UNIVERSITÀ
di VERONA

Machine Learning

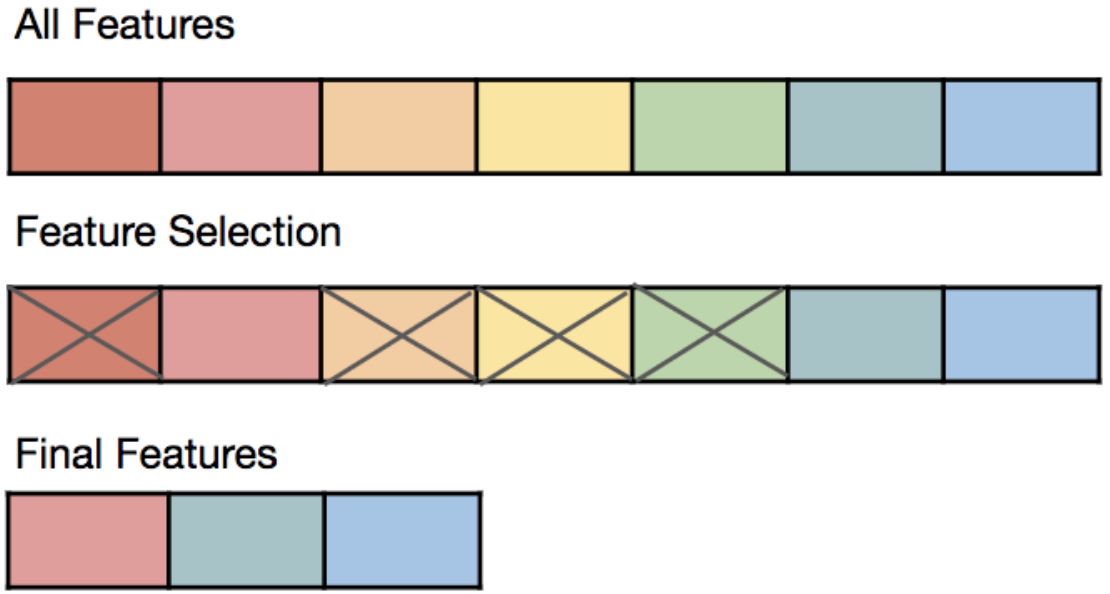
Feature Selection (extra Material)

