Class 6- Machine Learning concepts
Part III (putting it together!)

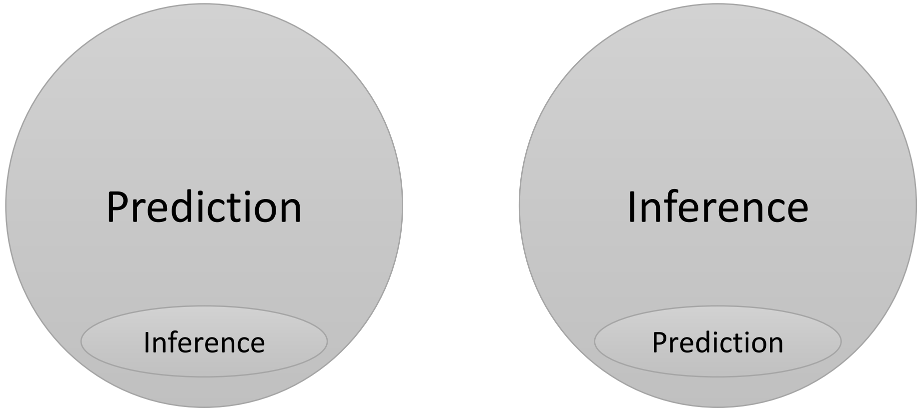
Agenda

- ❑ Slides

Motivation

Machine learning fundamental concepts:

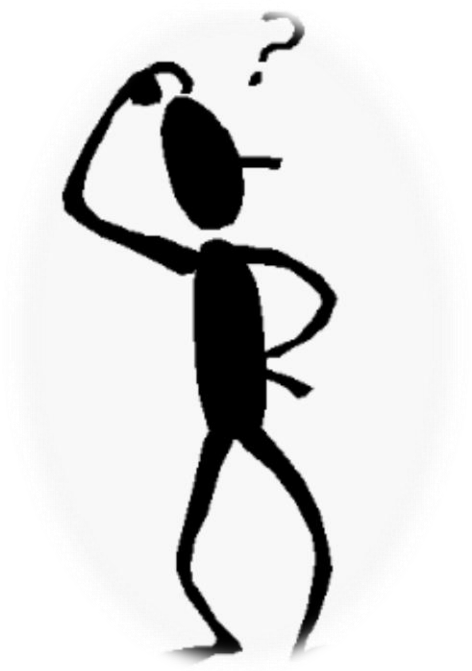
- Inference and prediction
- Part I: The Model
- Part II: Evaluation metrics
- Part III: Bias-Variance tradeoff
- Part IV: Resampling methods
- Part V: Solvers/learners (GD, SGD)
- Part VI: How do machines learn?
- Part VII: Scaling the features



Part VI
How do machines learn?

What is Machine Learning?

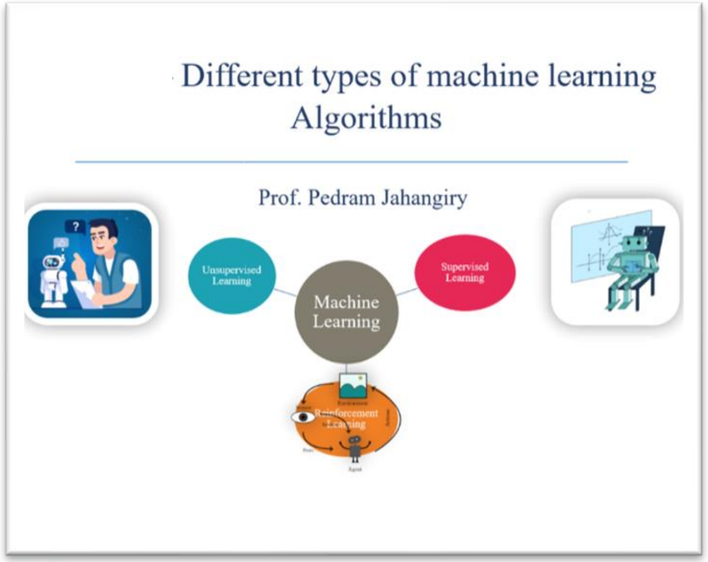
“A ML algorithm learns complex patterns in a high dimensional space without being specifically directed”

