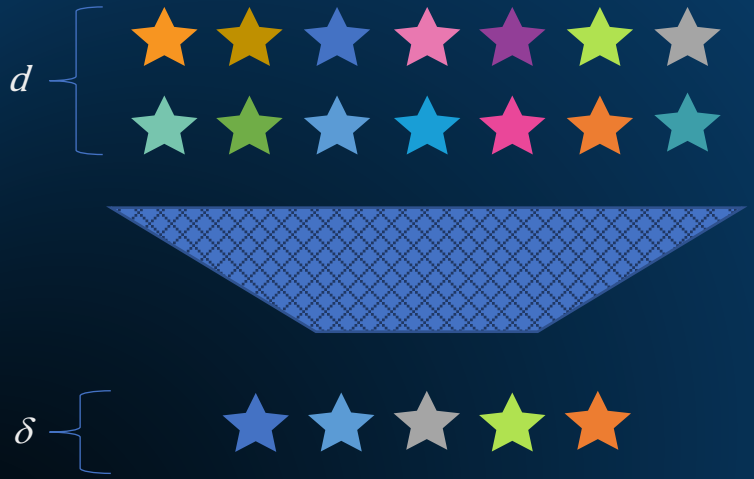
Feature Selection

Cláudia Antunes

Instituto Superior Técnico – Universidade de Lisboa

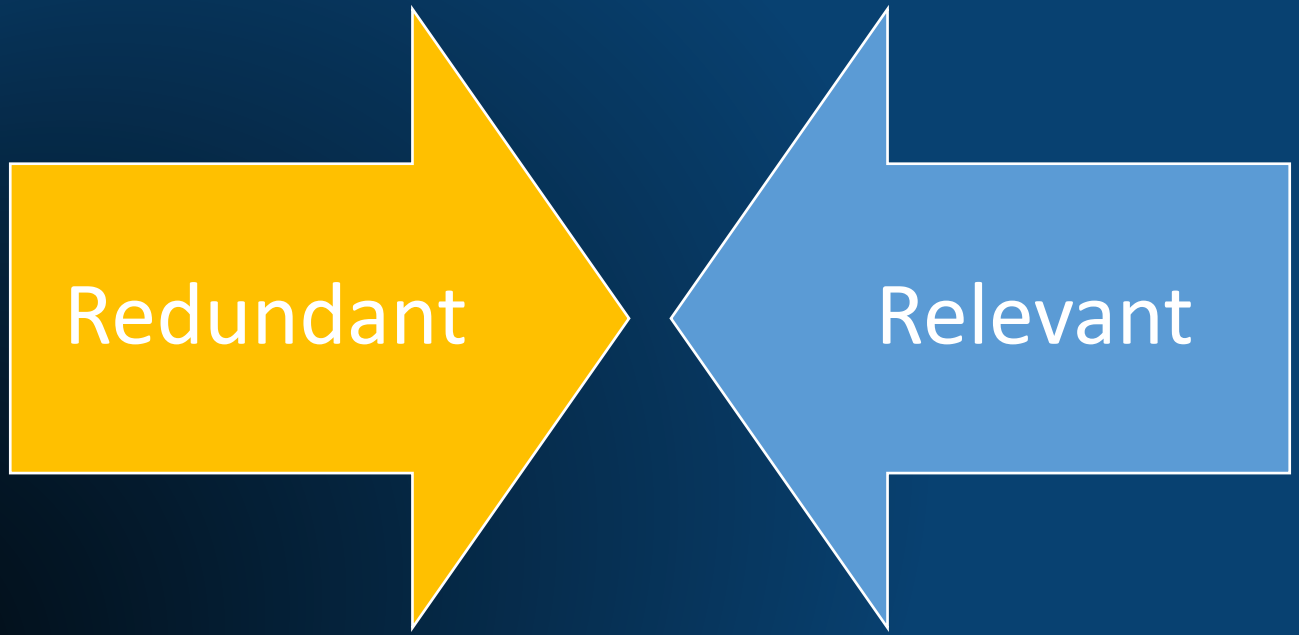
Feature Selection



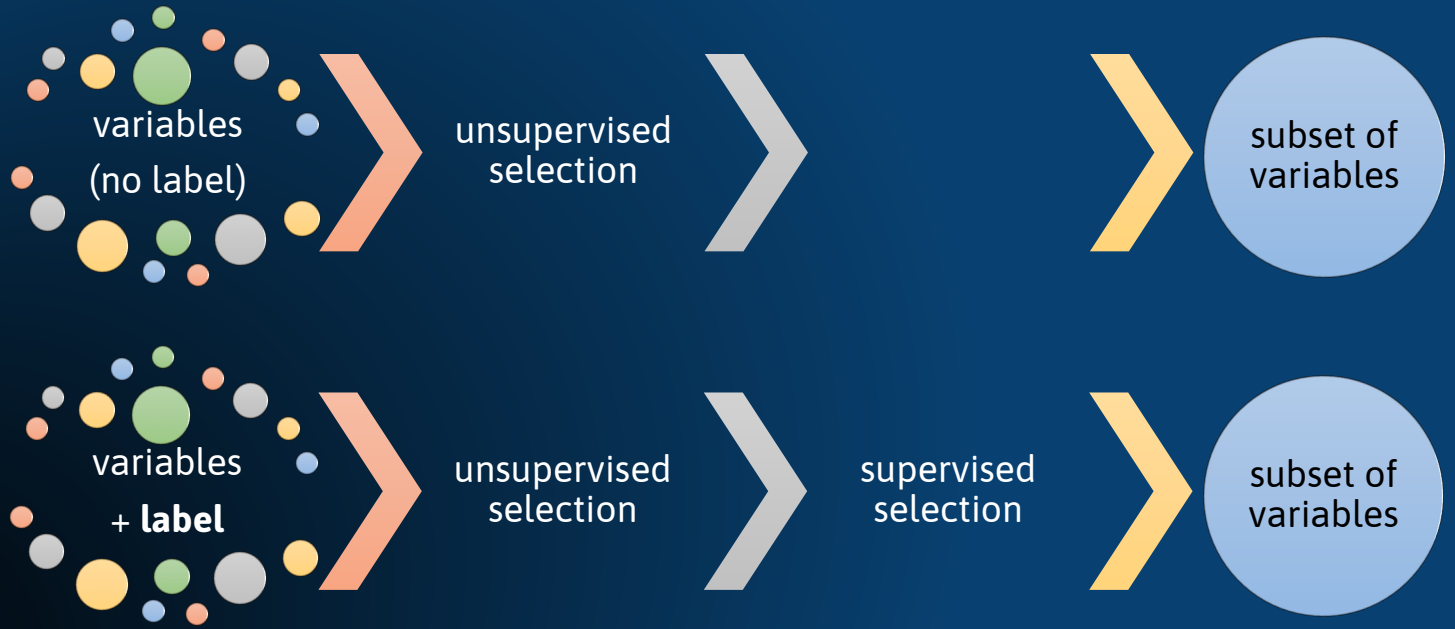
$$v_1, v_2, \dots, v_d$$