Cigdem Beyan

A. Y. 2024/2025

Feature selection

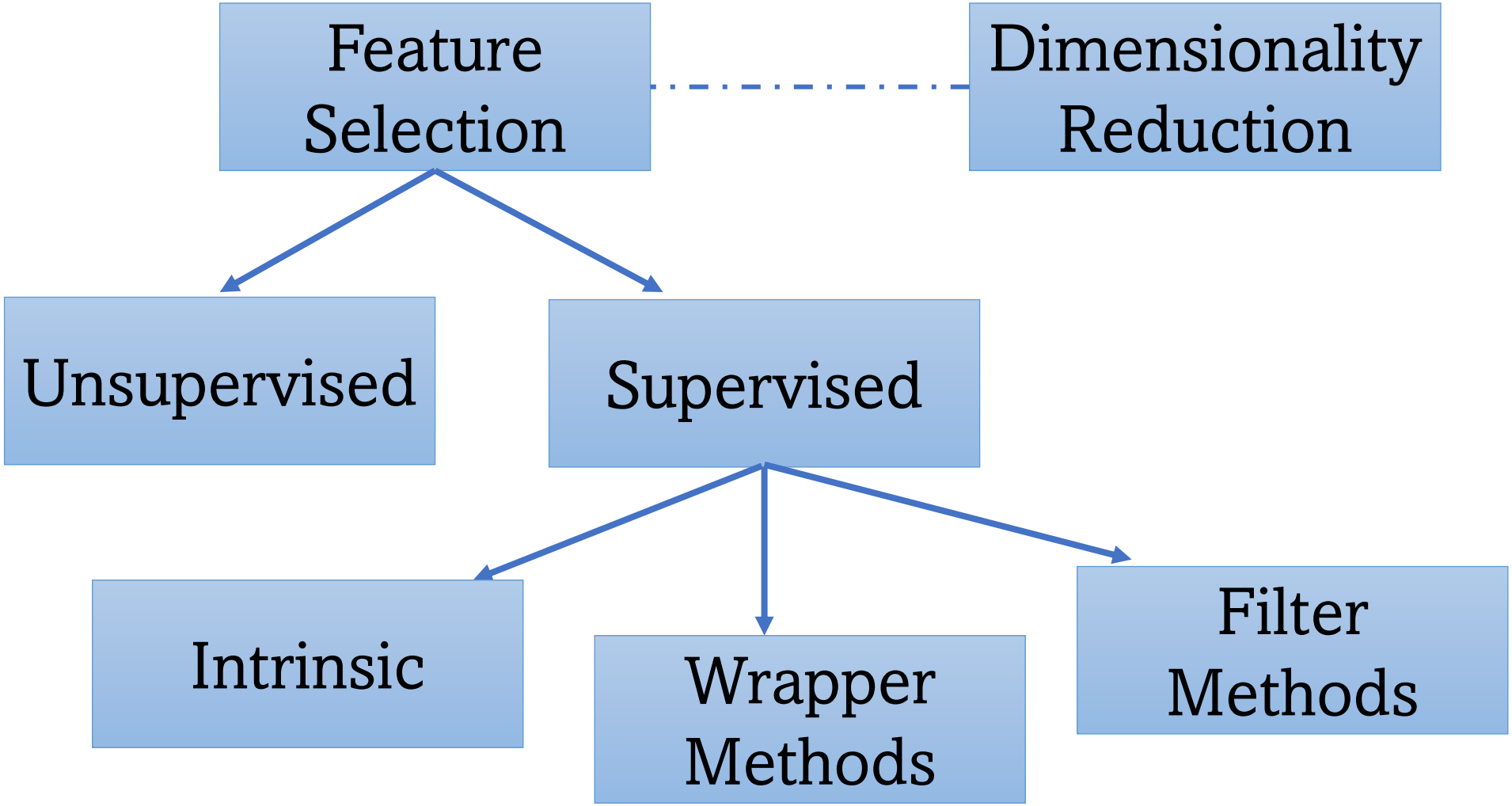
- The process of selecting the input variable to your model by using only relevant data and getting rid of “noise” in data.
- Because, the noisy (irrelevant) attributes can mislead your model, thus decrease its performance.



