Machine Learning In Python

Subject: Feature Selection

Lecturer : Reza Akbari Movahed

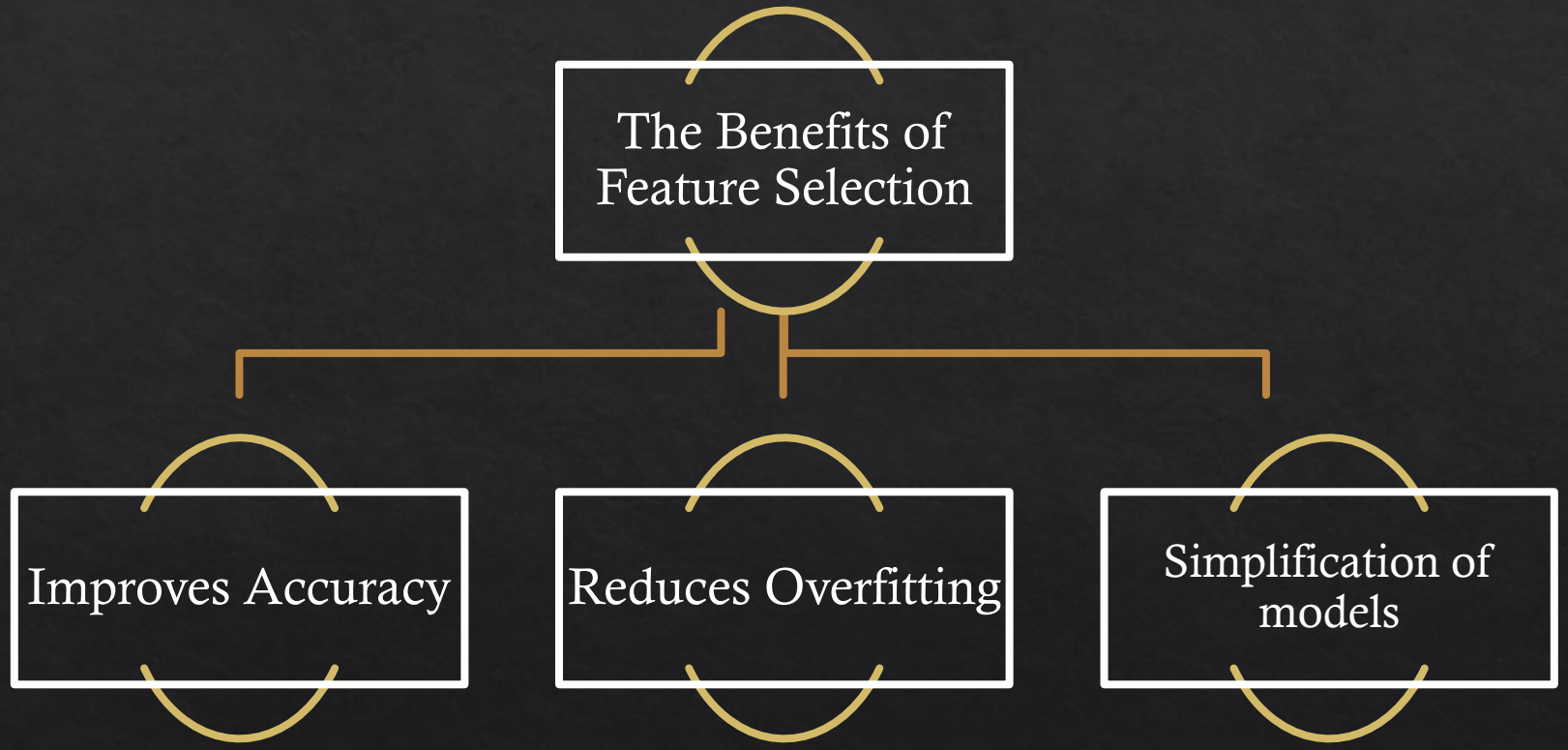
Hamedan University of Technology

Spring 2020

Feature Selection

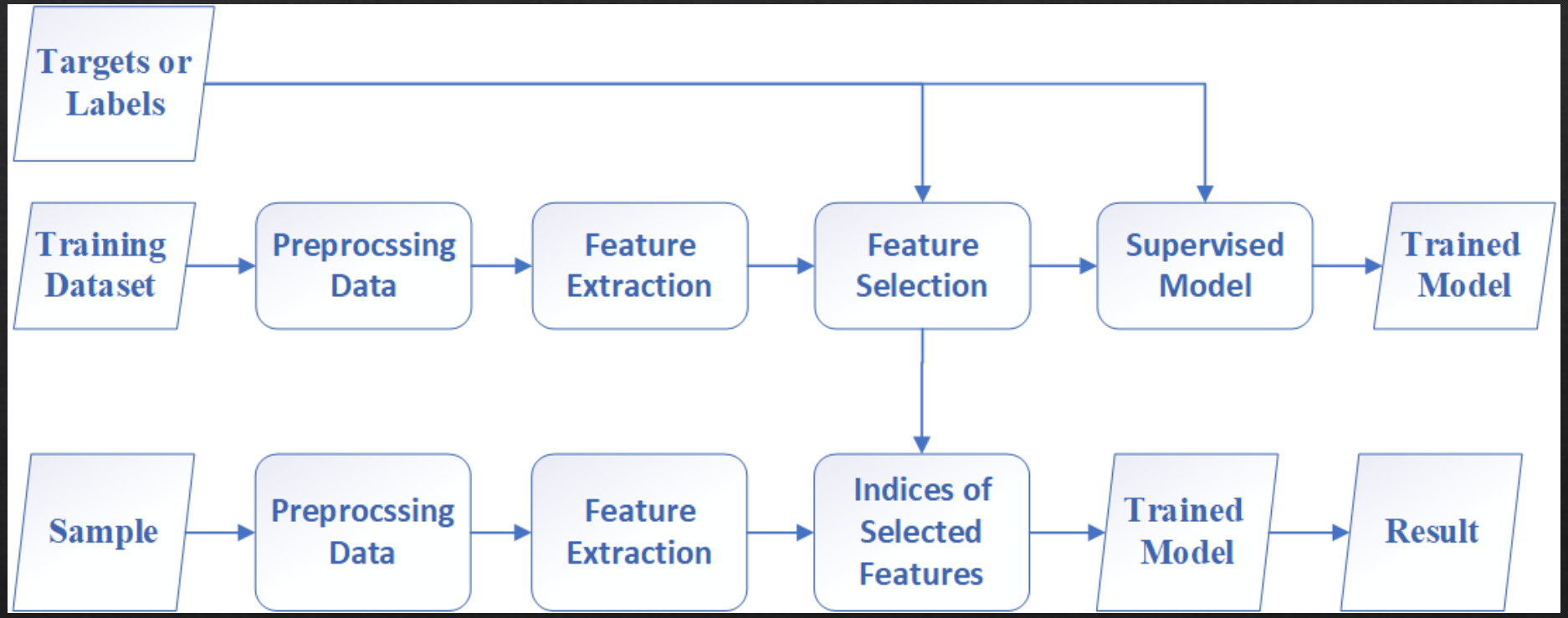
- The data features that you use to train your machine learning models have a huge influence on the performance you can achieve.
- In machine learning applications, we may face different issues such as overfitting, identifying the related features from a set of data and removing the irrelevant or less important features.
- Irrelevant features and partially relevant can negatively impact model performance.
- Feature Selection is one of the core concepts in machine learning which hugely impacts the performance of your model.
- Feature Selection is the process of selecting the best subset of features for use in model construction.

