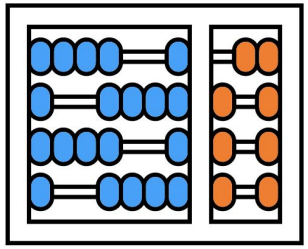


Instituto de Computação

UNIVERSIDADE ESTADUAL DE CAMPINAS

Capacitação profissional em tecnologias de Inteligência Artificial

Machine Learning Overview

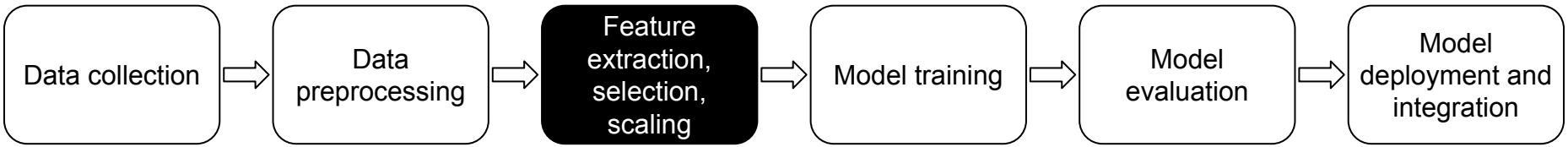
Prof. Edson Borin

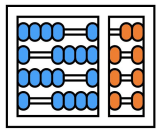
https://www.ic.unicamp.br/~edson

Institute of Computing - UNICAMP

ML Process

Feature extraction/selection/scaling

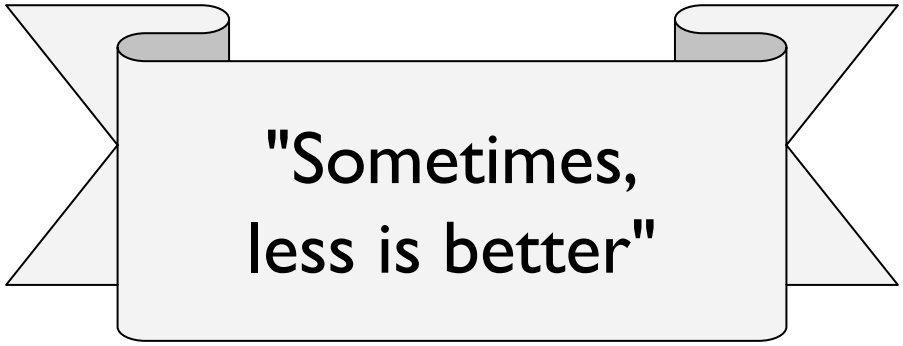


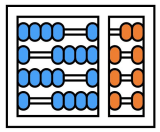


ML Process - Feature extraction/selection

Selecting the proper features to train a machine learning model:

- enables the machine learning algorithm to train faster
- reduces the complexity of a model and makes it easier to interpret
- improves the accuracy of a model if the right subset is chosen
- reduces overfitting





ML Process - Feature extraction/selection

Selecting the proper features to train a machine learning mod

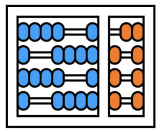
- enable ine learning algorithm to train
- it
- subset is chosen
- reduces overfitting

Main goal

have features that are

informative and **non-redundant**

"Sometimes,
less is better"



ML Process - Feature extraction/selection

Feature selection: identify the features that are most useful for the problem under examination

- Ex: find out that the *size* feature is more useful to predict a house *price* than the *color* feature