$$v_{i1}, v_{i2}, \dots, v_{i\delta}$$
$$with\ \delta \leq d$$

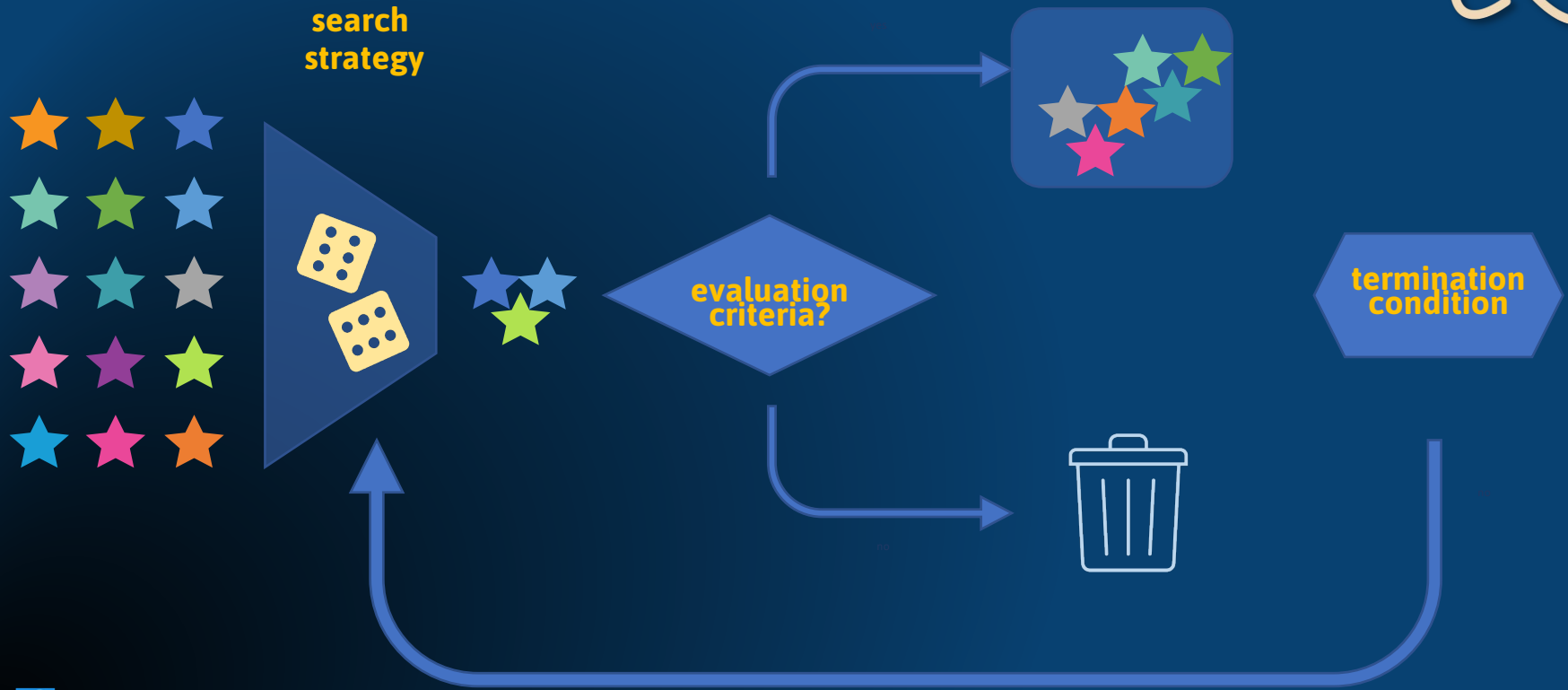
TWO NOTIONS



Supervised x Unsupervised

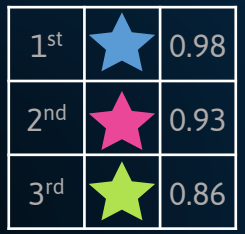


FEATURE SELECTION

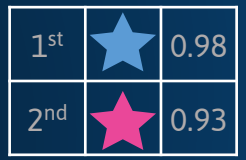