Feature Selection



Feature Selection

Supervised Learning Framework by Considering Feature Selection



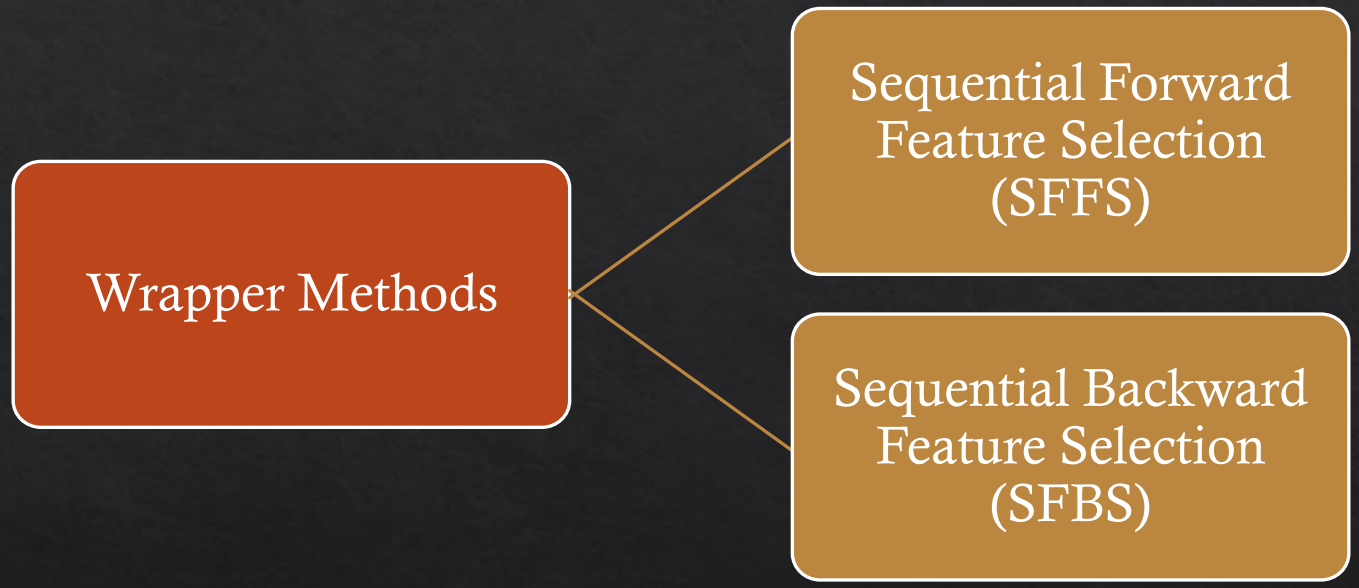
Feature Selection

Different Types of Feature Selection Methods

Filter methods	Wrapper methods	Embedded methods
Generic set of methods which do not incorporate a **specific machine learning algorithm.**	Evaluates on a **specific machine learning algorithm** to find optimal features.	Embeds (fix) features during **model building process.** Feature selection is done by observing each iteration of model training phase.
Much **faster** compared to Wrapper methods in terms of time complexity	**High computation time** for a dataset with many features	Sits **between Filter methods and Wrapper methods** in terms of time complexity
Examples – **Correlation, Chi-Square test, ANOVA, Information gain** etc.	Examples - **Forward Selection, Backward elimination, Stepwise selection** etc.	Examples - **LASSO, Elastic Net, Ridge Regression** etc.

Feature Selection

Wrapper Methods



Feature Selection

Sequential Forward Feature Selection (SFFS)

SFS is the simplest greedy search algorithm

- Starting from the empty set, sequentially add the feature x^+ that maximizes $J(Y_k + x^+)$ when combined with the features Y_k that have already been selected

