branch of applied statistics that emphasizes

1. the context of the data,
2. the spatial and time dependent relationship between data
3. the different relative value and precision of the data.

Spatial and Temporal Data

Statistics is collecting, organizing, and interpreting data

//

Spatial and temporal statistics is a branch of applied statistics that emphasizes

1. the context of the data,
2. the spatial and time dependent relationship between data
3. the different relative value and precision of the data.

Quantifying uncertainty

On the data sources side

- Confidence intervals
- Relevance
- Significance
- Correlation
- Causation
- Data Filters
- Biases identification

Quantifying uncertainty

On the data sources side

- Confidence intervals
- Relevance
- Significance
- Correlation
- Causation
- Data Filters
- Biases identification

Quantifying uncertainty

On the data sources side

- Confidence intervals
- Relevance
- Significance
- Correlation
- Causation
- Data Filters
- Biases identification

Quantifying uncertainty

On the data sources side

- Confidence intervals
- Relevance
- Significance
- Correlation
- Causation
- Data Filters
- Biases identification

Quantifying uncertainty

On the data sources side

- Confidence intervals
- Relevance
- Significance
- Correlation
- Causation
- Data Filters
- Biases identification

Quantifying uncertainty

On the data sources side

- Confidence intervals
- Relevance
- Significance
- Correlation
- Causation
- Data Filters
- Biases identification

Quantifying uncertainty

On the data sources side

- Confidence intervals
- Relevance
- Significance
- Correlation
- Causation
- Data Filters
- Biases identification

Quantifying uncertainty

On the data sources side

- Confidence intervals
- Relevance
- Significance
- Correlation
- Causation
- Data Filters
- Biases identification

1 Personal Intro

2 Intro to MOD550

3 Modeling

4 Recaps: Data concepts, Data types and Models

5 Descriptive and Predictive statistics

6 Simulations

7 Statistics, Machine learning or Artificial intelligence?

8 Metadata

On the modelling side

- Regression
- Principal components
- Decision tree (random forests)
- Neural network
- Clustering
- Performance metrics

Quantifying uncertainty

On the modelling side

- Regression
- Principal components
- Decision tree (random forests)
- Neural network
- Clustering
- Performance metrics

Quantifying uncertainty

On the modelling side

- Regression
- Principal components
- Decision tree (random forests)
- Neural network
- Clustering
- Performance metrics

Quantifying uncertainty

On the modelling side

- Regression
- Principal components
- Decision tree (random forests)
- Neural network
- Clustering
- Performance metrics

Quantifying uncertainty

On the modelling side

- Regression
- Principal components
- Decision tree (random forests)
- Neural network
- Clustering
- Performance metrics

Quantifying uncertainty

On the modelling side

- Regression
- Principal components
- Decision tree (random forests)
- Neural network
- Clustering
- Performance metrics

Quantifying uncertainty

On the modelling side

- Regression
- Principal components
- Decision tree (random forests)
- Neural network
- Clustering
- Performance metrics

Fields of application

- Spatial estimation of energy and mineral resources
- Weather modeling: from aviation to agriculture
- Maintenance forecasting
- Commodity, currency, stock and financial markets
- Market analysis
- Risk analysis
- ... and much much more!

Fields of application

- Spatial estimation of energy and mineral resources
- Weather modeling: from aviation to agriculture
- Maintenance forecasting
- Commodity, currency, stock and financial markets
- Market analysis
- Risk analysis
- ... and much much more!

Fields of application

- Spatial estimation of energy and mineral resources
- Weather modeling: from aviation to agriculture
- Maintenance forecasting
- Commodity, currency, stock and financial markets
- Market analysis
- Risk analysis
- ... and much much more!

Fields of application

- Spatial estimation of energy and mineral resources
- Weather modeling: from aviation to agriculture
- Maintenance forecasting
- Commodity, currency, stock and financial markets
- Market analysis
- Risk analysis
- ... and much much more!

Fields of application

- Spatial estimation of energy and mineral resources
- Weather modeling: from aviation to agriculture
- Maintenance forecasting
- Commodity, currency, stock and financial markets
- Market analysis
- Risk analysis
- ... and much much more!