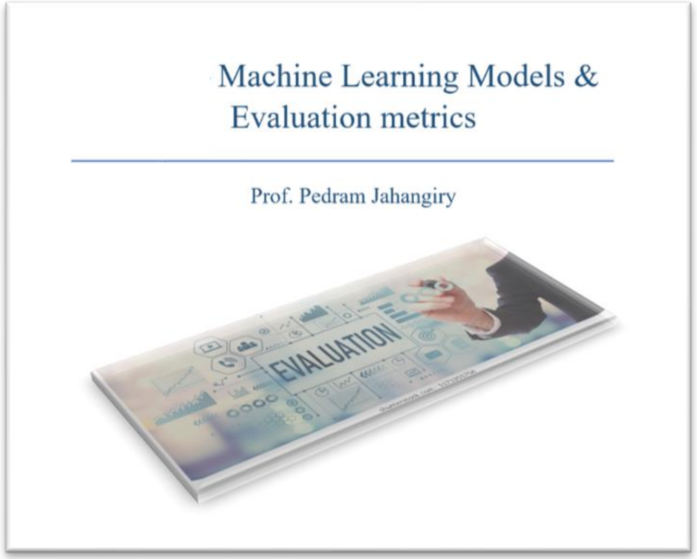


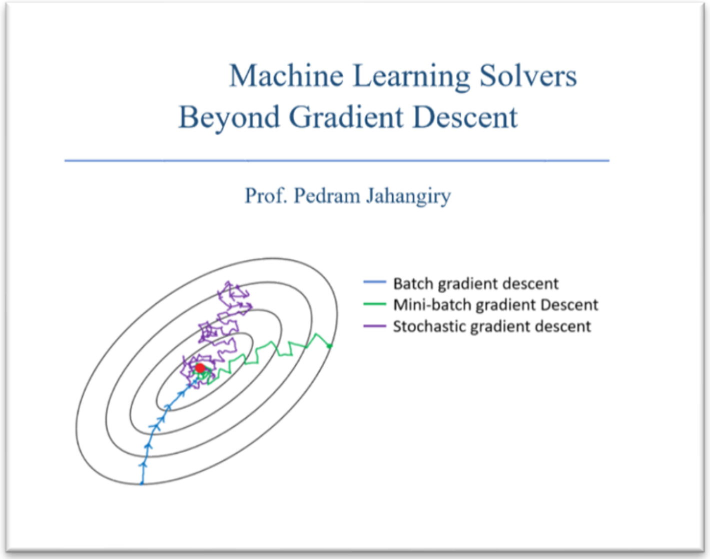
How do machines learn?

The short answer: by Algorithms!

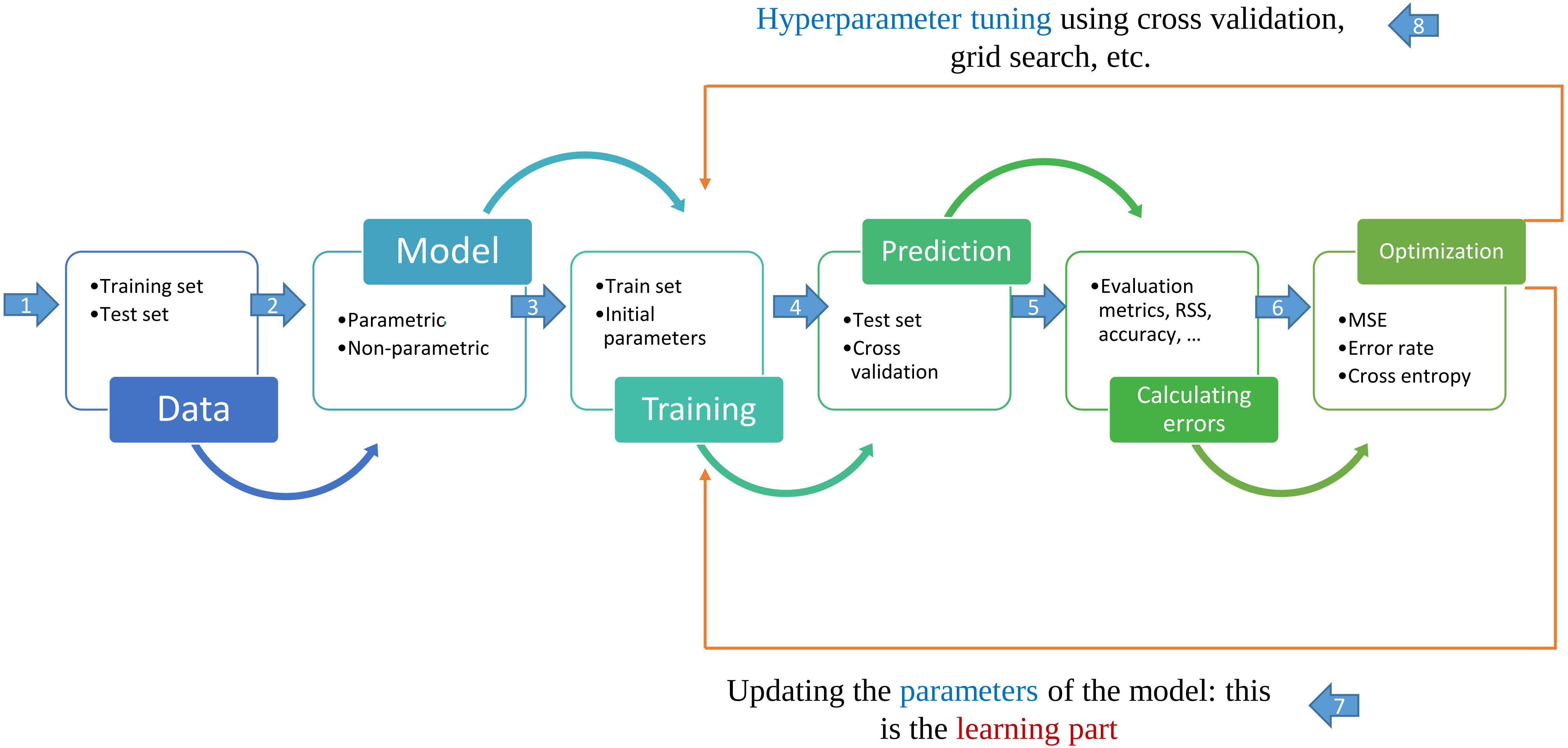
- Algorithm: a process or <u>set of rules </u>to be followed in calculations or other problem-solving operations, especially by a computer.
- Generally, the more data a machine learning algorithm is provided with, the more accurate it becomes.