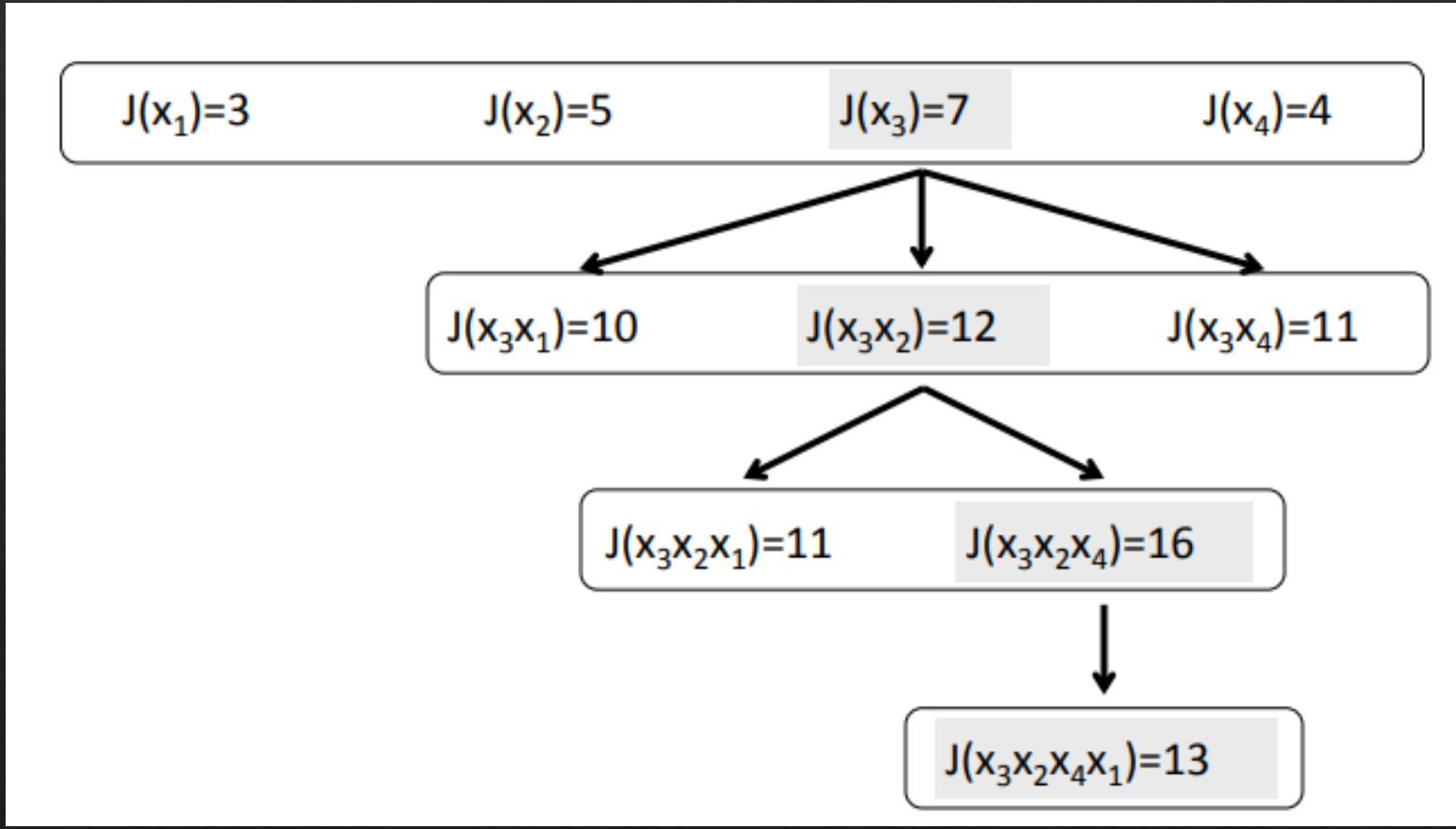
1. Start with the empty set $Y_0 = \{\emptyset\}$
2. Select the next best feature $x^+ = \underset{x \notin Y_k}{\arg\max}\, J(Y_k + x)$
3. Update $Y_{k+1} = Y_k + x^+;\ k = k + 1$
4. Go to 2

Notes

- SFS performs best when the optimal subset is small

Feature Selection

Sequential Forward Feature Selection (SFFS)



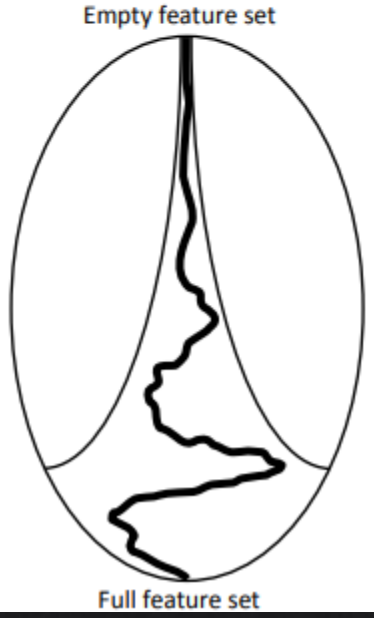
Feature Selection

Sequential Backward Feature Selection (SBFS)

SBS works in the opposite direction of SFS

- Starting from the full set, sequentially remove the feature x^- that least reduces the value of the objective function $J(Y - x^-)$
  - Removing a feature may actually increase the objective function $J(Y_k - x^-) > J(Y_k)$; such functions are said to be non-monotonic (more on this when we cover Branch and Bound)

1. Start with the full set $Y_0 = X$
2. Remove the worst feature $x^- = \arg\max_{x \in Y_k} J(Y_k - x)$
3. Update $Y_{k+1} = Y_k - x^-;\ k = k + 1$
4. Go to 2



Notes

- SBS works best when the optimal feature subset is large, since SBS spends most of its time visiting large subsets