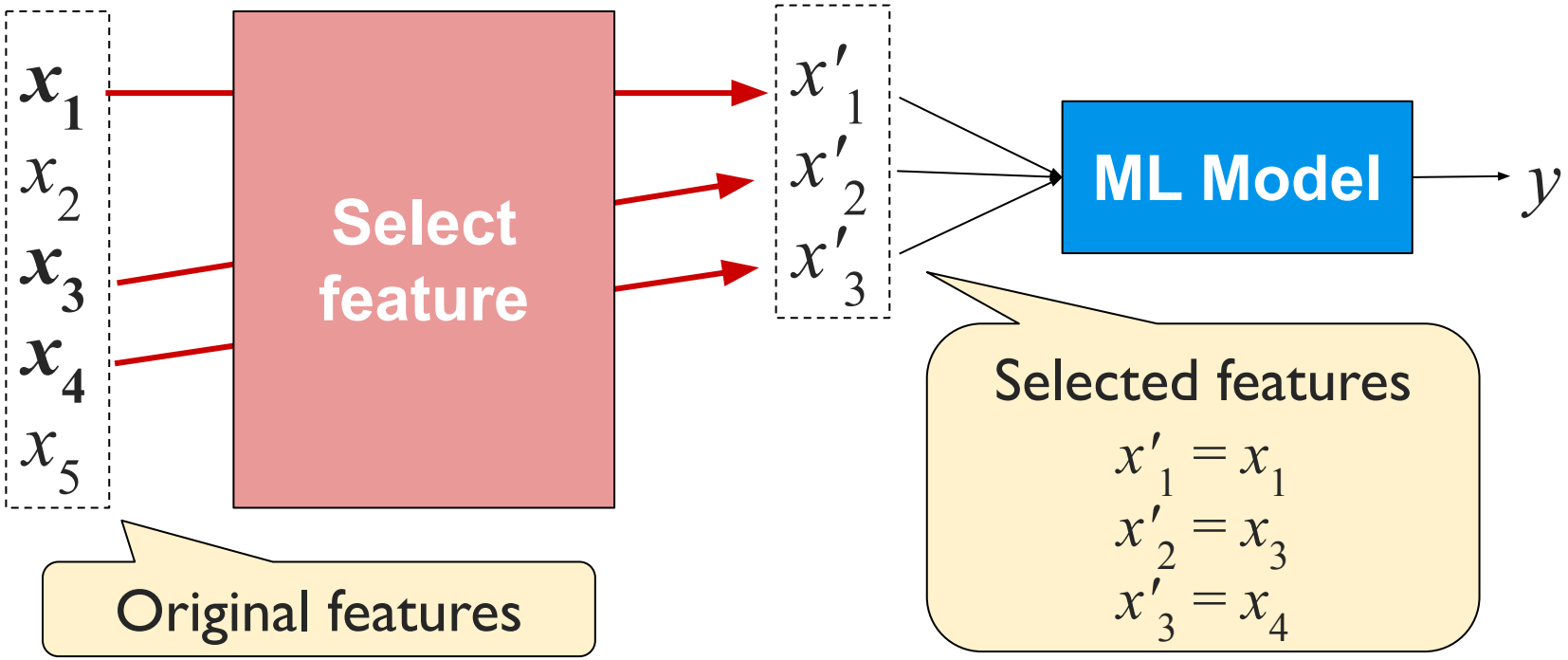
Feature extraction: Derive (create) new features based on the existing ones

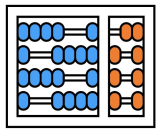
- Ex: transform individual features (e.g., square them), or combine multiple features (e.g. by multiplying them) to create new features

ML Process - Feature extraction/selection

Feature selection: identify the features that are most useful for the problem under examination

Approach using feature selection



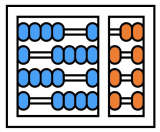


ML Process - Feature extraction/selection

Feature selection: identify the features that are most useful for the problem under examination

Common approaches for feature selection:

- Filter methods
- Wrapper methods
- Embedded methods



ML Process - Feature extraction/selection

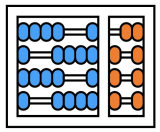
Filter methods: features are selected based on statistical test scores that measure their correlation with the outcome variable (label)

Common methods:

- Pearson correlation coefficient
- Chi-square coefficient
- Mutual information

Limitations:

- Relationship between features is not considered => tend to select redundant features



ML Process - Feature extraction/selection

Filter methods: features are selected based on statistical test scores that measure their correlation with the outcome variable (label)

Common methods:

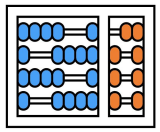
- Pearson correlation coefficient
- Chi-square coefficient
- Mutual information

Limitations:

- Relationship between features is select redundant features

Model agnostic

Selection of features is independent of ML algorithm.



ML Process - Feature extraction/selection

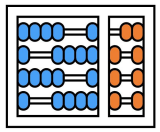
Wrapper methods: measure the usefulness of a subset of feature by actually training a model on it

Common methods:

- Forward Selection
- Backward Elimination
- Recursive feature elimination (RFE)

Limitations:

- Requires training a new model for each subset => huge number of computations
- Best feature set is usually associated with a specific type of model



ML Process - Feature extraction/selection

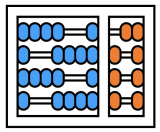
Embedded methods: consider feature selection as a part of the model training.