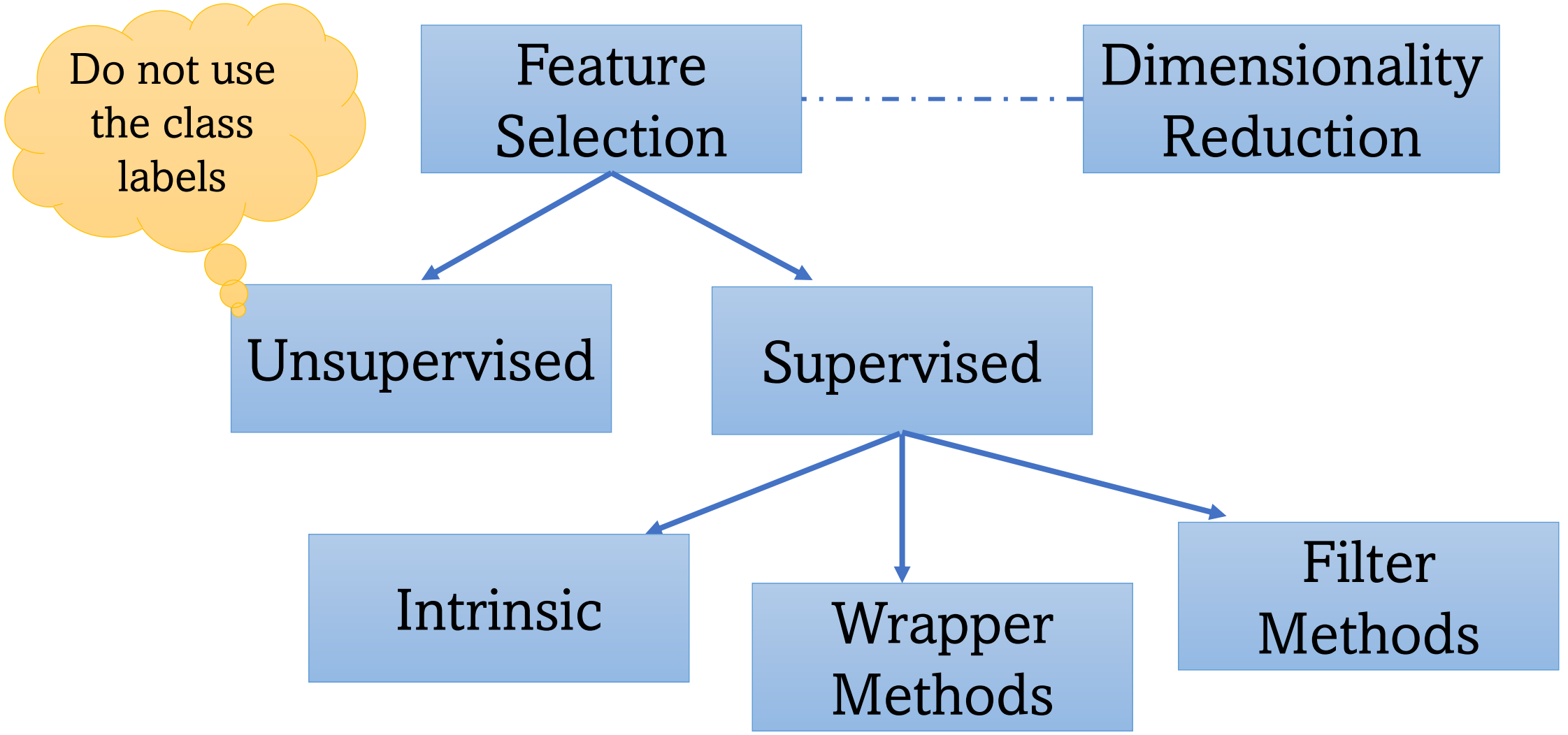
Feature selection

- High dimensional data suffers from: *Curse of Dimensionality*
- Observations in a high-dimensional space are more sparse and less representative than those in a low-dimensional space.
- Using feature selection, we can optimize our model in several ways:
  - Prevent learning from noise (overfitting)
  - Improved performance, e.g., accuracy
  - Reduce training time (more features, typically means more training time)

