Machine Learning: short introduction (trees and ANN), part 1

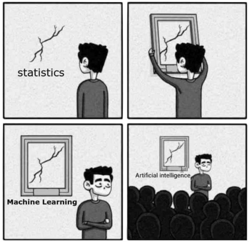
Sergey Korpachev on behalf of Dépôt

Outline

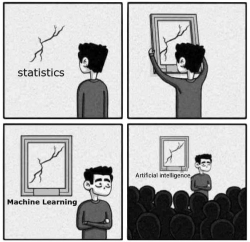
- Machine learning (ML)
- Data
- Features in ML
- ML pipeline
- Linear model
- Decision tree
- Neural network
- Summary

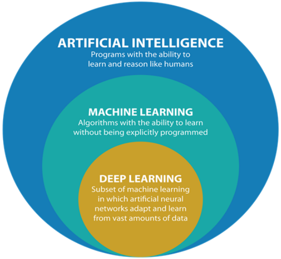
Machine learning

Machine learning — what is it?

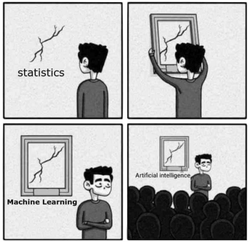


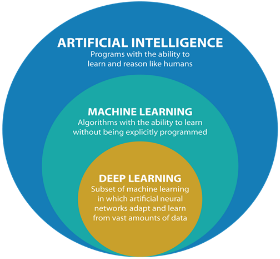
Machine learning — what is it?





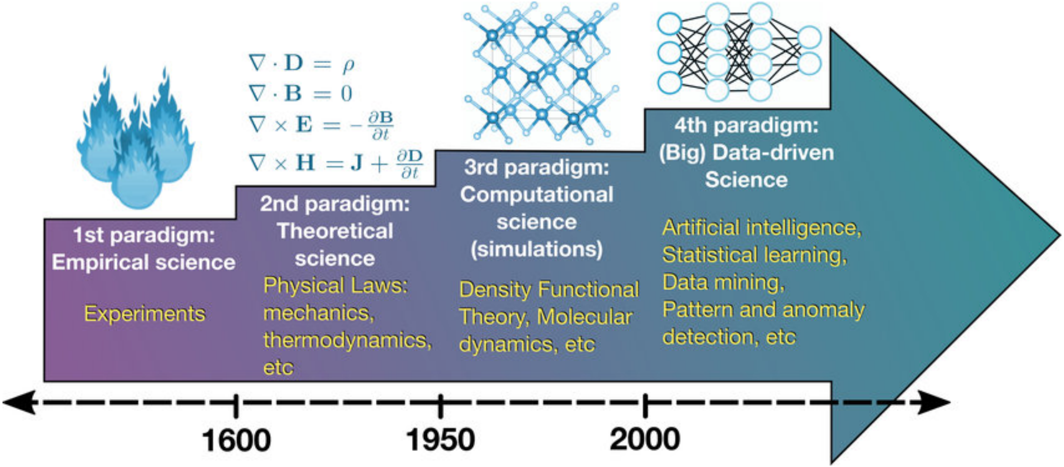
Machine learning — what is it?





Machine learning

The Fourth Paradigm (Tony Hey, ...)



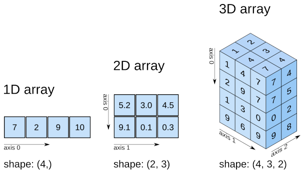
Data

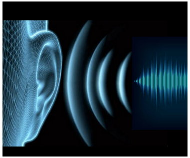
Data

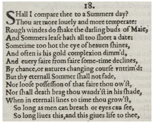
what is data?

Anything can be data:

1. Numbers
2. Text
3. Images
4. Sound
5. Geomap
6. Particle collisions
7. Knowledge
8. You name it

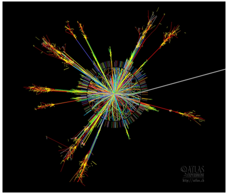






18.