Fields of application

- Spatial estimation of energy and mineral resources
- Weather modeling: from aviation to agriculture
- Maintenance forecasting
- Commodity, currency, stock and financial markets
- Market analysis
- Risk analysis
- ... and much much more!

Fields of application

- Spatial estimation of energy and mineral resources
- Weather modeling: from aviation to agriculture
- Maintenance forecasting
- Commodity, currency, stock and financial markets
- Market analysis
- Risk analysis
- ... and much much more!

Fields of application

- Spatial estimation of energy and mineral resources
- Weather modeling: from aviation to agriculture
- Maintenance forecasting
- Commodity, currency, stock and financial markets
- Market analysis
- Risk analysis
- ... and much much more!

Hard and soft modeling

Models allow us to predict 'the future', or describe the past and present (what is the present...?)

Last life lesson for today

Models are always wrong, but some are useful. (George Box)

Three main families:

1. Hard models (physics)
2. Soft models (statistic)
3. Machine learning

Hard and soft modeling

Models allow us to predict 'the future', or describe the past and present (what is the present...?)

Last life lesson for today

Models are always wrong, but some are useful. (George Box)

Three main families:

1. Hard models (physics)
2. Soft models (statistic)
3. Machine learning

Hard and soft modeling

Models allow us to predict 'the future', or describe the past and present (what is the present...?)

Last life lesson for today

Models are always wrong, but some are useful. (George Box)

Three main families:

1. Hard models (physics)
2. Soft models (statistic)
3. Machine learning

Hard modeling

- Based on an accurate physical description of the system and mathematical modeling (e.g. differential equations). Hard models are often deterministic.
- Hard modeling methods usually use optimization methods to find out the best values for the parameters of the model.
- Hard modeling is preferable in laboratory experiments, where all the variables are controlled and the physicochemical nature of the dynamic model is known and can be fully described using a known mathematical model.
- Hard modeling, if successful, usually gives better understanding of a system and better extrapolations. Wrong assumptions often lead to non-sense results.

Hard modeling

- Based on an accurate physical description of the system and mathematical modeling (e.g. differential equations). Hard models are often deterministic.
- Hard modeling methods usually use optimization methods to find out the best values for the parameters of the model.
- Hard modeling is preferable in laboratory experiments, where all the variables are controlled and the physicochemical nature of the dynamic model is known and can be fully described using a known mathematical model.
- Hard modeling, if successful, usually gives better understanding of a system and better extrapolations. Wrong assumptions often lead to non-sense results.

Hard modeling

- Based on an accurate physical description of the system and mathematical modeling (e.g. differential equations). Hard models are often deterministic.
- Hard modeling methods usually use optimization methods to find out the best values for the parameters of the model.
- Hard modeling is preferable in laboratory experiments, where all the variables are controlled and the physicochemical nature of the dynamic model is known and can be fully described using a known mathematical model.
- Hard modeling, if successful, usually gives better understanding of a system and better extrapolations. Wrong assumptions often lead to non-sense results.

Hard modeling

- Based on an accurate physical description of the system and mathematical modeling (e.g. differential equations). Hard models are often deterministic.
- Hard modeling methods usually use optimization methods to find out the best values for the parameters of the model.
- Hard modeling is preferable in laboratory experiments, where all the variables are controlled and the physicochemical nature of the dynamic model is known and can be fully described using a known mathematical model.
- Hard modeling, if successful, usually gives better understanding of a system and better extrapolations. Wrong assumptions often lead to non-sense results.

Hard modeling

- Based on an accurate physical description of the system and mathematical modeling (e.g. differential equations). Hard models are often deterministic.
- Hard modeling methods usually use optimization methods to find out the best values for the parameters of the model.
- Hard modeling is preferable in laboratory experiments, where all the variables are controlled and the physicochemical nature of the dynamic model is known and can be fully described using a known mathematical model.
- Hard modeling, if successful, usually gives better understanding of a system and better extrapolations. Wrong assumptions often lead to non-sense results.