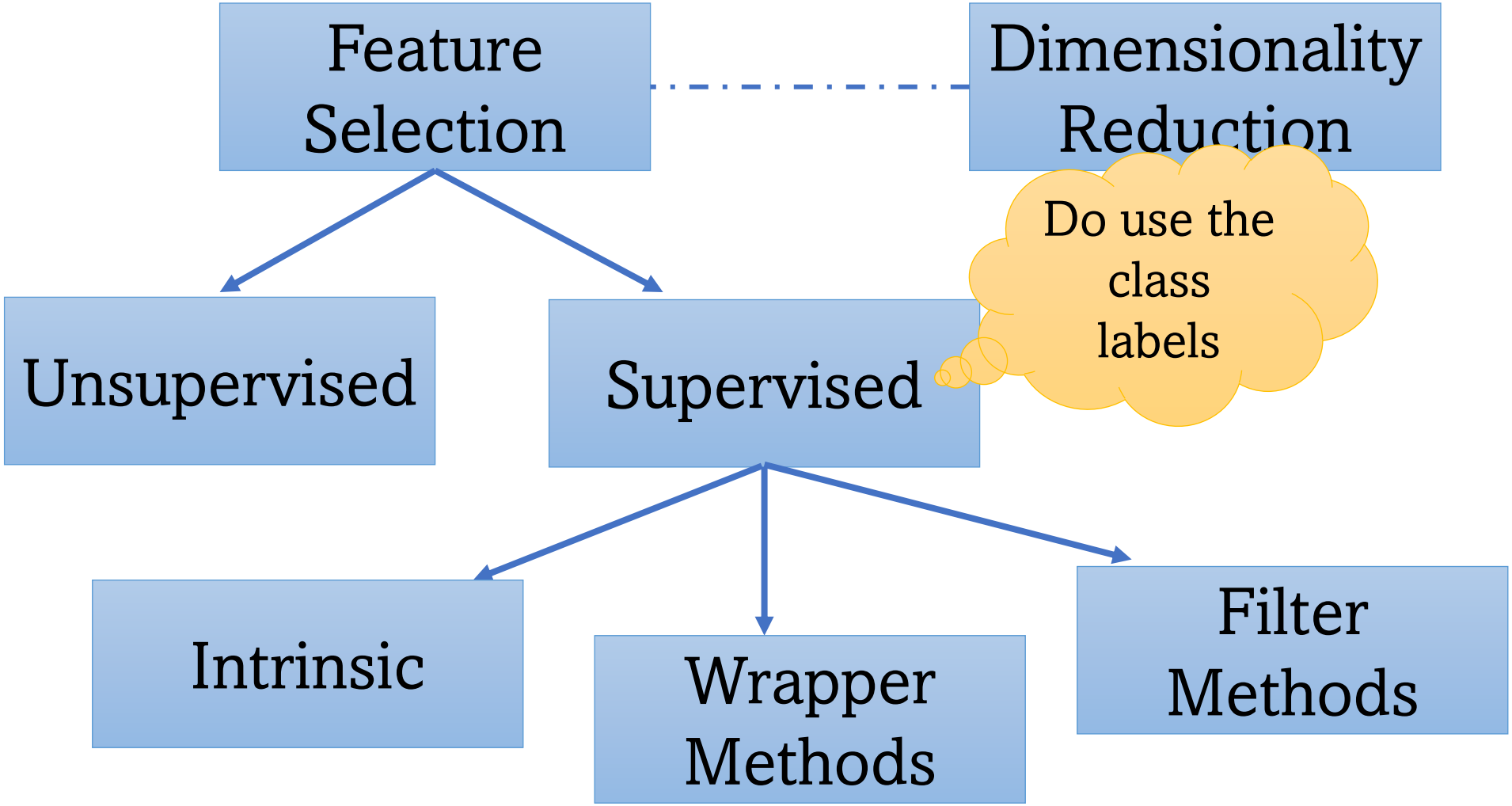
Feature selection methods



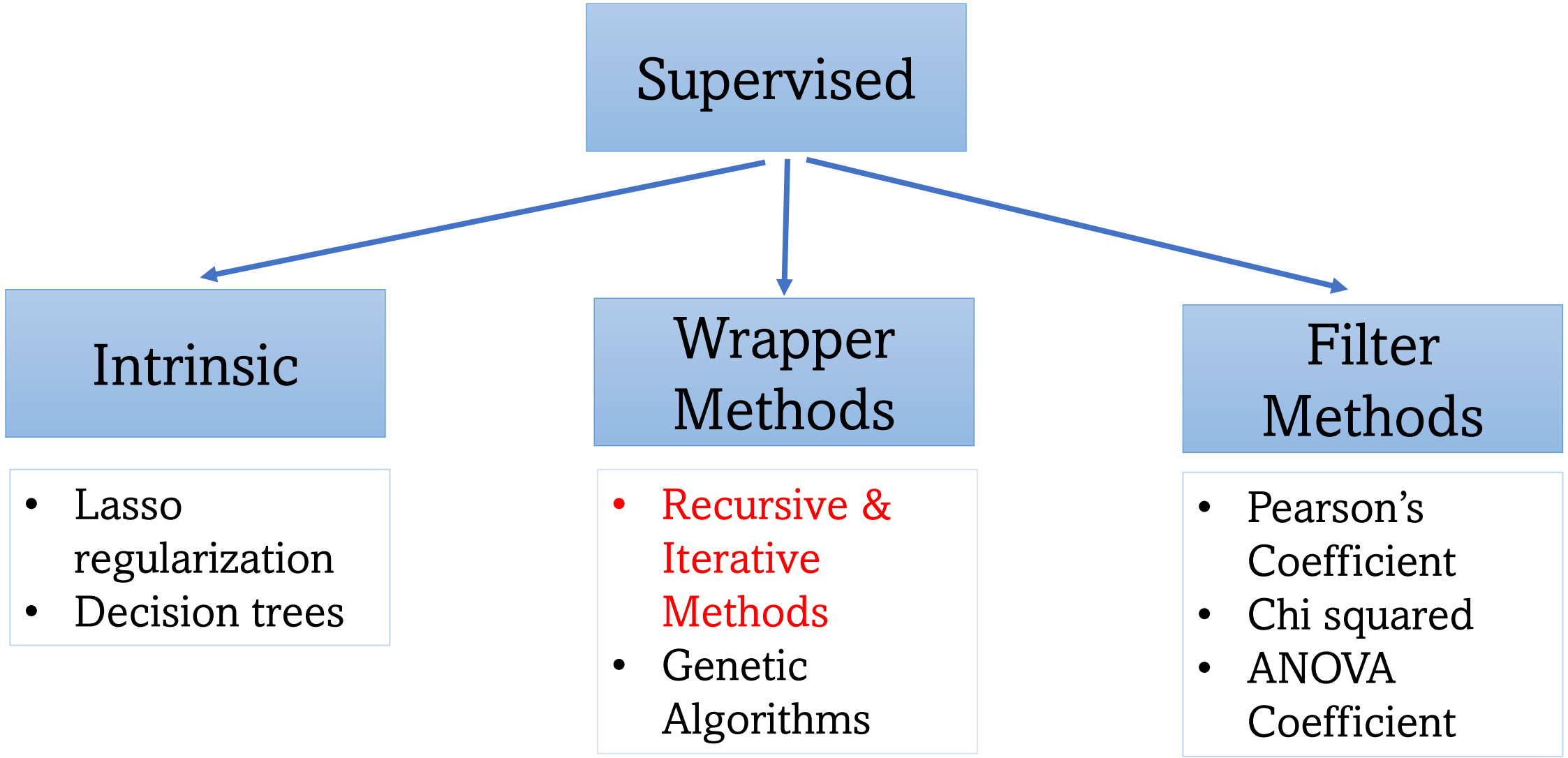
Feature selection methods



Feature selection methods