SHall I compare thee to a Summers day?
Thou art more louely and more temperate:
Rough windes do ſhake the darling buds of Maie,
And Sommers leaſe hath all too ſhort a date:
Sometime too hot the eye of heauen ſhines,
And often is his gold complexion dimm'd,
And euery faire from faire ſome-time declines,
By chance,or natures changing courſe vntrim'd:
But thy eternall Sommer ſhall not fade,
Nor looſe poſſeſſion of that faire thou ow'ſt,
Nor ſhall death brag thou wandr'ſt in his ſhade,
When in eternall lines to time thou grow'ſt,
So long as men can breath or eyes can ſee,
So long liues this,and this giues life to thee,



Features in ML

Features in ML

not HEP

Supervised learning

Classification

- cat, dog or muffin
- relevant or spam
- disease or not
- good or bad
- ...

Regression

- rent price
- temperature
- annual profit
- driving time
- ...

Don't forget about **unsupervised learning**, **reinforcement learning**, **semi-supervised learning** and so on.

Features in ML — HEP

Supervised learning

Classification

- b , c, uds jet
- π, K, μ particle
- $t\bar{t}$ or QCD event
- select or reject trigger candidate
- ...

Regression

- energy resolution
- pile-up mitigation
- ...

Don't forget about **unsupervised learning**, **reinforcement learning**, **semi-supervised learning** and so on.

Features in ML

- Precise and fast particle tracking (single tracks, shower, jets)
- Particle identification
- Fast and accurate online data processing and filtering
- Anomaly detection (data quality monitoring, infrastructure monitoring)
- Detector design optimization (bayesian optimization, surrogate modelling)
- Data analysis (signal from background separation, ...)
- Simulation (speed-up simulation using generative models, simulator parameters optimization - tuning)
- ...

Features in ML

ML challenges: slide 2 (A. Ustyuzhanin)

Tracking system features

- Particle momentum
- Particle charge
- Track parameters
- Quality of track fit
- Number of track hits
- ...

RICH features

- Angle θ
- Quality of angle reconstruction
- Reconstructed particle type
- Reconstructed particle energy
- Light intensity
- ...

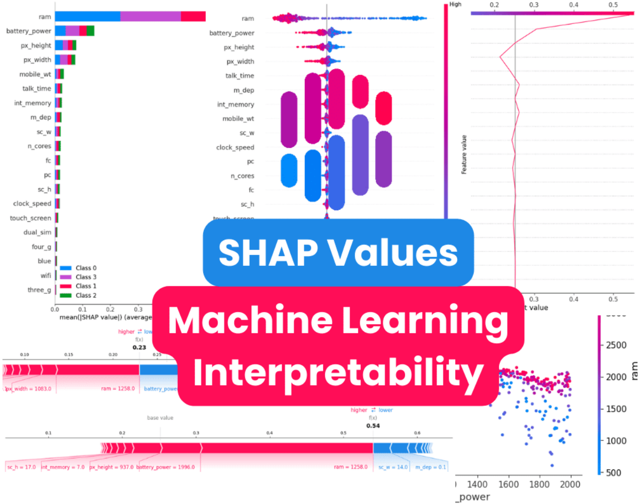
Calorimeter features

- Measured particle energy
- Shower parameters: length, width, ...
- Number of clusters in each layer Intensity of the clusters
- Intensity of the clusters
- Distance form track of the original particle
- ...

Muon detector features

- Muon track parameters
- Quality of track fit
- Number of active layers
- Distance between the track and the active layers
- Length of shower
- ...

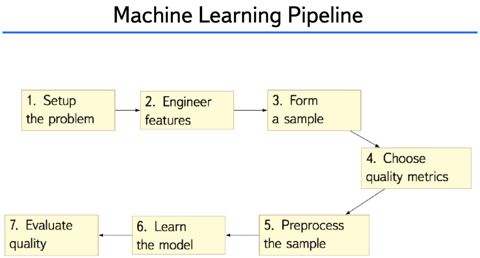
Features in ML → SHAP value



ML pipeline

ML pipeline

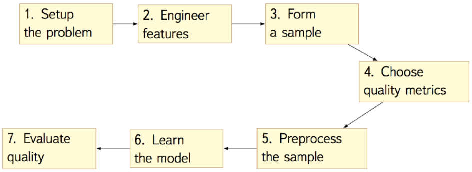
what is ML pipeline?