Termination Condition

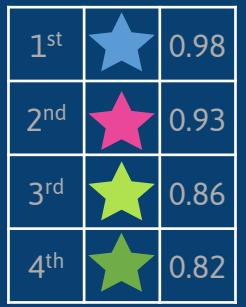
k=3



80th percentile

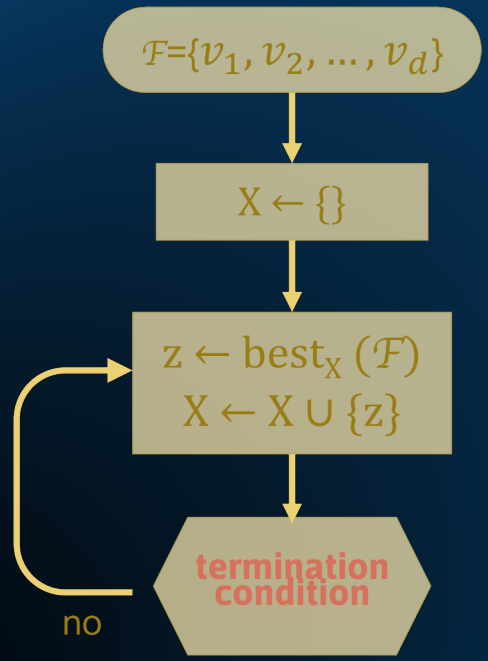


score > 0.8

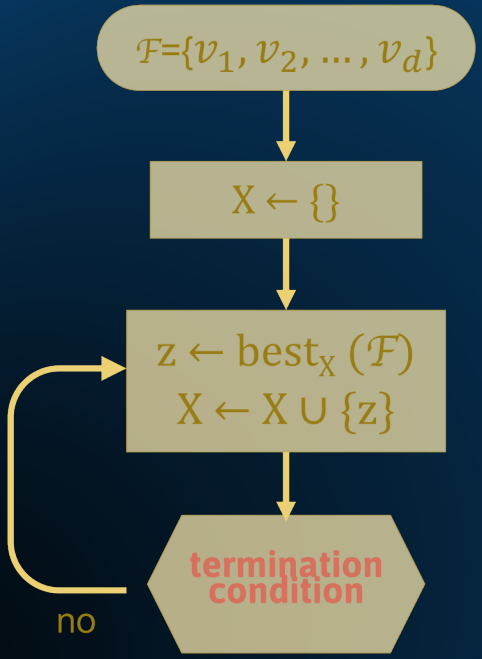


Search Strategies

Sequential Forward Selection



Sequential Forward Selection