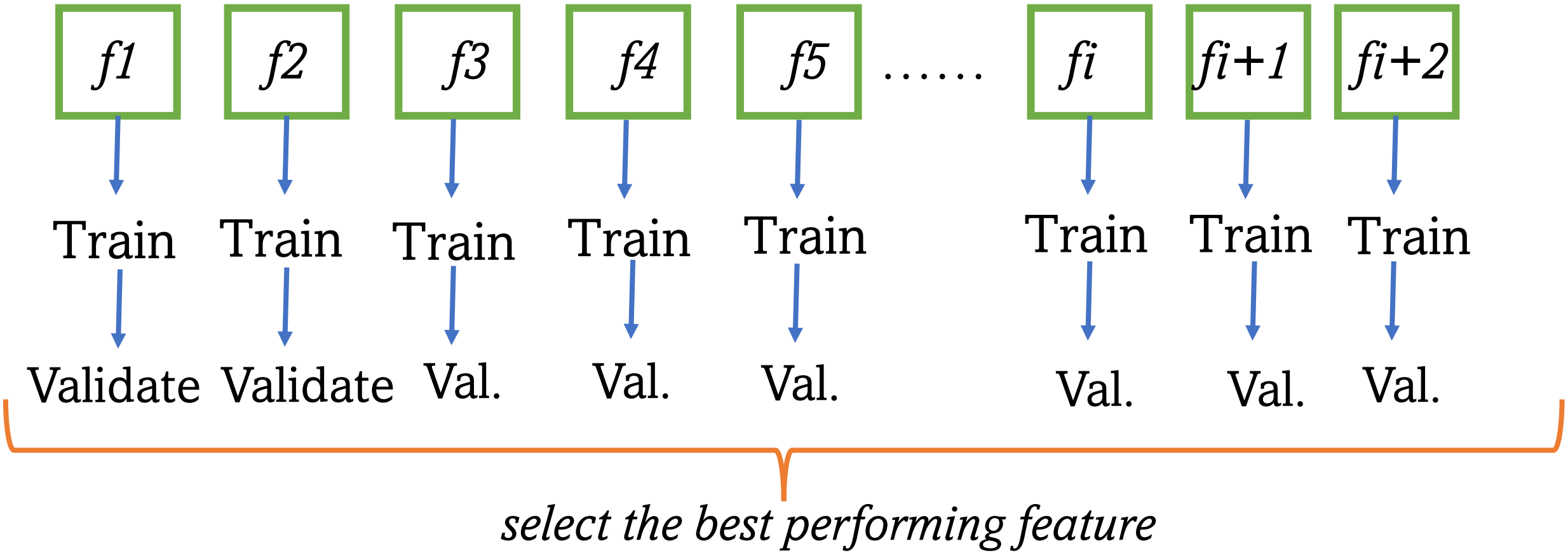
Feature selection methods



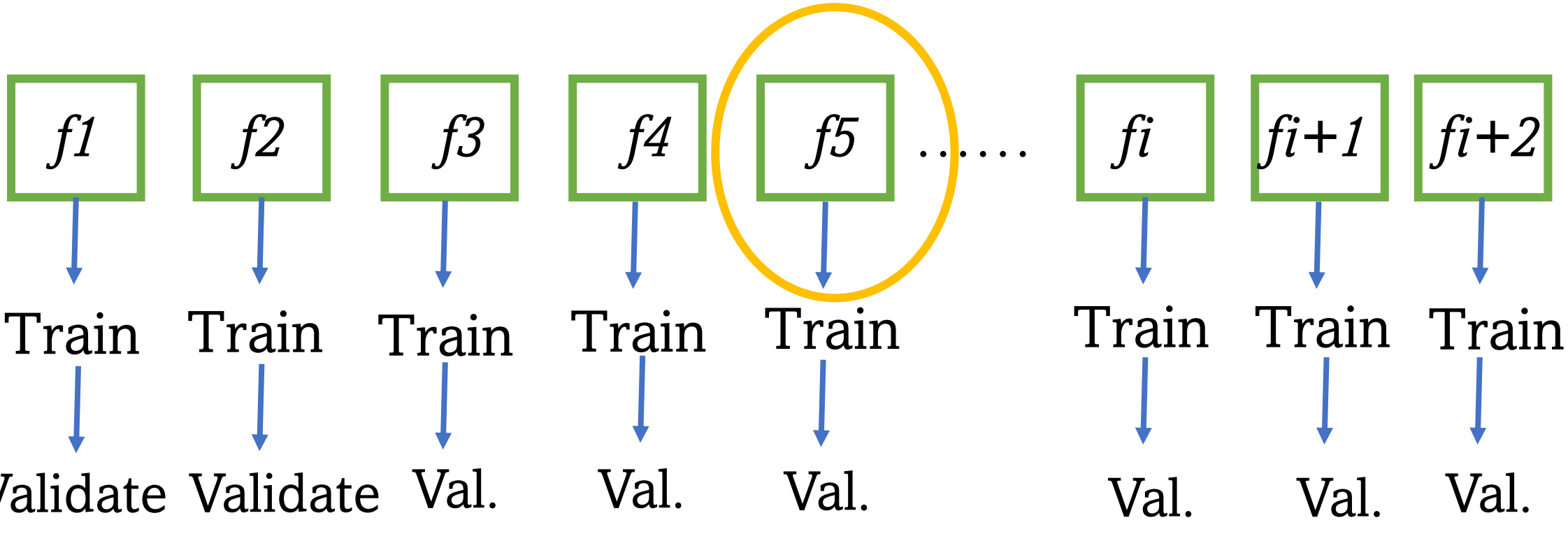
Forward feature selection

- An iterative method in which we start with having a single feature in the model.
- In each iteration, we keep adding the feature which improves our model the most, till an addition of a new variable does not improve the performance of the model.