Example: Regularization method

- Lasso regression: $\text{Cost Func.} = \text{LSq}_\theta(X,Y) + \sum |\boldsymbol{\theta_i}|$

Limitations:

- Specific to model (e.g., embedded into the cost function)

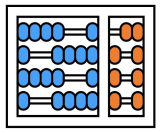


ML Process - Feature extraction/selection

Feature selection: Main benefits

Benefits:

- Simplify models to make them easy for users to interpret
- Reduce the model training time
- Avoid dimension explosion
- Improve model generalization and avoid overfitting



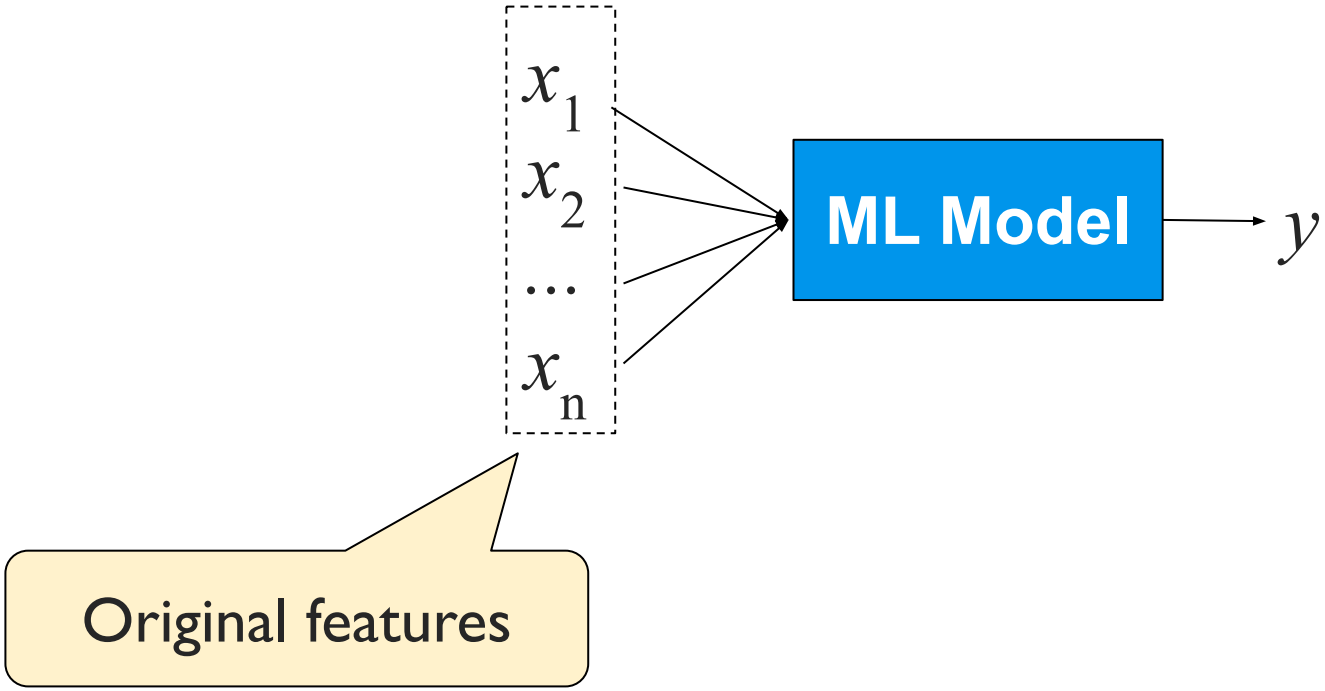
ML Process - Feature extraction/selection