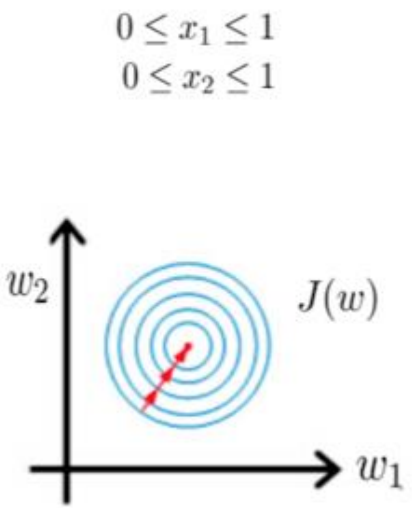
How do machines learn?



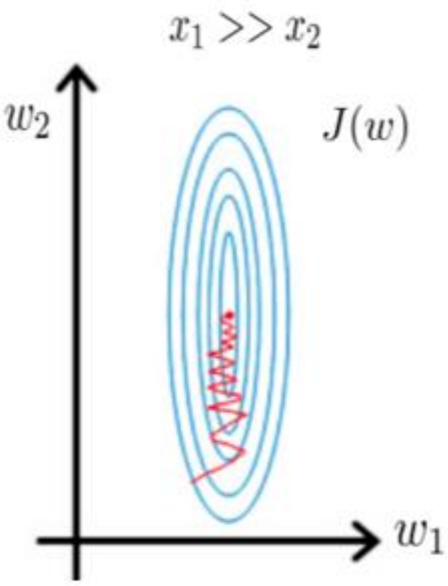
Part VII
Scaling the features!

Why feature scaling?

- Feature scaling in machine learning is a critical step during the pre-processing of data before creating a machine learning model.
- Feature scaling is essential for machine learning models that calculate distances between data.
- Feature scaling could:
    - Avoid numerical overflow and speed up the algo
    - Reduce dominant effects of specific variables