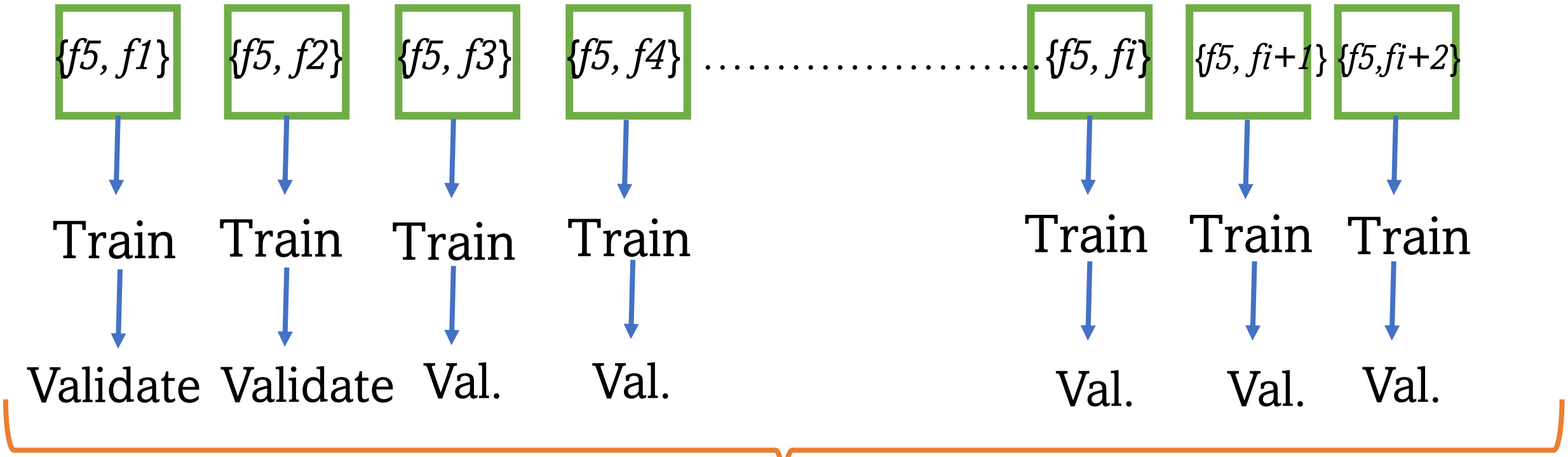
Forward feature selection – iteration 1



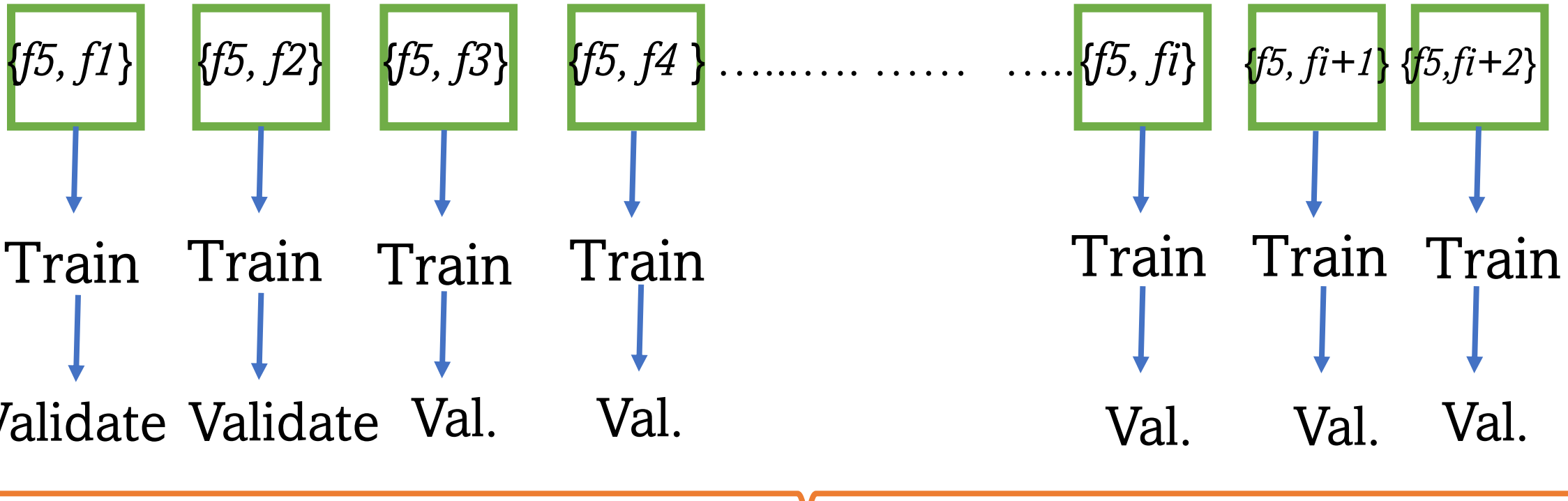
Forward feature selection – iteration 1



Forward feature selection-Iteration 2



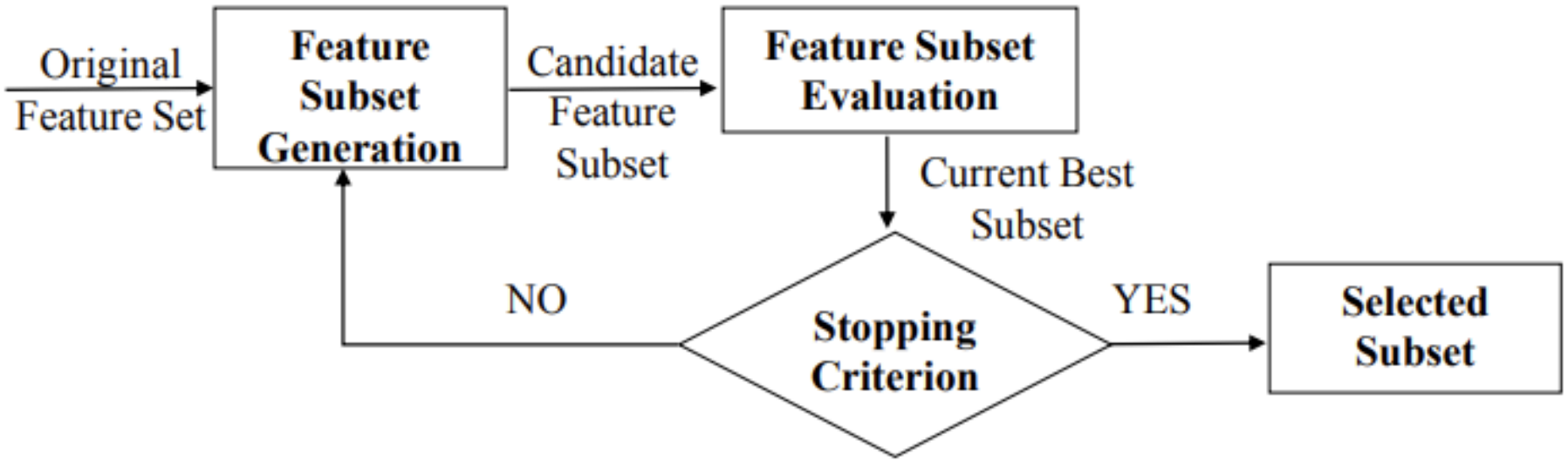
Forward feature selection-Iteration 2



select the best performing feature

if: performance(f5) << performance(f5,new) continue iteration, otherwise stop!

Forward feature selection



BACKWARD FEATURE ELIMINATION

- An iterative method in which we start with all features, and we remove the least significant feature at each iteration such that removing it increases (rarely not changes) the performance of the model. We repeat this until no improvement is observed on removal of features.

Coding tutorial

UNIVERSITÀ di VERONA

- Feature Selection: https://scikit-learn.org/stable/modules/feature_selection.html

1.13.5. Sequential Feature Selection