Feature extraction: Derive (create) new features based on the existing ones

ML Process - Feature extraction/selection

Feature extraction: Derive (create) new features based on the existing ones

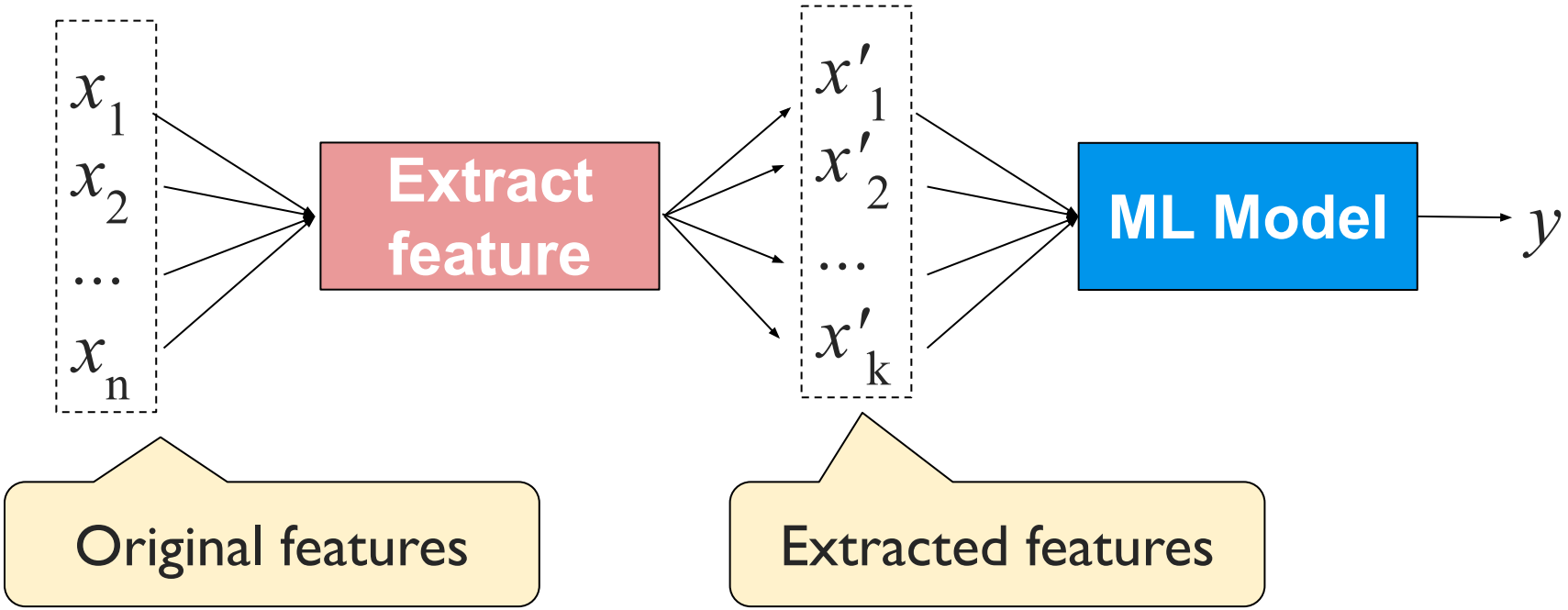
Approach without feature extraction

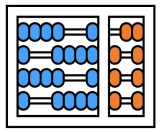


ML Process - Feature extraction/selection

Feature extraction: Derive (create) new features based on the existing ones

Approach using feature extraction



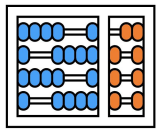


ML Process - Feature extraction/selection

Feature extraction: Derive (create) new features based on the existing ones

Approaches for feature extraction:

- Create new features
  - Ex: Feature cross, kernel-tricks…
- Dimensionality reduction techniques
  - Focus on reducing dimensionality by combining existing features into a smaller set of new features
  - Ex: Isomap, UMAP, PCA, Autoencoder, …

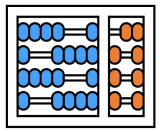


ML Process - Feature extraction/selection

Feature scaling:

Many ML algorithms do not perform well when the input numerical attributes have very different scales

- Ex: #bedrooms vs. house price



ML Process - Feature extraction/selection

Feature scaling:
Many ML algorithms do not perform well when the input numerical attributes have very different scales

- Ex: #bedrooms vs. house price

Two common approaches

- min-max scaling (a.k.a. normalization)
  - Normalize data to the interval [0-1]
- standardization
  - Subtract the mean value and divide by the standard deviation