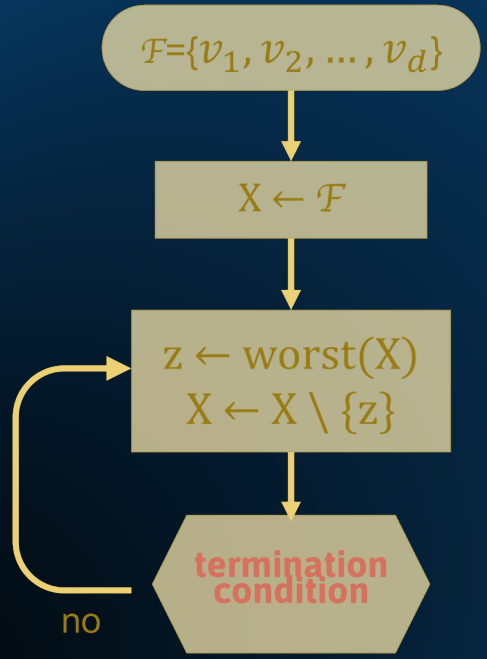
k=3
$X \leftarrow \{\}$

$X \leftarrow \{$ ★ $\}$

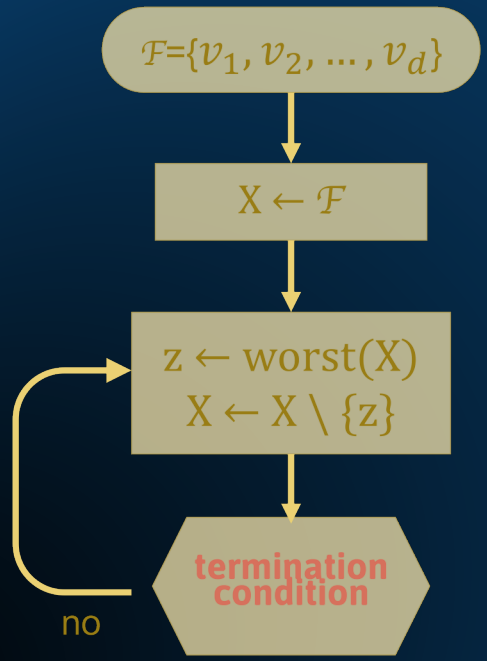
$X \leftarrow \{$ ★ ★ $\}$

$X \leftarrow \{$ ★ ★ ★ $\}$

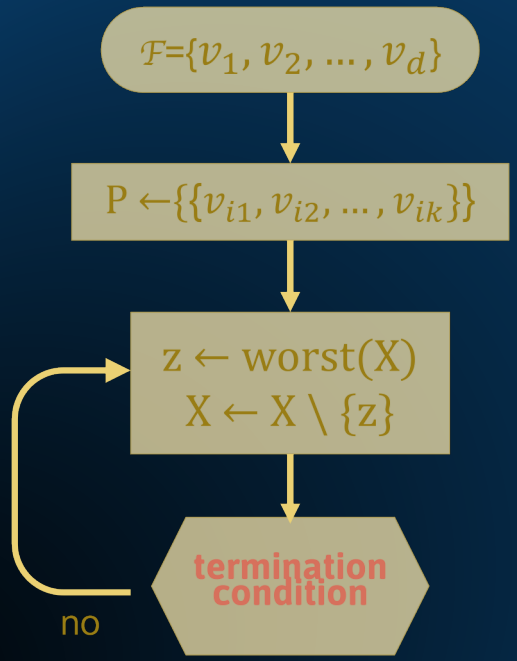
Sequential Backward Selection