Both parameters could be updated in equal proportions



Gradient of larger parameters dominates the updates

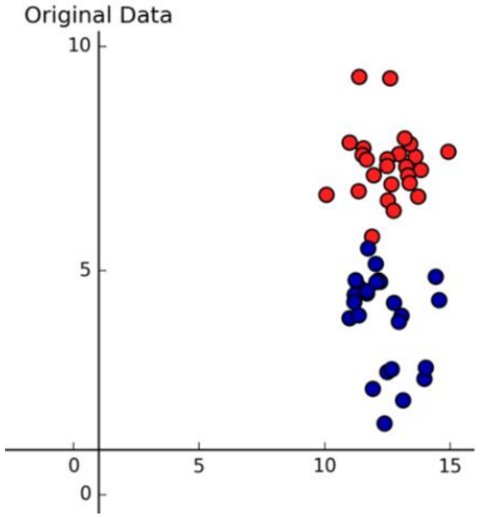
Scaling the features

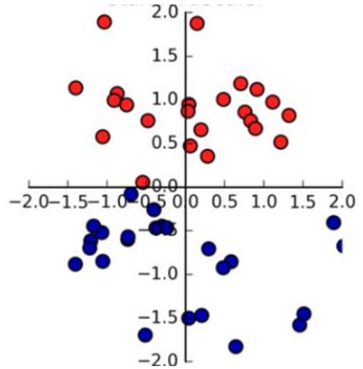
Let us use x_i for raw input and $\widetilde{x_i}$ for the transformed data.
Common scaling practices include:

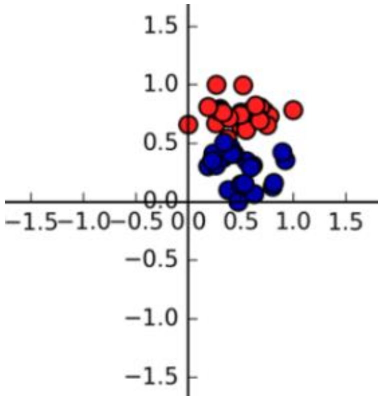
- Standardization (Z-score): $\widetilde{x_i} = \left(\frac{x_i - \mu_x}{\sigma_x}\right)$

- Normalization:
  - Min-Max scaler over [0,1]: $\widetilde{x_i} = \left(\frac{x_i - \mathrm{Min}(X)}{Max(X) - Min(X)}\right)$
  - Min-Max scaler over [-1,1]: $\widetilde{x_i} = 2 * \left(\frac{x_i - \mathrm{Min}(X)}{Max(X) - Min(X)}\right) - 1$
  - Mean normalization: $\widetilde{x_i} = \left(\frac{x_i - \mathrm{Mean}(X)}{Max(X) - Min(X)}\right)$

Scaling the features (quiz)







aim for about $-1 \le x_j \le 1$ for each feature x_j

$-3 \le x_j \le 3$

$-0.3 \le x_j \le 0.3$

} acceptable ranges

$0 \le x_1 \le 3$

$-2 \le x_2 \le 0.5$

$-100 \le x_3 \le 100$

$-0.001 \le x_4 \le 0.001$

$98.6 \le x_5 \le 105$

Which of these ranges need to be scaled?

Normalization vs Standardization

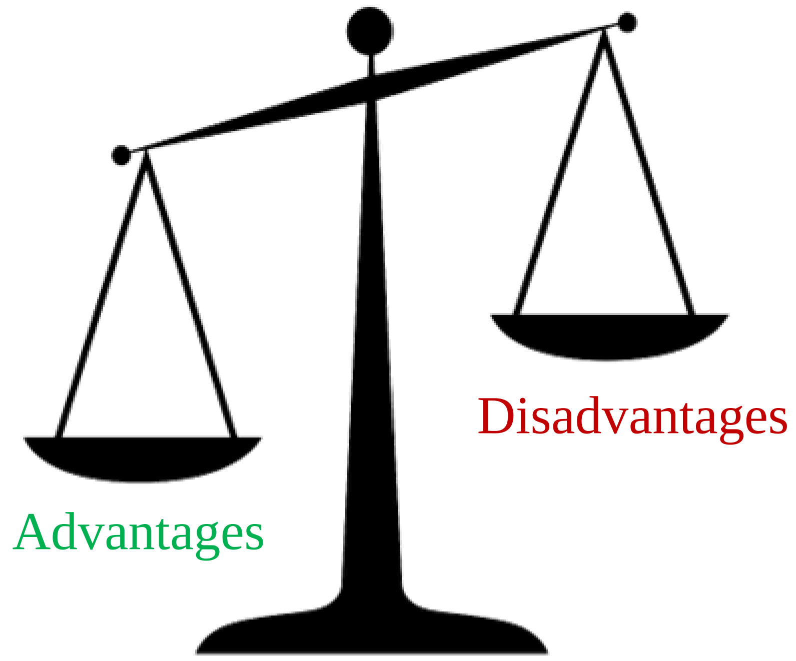
- Normalization is good to use when the distribution of the data does not follow a Normal distribution.
- Standardization, can be helpful in cases where the data follows a Normal distribution. However, this does not have to be necessarily true.
- Unlike normalization, standardization does not have a bounding range.
- The choice of using normalization or standardization will depend on your problem and the machine learning algorithm you are using