Soft modeling

- Soft-modeling describes systems without the need of an *a priori* physical or (bio)chemical model postulation. They are **data driven** models.
- Soft models are much easier to make than hard models.
- Soft modeling can be used to understand complex relationships.
- Soft modeling needs (much) more data than hard-modeling.
- Soft models have a poor extrapolating capabilities (compared with hard-modeling)

Soft modeling

- Soft-modeling describes systems without the need of an *a priori* physical or (bio)chemical model postulation. They are **data driven** models.
- Soft models are much easier to make than hard models.
- Soft modeling can be used to understand complex relationships.
- Soft modeling needs (much) more data than hard-modeling.
- Soft models have a poor extrapolating capabilities (compared with hard-modeling)

Soft modeling

- Soft-modeling describes systems without the need of an *a priori* physical or (bio)chemical model postulation. They are **data driven** models.
- Soft models are much easier to make than hard models.
- Soft modeling can be used to understand complex relationships.
- Soft modeling needs (much) more data than hard-modeling.
- Soft models have a poor extrapolating capabilities (compared with hard-modeling)

Soft modeling

- Soft-modeling describes systems without the need of an *a priori* physical or (bio)chemical model postulation. They are **data driven** models.
- Soft models are much easier to make than hard models.
- Soft modeling can be used to understand complex relationships.
- Soft modeling needs (much) more data than hard-modeling.
- Soft models have a poor extrapolating capabilities (compared with hard-modeling)

Soft modeling

- Soft-modeling describes systems without the need of an *a priori* physical or (bio)chemical model postulation. They are **data driven** models.
- Soft models are much easier to make than hard models.
- Soft modeling can be used to understand complex relationships.
- Soft modeling needs (much) more data than hard-modeling.
- Soft models have a poor extrapolating capabilities (compared with hard-modeling)

Soft modeling

- Soft-modeling describes systems without the need of an *a priori* physical or (bio)chemical model postulation. They are **data driven** models.
- Soft models are much easier to make than hard models.
- Soft modeling can be used to understand complex relationships.
- Soft modeling needs (much) more data than hard-modeling.
- Soft models have a poor extrapolating capabilities (compared with hard-modeling)

How to create hard models?

After understanding the problem to be solved we need to:

1. Link mathematics to physics.
2. Define boundary conditions and constitutive equations.
3. Make tons of assumptions.
4. Solve the constitutive equation in space and time.
5. Check solution stability and sensitivity analysis.
6. A long set of judicious approximations have to be taken.
7. It is hard (but we are engineers!).
8. Get quite some money for the awesome job.

How to create soft models?

After understanding the general problem to be solved we need to:

- Determine a suitable **numerical description** .
- Choose a suitable **model** to which parameters are fitted.
- Train, test, validate the model.
- Perform **data analysis** with chosen method(s).
- Link predictions with expectations.

Hard models vs Soft models

Requirements

Deterministic:

- physics and expert knowledge
- integration of various information sources
- very complicated

Statistical:

- quick
- uncertainty assessment
- data driven approach
- physics can be included
- stochastic modeling

Hard models vs Soft models

Behaviour

Deterministic:

- predictable
- defined error

Statistical:

- outcome uncertainty
- undefined error
- sampling resolution issues

Spatial and Temporal Modeling

It is a branch of statistical analysis and model that uses spatial and time dependent data.

- Only a subset of statistical models can be fed with time dependent data

(most standard statistical method assume independent, identically distributed, data)

- Spatial and time related data come at a different range of scales

Frequency of data collection can be dependent of time and space, resulting in different representativity of a sample.

Spatial and Temporal Modeling

It is a branch of statistical analysis and model that uses spatial and time dependent data.

- Only a subset of statistical models can be fed with time dependent data

(most standard statistical method assume independent, identically distributed, data)

- Spatial and time related data come at a different range of scales

Frequency of data collection can be dependent of time and space, resulting in different representativity of a sample.

Spatial and Temporal Modeling

It is a branch of statistical analysis and model that uses spatial and time dependent data.

- Only a subset of statistical models can be fed with time dependent data

(most standard statistical method assume independent, identically distributed, data)

- Spatial and time related data come at a different range of scales

Frequency of data collection can be dependent of time and space, resulting in different representativity of a sample.