Sequential Backward Selection

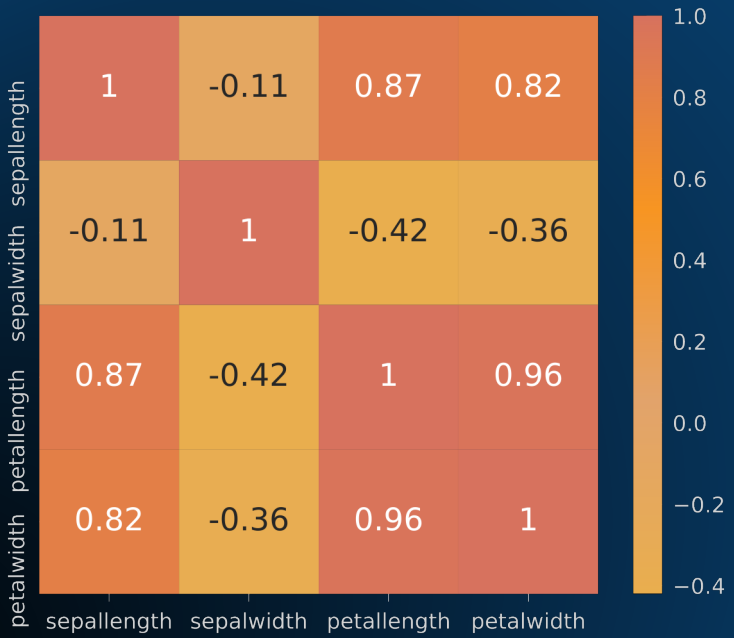


Genetic Algorithms Search

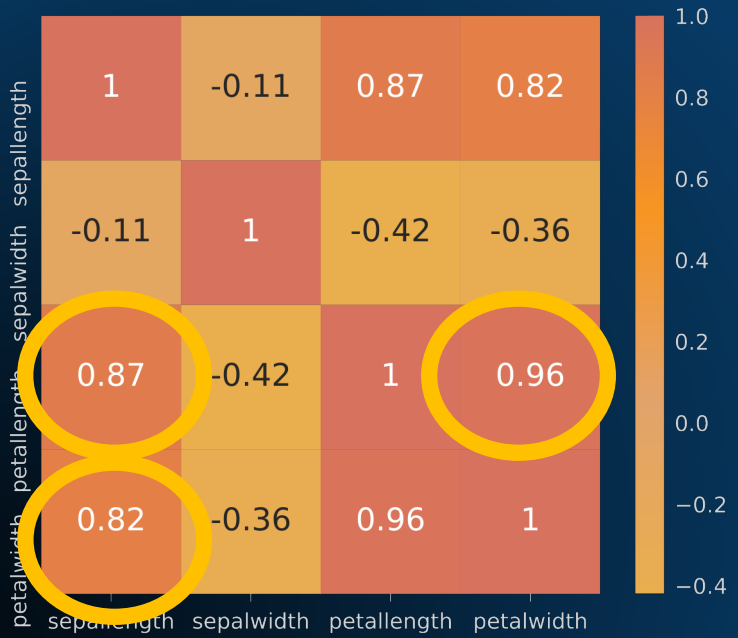


Unsupervised Selection

Unsupervised Selection



UNSUPERVISED SELECTION

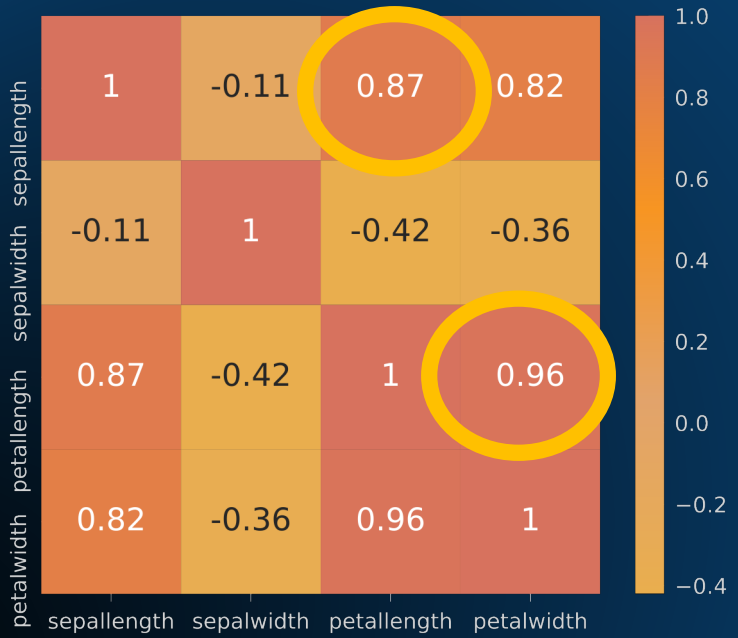