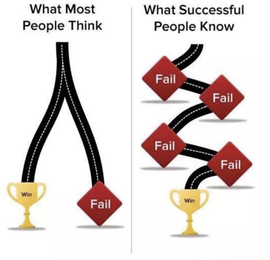
ML pipeline

what is ML pipeline?

Machine Learning Pipeline



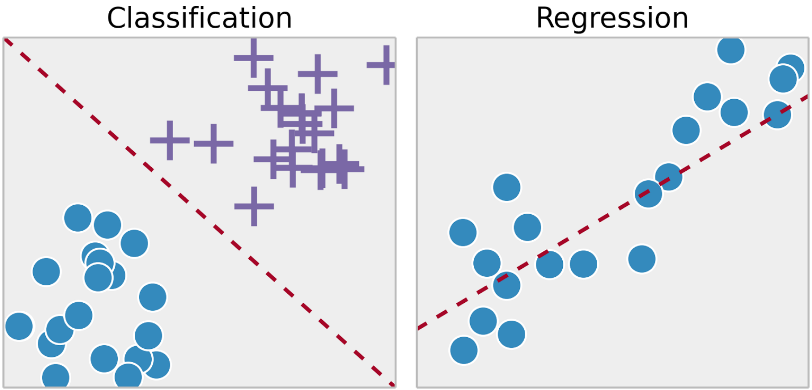
Machine Learning Pipeline



Linear model

Linear model

classification and regression



Linear model — regression

- $Y = \mathbb{R}$
- N objects with K real features: $D = \mathbb{R}^K$
- $a(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_0 + \sum\limits_{j=1}^{K} f_j(\boldsymbol{x}) \cdot \theta_j$
- Extend and reassign:
  $[1, f_1(\boldsymbol{x}), \ldots, f_K(\boldsymbol{x})] \equiv \boldsymbol{x}$
  $[\theta_0, \ldots, \theta_K] \equiv \boldsymbol{\theta}$
- Then $a(\boldsymbol{x}, \boldsymbol{\theta}) = \langle \boldsymbol{x}, \boldsymbol{\theta} \rangle$

- Minimization problem:
  - $\mathcal{L}(\boldsymbol{\theta}) = \big(\langle \boldsymbol{x}, \boldsymbol{\theta} \rangle - y\big)^2$
  - $Q(\boldsymbol{\theta}) = \frac{1}{N} \sum\limits_{i=1}^{N} \big(\langle \boldsymbol{x_i}, \boldsymbol{\theta} \rangle - y_i\big)^2$
- Then we need to minimize $Q(\boldsymbol{\theta})$ by varying $\boldsymbol{\theta}$:
  - $\hat{a} = \arg\min\limits_{\boldsymbol{\theta} \in \Theta} Q(\boldsymbol{\theta})$

Linear model — classification

- $Y = \{-1, +1\}$
- N objects with K real features: $D = \mathbb{R}^K$
- $a(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_0 + \sum\limits_{j=1}^{K} f_j(\boldsymbol{x}) \cdot \theta_j$
- Extend and reassign:
  $[1, f_1(\boldsymbol{x}), \ldots, f_K(\boldsymbol{x})] \equiv [1, x_1, \ldots, x_K] \equiv \boldsymbol{x}$
  $[\theta_0, \ldots, \theta_K] \equiv \boldsymbol{\theta}$
- Then $a(\boldsymbol{x}, \boldsymbol{\theta}) = \langle \boldsymbol{x}, \boldsymbol{\theta} \rangle$

- Minimization problem:
  - $\mathcal{L}(\boldsymbol{\theta}) = \left[\operatorname{sign}\langle \boldsymbol{x}, \boldsymbol{\theta} \rangle \neq y\right]$
  - $Q(\boldsymbol{\theta}) = \frac{1}{N} \sum\limits_{i=1}^{N} \left[\operatorname{sign}\langle \boldsymbol{x_i}, \boldsymbol{\theta} \rangle \neq y_i\right]$
- Then we need to minimize $Q(\boldsymbol{\theta})$ by varying $\boldsymbol{\theta}$:
  - $\hat{a} = \arg\min\limits_{\boldsymbol{\theta} \in \Theta} Q(\boldsymbol{\theta})$