1 Personal Intro

2 Intro to MOD550

3 Modeling

4 Recaps: Data concepts, Data types and Models

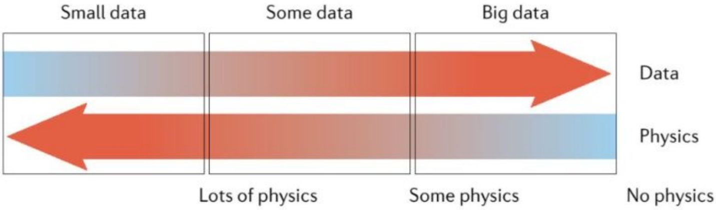
5 Descriptive and Predictive statistics

6 Simulations

7 Statistics, Machine learning or Artificial intelligence?

8 Metadata

Data vs Physics



Uncertainty

- Def 1: Not knowing if an event is true or false. (Useful)
- Def 2: Things that cannot be measured. (Not useful)

Probability is how Uncertainty is quantified!

- Clarity test
- Assign a number between 0 and 1 to our degree of belief
- Error definition

Sentence also good for fortune cookies

Uncertainty is the only certainty

Uncertainty

- Def 1: Not knowing if an event is true or false. (Useful)
- Def 2: Things that cannot be measured. (Not useful)

Probability is how Uncertainty is quantified!

- Clarity test
- Assign a number between 0 and 1 to our degree of belief
- Error definition

Sentence also good for fortune cookies

Uncertainty is the only certainty

Uncertainty

- Def 1: Not knowing if an event is true or false. (Useful)
- Def 2: Things that cannot be measured. (Not useful)

Probability is how Uncertainty is quantified!

- Clarity test
- Assign a number between 0 and 1 to our degree of belief
- Error definition

Sentence also good for fortune cookies

Uncertainty is the only certainty

Uncertainty and Probability

Random quotes

- Probability: there is not science more worthy in out contemplations nor a more useful one for admission to our system of public education
- The theory of probabilities is at the bottom of nothing but common sense reduced to calculus.

What is Statistics

Clarity test. Beer drinker?

Rain in Stavanger?

Random quotes

- Probability: there is not science more worthy in out contemplations nor a more useful one for admission to our system of public education
- The theory of probabilities is at the bottom of nothing but common sense reduced to calculus.

What is Statistics

Clarity test. Beer drinker?

Rain in Stavanger?

Data properties

1 D

logs

2 D: maps

Quite limited but great for visualization

3 D

3d maps, seismic cubes. More informative, mostly ok in digital formats.

4 D

Trajectories

x D

Data realm

Data properties

1 D

logs

2 D: maps

Quite limited but great for visualization

3 D

3d maps, seismic cubes. More informative, mostly ok in digital formats.

4 D

Trajectories

x D

Data realm

Data properties

1 D

logs

2 D: maps

Quite limited but great for visualization

3 D

3d maps, seismic cubes. More informative, mostly ok in digital formats.

4 D

Trajectories

x D

Data realm

Data properties

1 D

logs

2 D: maps

Quite limited but great for visualization

3 D

3d maps, seismic cubes. More informative, mostly ok in digital formats.

4 D

Trajectories

x D

Data realm

Data properties

1 D
logs

2 D: maps
Quite limited but great for visualization

3 D
3d maps, seismic cubes. More informative, mostly ok in digital formats.

4 D
Trajectories

x D
Data realm

Types of data

- Categorical / Nominal (classes)
- Categorical / Ordinal
- Continuous / Interval (e.g. Celsius)
- Continuous / Ratio
- Discrete: binned/grouped data
- Hard data: direct measurements
- Soft data: indirect measurements, very uncertain
- Primary data: variable(s) of interest
- Secondary data: descriptors
- Collective variables
- Latent variables

Types of data

- Categorical / Nominal (classes)
- Categorical / Ordinal
- Continuous / Interval (e.g. Celsius)
- Continuous / Ratio
- Discrete: binned/grouped data
- Hard data: direct measurements
- Soft data: indirect measurements, very uncertain
- Primary data: variable(s) of interest
- Secondary data: descriptors
- Collective variables
- Latent variables