correlation >0.80

(petallength, petalwidth)
(sepallength, petallength)
(sepallength, petalwidth)

X ← {sepalwidth, petalwidth }

Unsupervised Selection

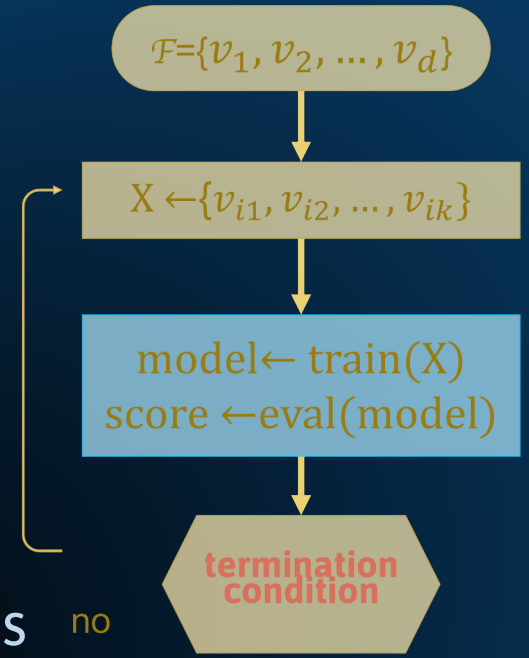


k=2

(petallength, petalwidth)
(sepallength, petallength)