Some general hints with scaling

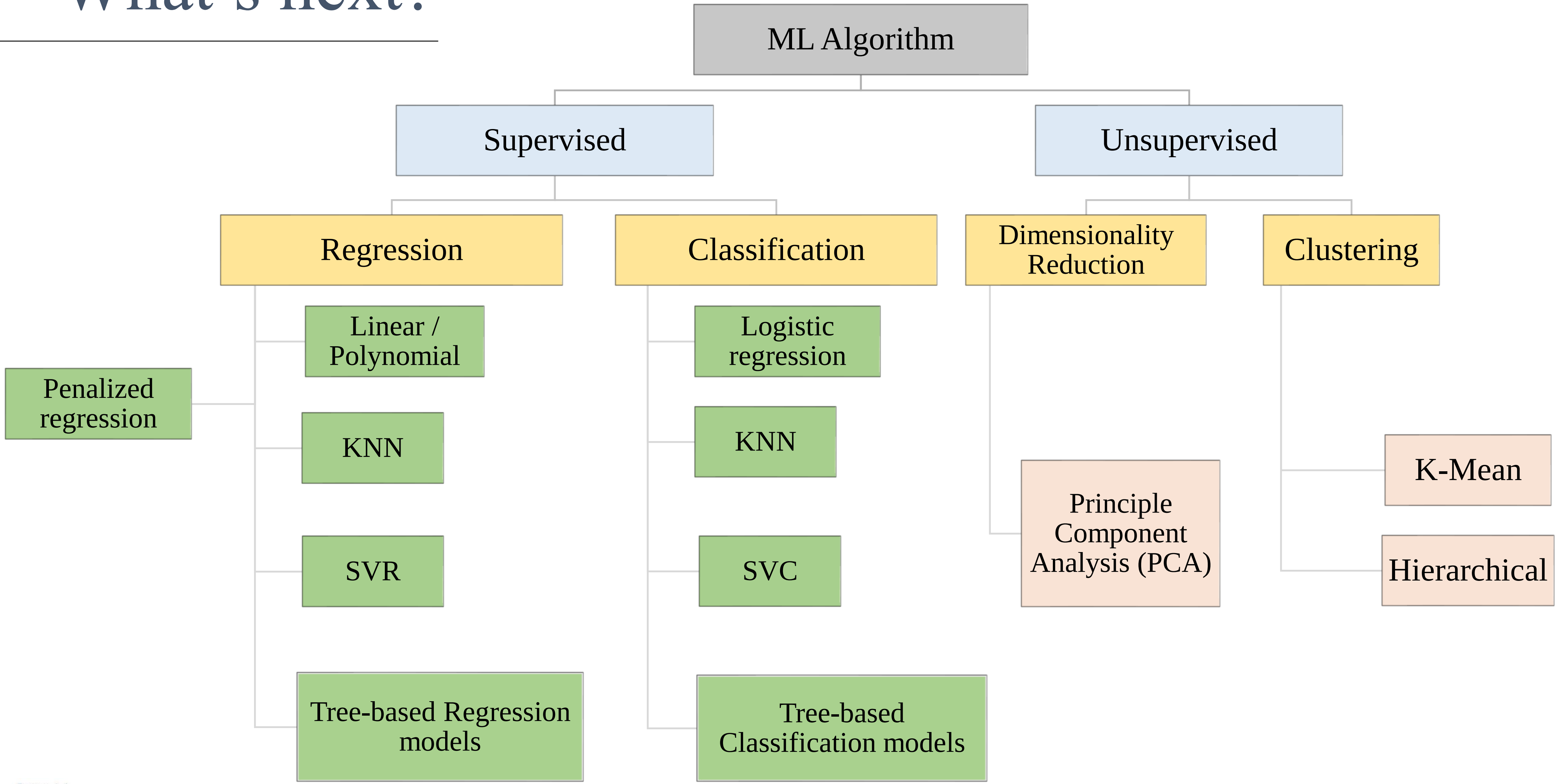
- Be careful when scaling the time series data! Why?
- To avoid data leakage, It is a good practice to fit the scaler on the training data and then use it to transform the testing data.
- Scaling the data does NOT change the shape of the distributions.

Question of the day?

What are the disadvantages of feature scaling (if any)?



What’s next?


ML Algorithm
Supervised
Unsupervised
Regression
Classification
Dimensionality Reduction
Clustering
Linear / Polynomial
Penalized regression
KNN
SVR
Tree-based Regression models
Logistic regression
KNN
SVC
Tree-based Classification models
Principle Component Analysis (PCA)
K-Mean
Hierarchical


class
DISMISSED