1 Personal Intro

2 Intro to MOD550

3 Modeling

4 Recaps: Data concepts, Data types and Models

5 Descriptive and Predictive statistics

6 Simulations

7 Statistics, Machine learning or Artificial intelligence?

8 Metadata

Descriptive and Predictive statistics

Estimation

- Process of obtaining the best value or range of a property in an unsampled location
- Local accuracy takes precedence over global spatial variability
- Not appropriate for forecasting

Inference

- Predict unseen samples given assumptions about the population
- Test with a pre-trained model (ML definition)
- Generality versus Accuracy

Descriptive and Predictive statistics

Estimation

- Process of obtaining the best value or range of a property in an unsampled location
- Local accuracy takes precedence over global spatial variability
- Not appropriate for forecasting

Inference

- Predict unseen samples given assumptions about the population
- Test with a pre-trained model (ML definition)
- Generality versus Accuracy

Descriptive and Predictive statistics

Estimation

- Process of obtaining the best value or range of a property in an unsampled location
- Local accuracy takes precedence over global spatial variability
- Not appropriate for forecasting

Inference

- Predict unseen samples given assumptions about the population
- Test with a pre-trained model (ML definition)
- Generality versus Accuracy

Variables and Features, Labels and Instances

Population

Exhaustive, finite list of properties of interest over area of interest.

Generally the entire population is not accessible

Samples/experiments/instances

The set of values and location that have been measured.

How many experiments are needed?

Features

The values to be measured for each sample/experiment/instance.

How many features are needed?

Variables and Features, Labels and Instances

Population

Exhaustive, finite list of properties of interest over area of interest.

Generally the entire population is not accessible

Samples/experiments/instances

The set of values and location that have been measured.

How many experiments are needed?

Features

The values to be measured for each sample/experiment/instance.

How many features are needed?

Variables and Features, Labels and Instances

Population

Exhaustive, finite list of properties of interest over area of interest.

Generally the entire population is not accessible

Samples/experiments/instances

The set of values and location that have been measured.

How many experiments are needed?

Features

The values to be measured for each sample/experiment/instance.

How many features are needed?

Variables and Features, Labels and Instances

Predictors = input variables, X_1, ..., X_M

Response = output variables

Error

Deviation from ... exact value (or expected value, mean value, trend...?)

Errors without definitions are just numbers.

Predictor and Response Features

Given a model $Y = f(X_1, ..., X_M) + e$

!Here and error! But is it even an error?

Variables and Features, Labels and Instances

Predictors = input variables, X_1, ..., X_M

Response = output variables

Error

Deviation from ... exact value (or expected value, mean value, trend...?)

Errors without definitions are just numbers.

Predictor and Response Features

Given a model $Y = f(X_1, ..., X_M) + e$

!Here and error! But is it even an error?

Variables and Features, Labels and Instances

Predictors = input variables, X_1, ..., X_M

Response = output variables

Error

Deviation from ... exact value (or expected value, mean value, trend...?)

Errors without definitions are just numbers.

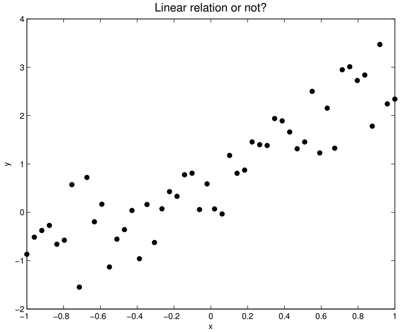
Predictor and Response Features

Given a model $Y = f(X_1, ..., X_M) + e$

!Here and error! But is it even an error?

Finding a suitable model

Soft modeling is in most cases based on **multivariate statistical methods**. Many of these methods may be viewed as sophisticated ways of performing curve fitting to data.



What would be the best model?

- Straight?: $y(x) = ax + b$
- Parabolic?: $y(x) = ax^2 + bx + c$
- Trigonometric?: $y(x) = asin(x) + bcos(x)$

Uncertainty Modeling

Given a model, Generate multiple simulation to represent uncertainty

- Realizations: for the same input parameters, different random numbers.
- Scenarios: different input parameters.

Sampling representative.

Random sampling

Each item of the population has an equal chance of being chosen.