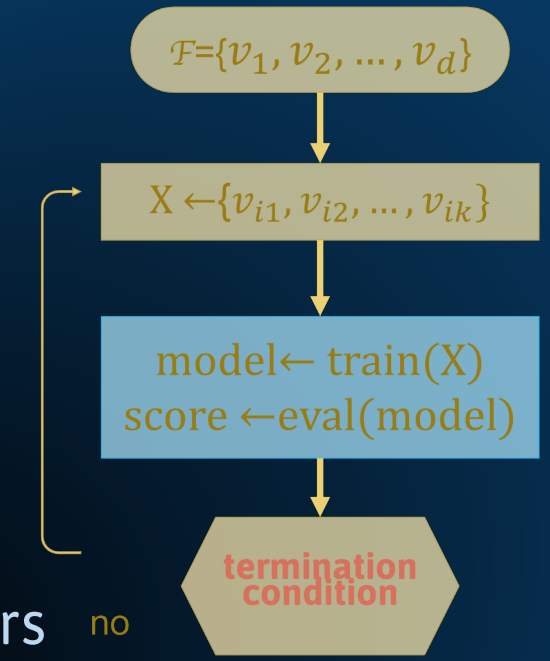
X ← {sepalwidth, petallength }

SUPERVISED SELECTION

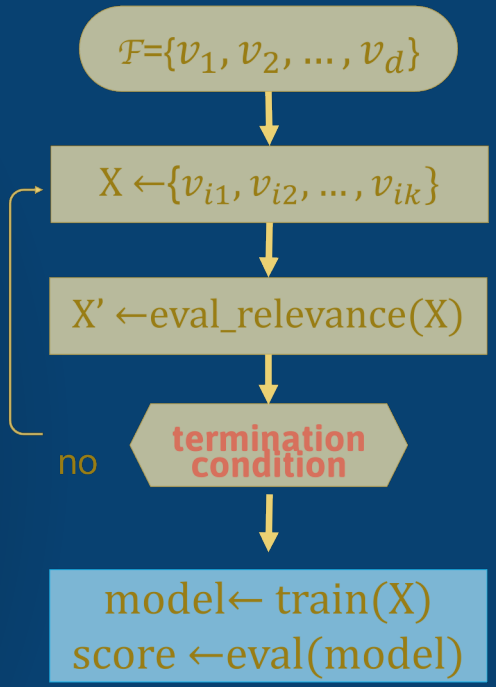


Wrappers

SUPERVISED SELECTION

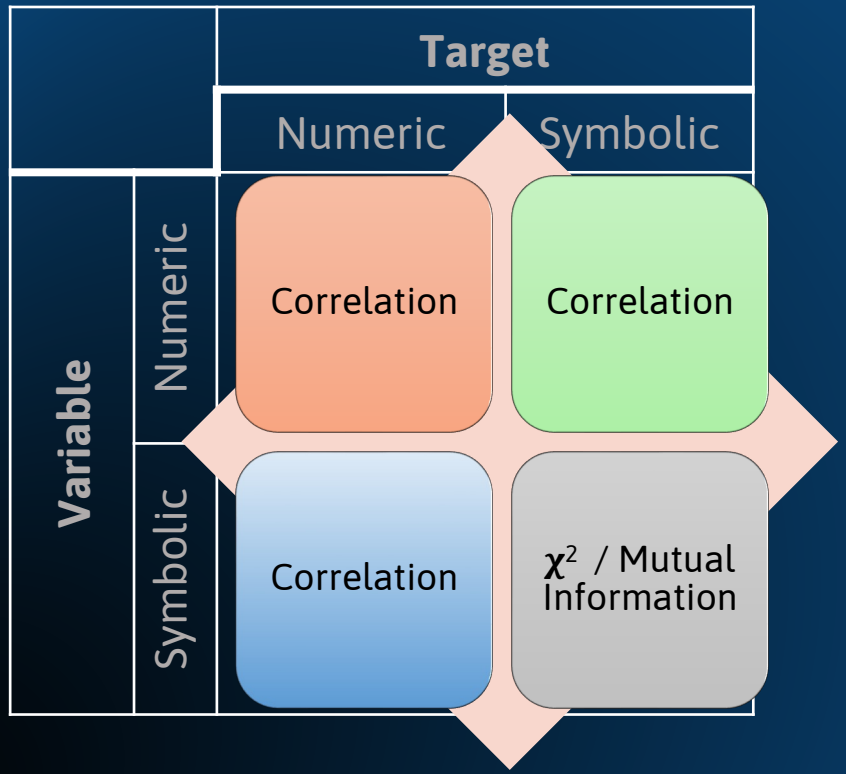


Wrappers



Filters

Relevance Measures



	Linear	Non-linear
Symbolic	ANOVA	Kendall's
Numeric	Pearson	Spearman

Thank you!

claudia.antunes@tecnico.ulisboa.pt