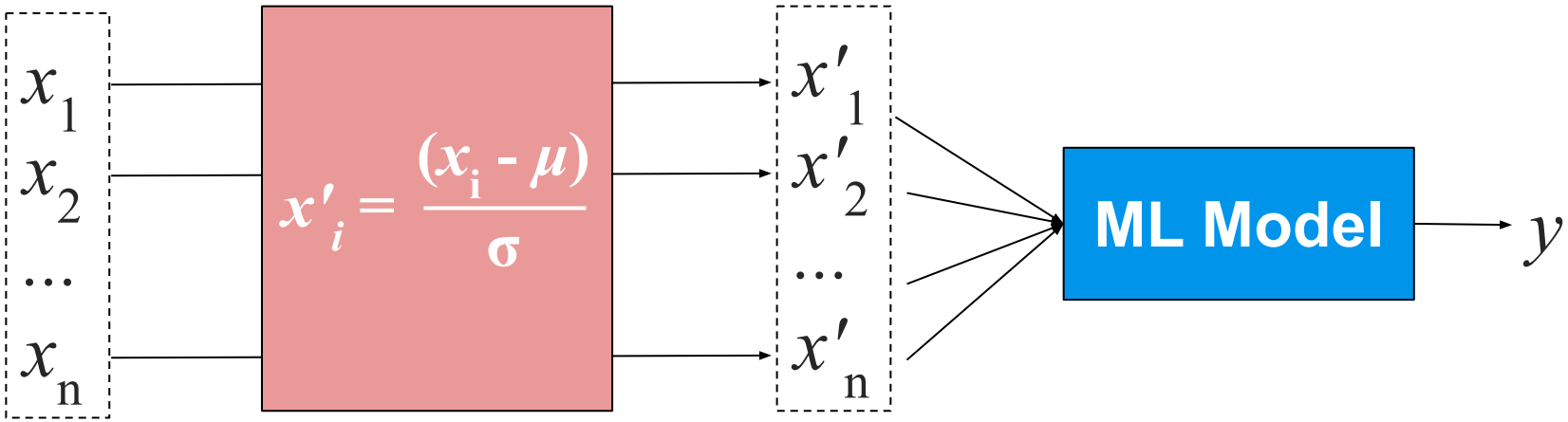
ML Process - Feature extraction/selection

Feature scaling:

Many ML algorithms do not perform well when the input numerical attributes have very different scales

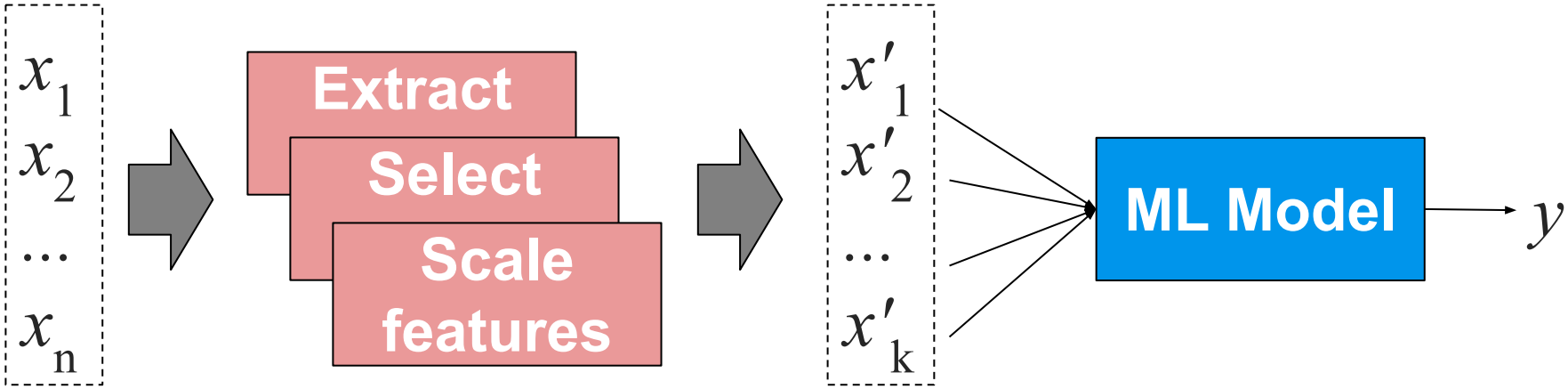
- Ex: #bedrooms vs. house price

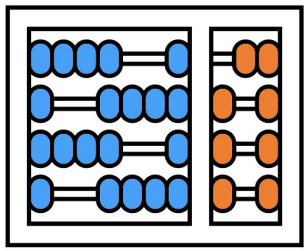
Approach using feature scaling (standardization)



ML Process - Feature extraction/selection

Feature extraction/selection/scaling: Techniques to improve the set features used on the machine learning model!





Instituto de Computação

UNIVERSIDADE ESTADUAL DE CAMPINAS

Capacitação profissional em tecnologias de Inteligência Artificial

Machine Learning Overview

Prof. Edson Borin

https://www.ic.unicamp.br/~edson

Institute of Computing - UNICAMP