- Very expensive
- Mostly not interesting
- Gives some global properties

Bias sampling

Selection of data is (arbitrarily) distorted

- Sample probability bias has to be corrected for
- Might not capture the global picture

Given a model, Generate multiple simulation to represent uncertainty

- Realizations: for the same input parameters, different random numbers.
- Scenarios: different input parameters.

Sampling representative.

Random sampling

Each item of the population has an equal chance of being chosen.

- Very expensive
- Mostly not interesting
- Gives some global properties

Bias sampling

Selection of data is (arbitrarily) distorted

- Sample probability bias has to be corrected for
- Might not capture the global picture

Cognitive biases

- anchoring: The first bits are over-considered
- availability: over-estimating the importance of info
- bandwagon: P increases with the number of people holding a belief
- blind spot: not seen biases
- choice supporting: commitment/decision dependent
- clustering illusion: seeing patterns in random events
- confirmation bias
- conservatism bias
- Recency bias
- Supervision bias
- Many many more!

Bias DO NOT cancel out! They sum up (or multiply?)

1. Personal Intro
2. Intro to MOD550
3. Modeling
4. Recaps: Data concepts, Data types and Models
5. Descriptive and Predictive statistics
6. Simulations
7. Statistics, Machine learning or Artificial intelligence?
8. Metadata

Simulations

Process of obtaining one or more values of a property

- Improved Global accuracy
- Better property distributions

Why simulations then?

- We need to capture the full distribution of properties, extremes matter!
- We need more realistic models.

Why not?

- High dimensionality level
- Computationally expensive
- Convergence limitations
- Constitutive equations need to be rather accurate.

Simulations

Process of obtaining one or more values of a property

- Improved Global accuracy
- Better property distributions

Why simulations then?

- We need to capture the full distribution of properties, extremes matter!
- We need more realistic models.

Why not?

- High dimensionality level
- Computationally expensive
- Convergence limitations
- Constitutive equations need to be rather accurate.

Simulations

Process of obtaining one or more values of a property

- Improved Global accuracy
- Better property distributions

Why simulations then?

- We need to capture the full distribution of properties, extremes matter!
- We need more realistic models.

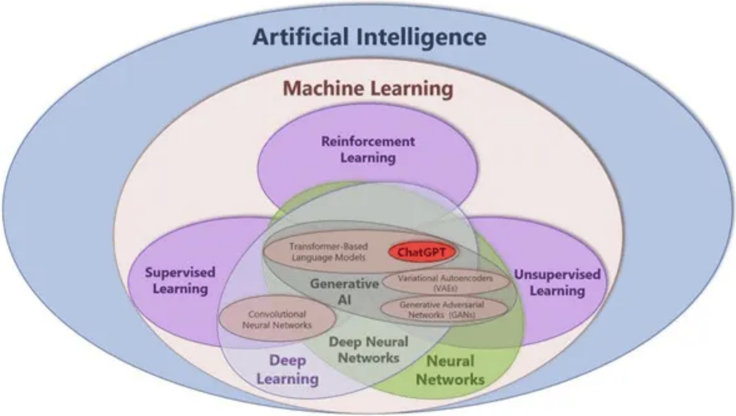
Why not?

- High dimensionality level
- Computationally expensive
- Convergence limitations
- Constitutive equations need to be rather accurate.

