

Lab 3

Feature selection



Escola Tècnica Superior d'Enginyeria
de Telecomunicació de Barcelona

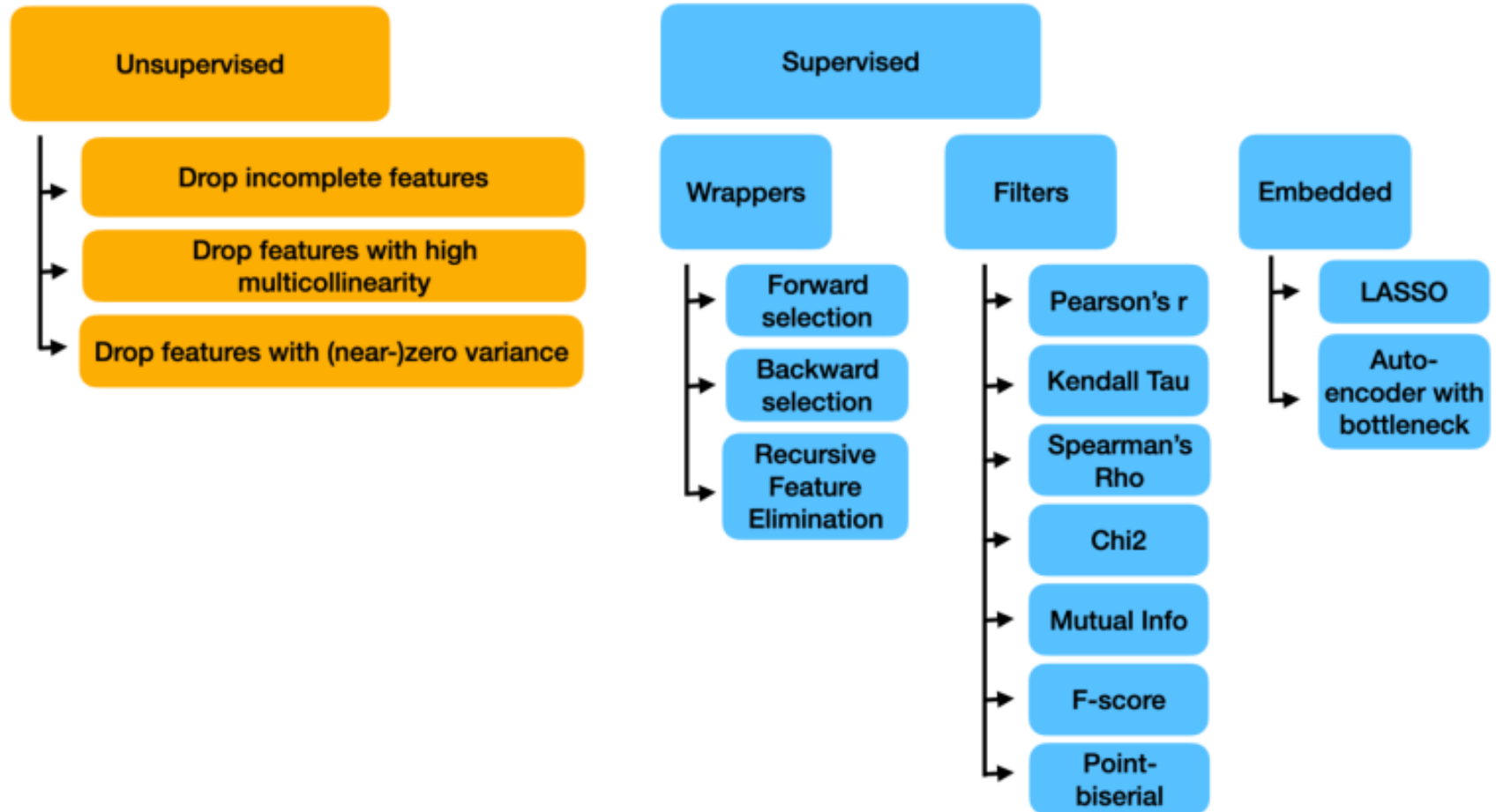
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Feature selection

- Feature Selection is the process of selecting a subset of relevant features for use in machine learning model building.
- It is not always the truth that the more data, the better the result will be.
- Including irrelevant features (the ones that are just unhelpful to the prediction) and redundant features (irrelevant in the presence of others) will only make the learning process overwhelmed and easy to cause overfitting.
- By employing feature selection, we can:
 - Simplify models: Leading to improved interpretability.
 - Reduce training time and computational cost.
 - Lower data collection costs.
 - Mitigate the curse of dimensionality.
 - Enhance generalization by reducing overfitting.

Feature selection techniques

Feature selection methods



Supervised Feature selection techniques

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
Less prone to over-fitting	High chances of over-fitting because it involves training of machine learning models with different combination of features	Generally used to reduce over-fitting by penalizing the coefficients of a model being too large.
Examples – Correlation, Chi-Square test, ANOVA, Information gain etc.	Examples - Forward Selection, Backward elimination, Stepwise selection etc.	Examples - LASSO, Elastic Net, Ridge Regression etc.