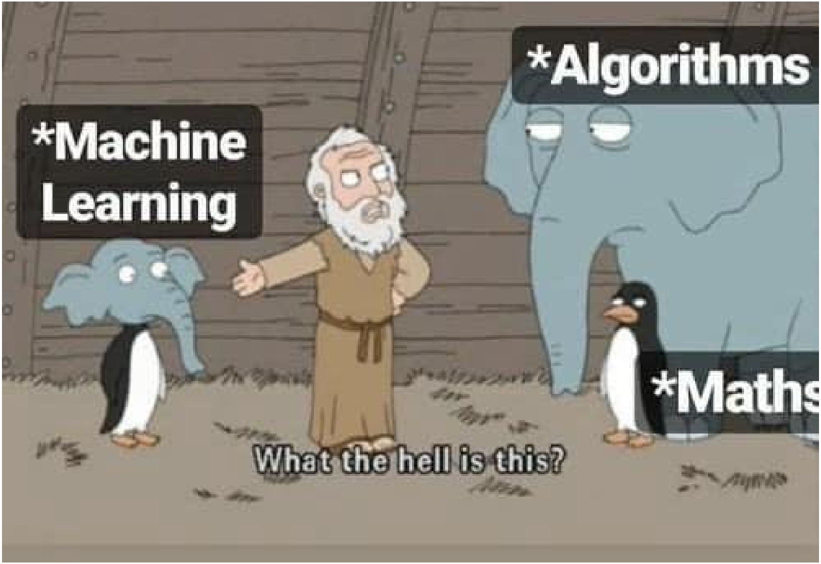
1. Personal Intro
2. Intro to MOD550
3. Modeling
4. Recaps: Data concepts, Data types and Models
5. Descriptive and Predictive statistics
6. Simulations
7. Statistics, Machine learning or Artificial intelligence?
8. Metadata

Statistics, Machine learning or Artificial intelligence?

What is the main difference between the three fields?



How Machine Learning Started?



Let's start from the definition

- Statistics (origin "description of a state/country") is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- It is conventional to begin with a statistical population or a statistical model to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments.[Wikipedia]

Let's start from the definition

- Statistics (origin "description of a state/country") is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- It is conventional to begin with a statistical population or a statistical model to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments.[Wikipedia]

Statistics

Let's start from the definition

- Statistics (origin "description of a state/country") is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- It is conventional to begin with a statistical population or a statistical model to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments.[Wikipedia]

Statistics

Let's start from the definition

- Statistics (origin "description of a state/country") is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- It is conventional to begin with a statistical population or a statistical model to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments.[Wikipedia]

Machine learning

Definitions:

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. [IBM]
- Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. [WIKI]
- Machine learning is a subfield of artificial intelligence that uses algorithms trained on data sets to create models that enable machines to perform tasks that would otherwise only be possible for humans, such as categorizing images, analyzing data, or predicting price fluctuations. [Coursera]

Machine learning

Definitions:

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. [IBM]
- Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. [WIKI]
- Machine learning is a subfield of artificial intelligence that uses algorithms trained on data sets to create models that enable machines to perform tasks that would otherwise only be possible for humans, such as categorizing images, analyzing data, or predicting price fluctuations. [Coursera]

Definitions:

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. [IBM]
- Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. [WIKI]
- Machine learning is a subfield of artificial intelligence that uses algorithms trained on data sets to create models that enable machines to perform tasks that would otherwise only be possible for humans, such as categorizing images, analyzing data, or predicting price fluctuations. [Coursera]

Machine learning

Definitions:

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. [IBM]
- Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. [WIKI]
- Machine learning is a subfield of artificial intelligence that uses algorithms trained on data sets to create models that enable machines to perform tasks that would otherwise only be possible for humans, such as categorizing images, analyzing data, or predicting price fluctuations. [Coursera]

One technical definition

Machine learning is a set of computer based statistical approaches that aim to minimise the loss function to maximise inference accuracy. [Enrico, 5.2.2024]

The loss function is the actual engine in machine learning.

Loss function

It quantifies the difference between the predicted outputs of a machine learning algorithm and the actual target values.

One technical definition

Machine learning is a set of computer based statistical approaches that aim to minimise the loss function to maximise inference accuracy. [Enrico, 5.2.2024]

The loss function is the actual engine in machine learning.

Loss function

It quantifies the difference between the predicted outputs of a machine learning algorithm and the actual target values.

Machine learning

One technical definition

Machine learning is a set of computer based statistical approaches that aim to minimise the loss function to maximise inference accuracy. [Enrico, 5.2.2024]

The loss function is the actual engine in machine learning.

Loss function

It quantifies the difference between the predicted outputs of a machine learning algorithm and the actual target values.

Machine learning

One technical definition

Machine learning is a set of computer based statistical approaches that aim to minimise the loss function to maximise inference accuracy. [Enrico, 5.2.2024]

The loss function is the actual engine in machine learning.

Loss function

It quantifies the difference between the predicted outputs of a machine learning algorithm and the actual target values.

Artificial intelligences

And more definitions:

- Artificial intelligence is the intelligence of machines or software, as opposed to the intelligence of humans or other animals. It is a field of study in computer science that develops and studies intelligent machines. [WIKI]
- Artificial intelligence (AI) is the theory and development of computer systems capable of performing tasks that historically required human intelligence, such as recognizing speech, making decisions, and identifying patterns [Coursera]
- It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. [IBM]

Artificial intelligences

And more definitions:

- Artificial intelligence is the intelligence of machines or software, as opposed to the intelligence of humans or other animals. It is a field of study in computer science that develops and studies intelligent machines. [WIKI]
- Artificial intelligence (AI) is the theory and development of computer systems capable of performing tasks that historically required human intelligence, such as recognizing speech, making decisions, and identifying patterns [Coursera]
- It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. [IBM]

Artificial intelligences

And more definitions:

- Artificial intelligence is the intelligence of machines or software, as opposed to the intelligence of humans or other animals. It is a field of study in computer science that develops and studies intelligent machines. [WIKI]
- Artificial intelligence (AI) is the theory and development of computer systems capable of performing tasks that historically required human intelligence, such as recognizing speech, making decisions, and identifying patterns [Coursera]
- It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. [IBM]

Artificial intelligences

And more definitions:

- Artificial intelligence is the intelligence of machines or software, as opposed to the intelligence of humans or other animals. It is a field of study in computer science that develops and studies intelligent machines. [WIKI]
- Artificial intelligence (AI) is the theory and development of computer systems capable of performing tasks that historically required human intelligence, such as recognizing speech, making decisions, and identifying patterns [Coursera]
- It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. [IBM]

1. Personal Intro
2. Intro to MOD550
3. Modeling
4. Recaps: Data concepts, Data types and Models
5. Descriptive and Predictive statistics
6. Simulations
7. Statistics, Machine learning or Artificial intelligence?
8. Metadata

Data properties

- All starts from data: what are data-properties?

- Are there such things as good data and bad data?

Main lesson (Exam question)

- Data DO NOT always have value.

- TRASH in TRASH out

Data properties

- All starts from data: what are data-properties?

- Are there such things as good data and bad data?

Main lesson (Exam question)

- Data DO NOT always have value.

- TRASH in TRASH out

Data properties

- All starts from data: what are data-properties?

- Are there such things as good data and bad data?

Main lesson (Exam question)

- Data DO NOT always have value.

- TRASH in TRASH out

Data properties

- All starts from data: what are data-properties?

- Are there such things as good data and bad data?

Main lesson (Exam question)

- Data DO NOT always have value.

- TRASH in TRASH out

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData aim

These characteristics shall all be considered when constructing data repositories.

They are a must for:

- Code repositories
- Data repositories

Please click on the link and explore. Those are just some of the open (scientific) repositories. The aim/hope is to allow people to extract information. How to make value from that information... is another story.

MetaData aim

These characteristics shall all be considered when constructing data repositories.

They are a must for:

- Code repositories
- Data repositories

Please click on the link and explore. Those are just some of the open (scientific) repositories. The aim/hope is to allow people to extract information. How to make value from that information... is another story.

MetaData aim

These characteristics shall all be considered when constructing data repositories.

They are a must for:

- Code repositories
- Data repositories

Please click on the link and explore. Those are just some of the open (scientific) repositories. The aim/hope is to allow people to extract information. How to make value from that information... is another story.

MetaData aim

These characteristics shall all be considered when constructing data repositories.

They are a must for:

- Code repositories
- Data repositories

Please click on the link and explore. Those are just some of the open (scientific) repositories. The aim/hope is to allow people to extract information. How to make value from that information... is another story.

MetaData aim

These characteristics shall all be considered when constructing data repositories.

They are a must for:

- Code repositories
- Data repositories

Please click on the link and explore. Those are just some of the open (scientific) repositories. The aim/hope is to allow people to extract information. How to make value from that information... is another story.

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

Good examples

- Norwegian offshore directorate
- Norway Statistics
- World statistics
- Code repositories
- Data repositories