Pattern Classification and Recognition: Feature Selection

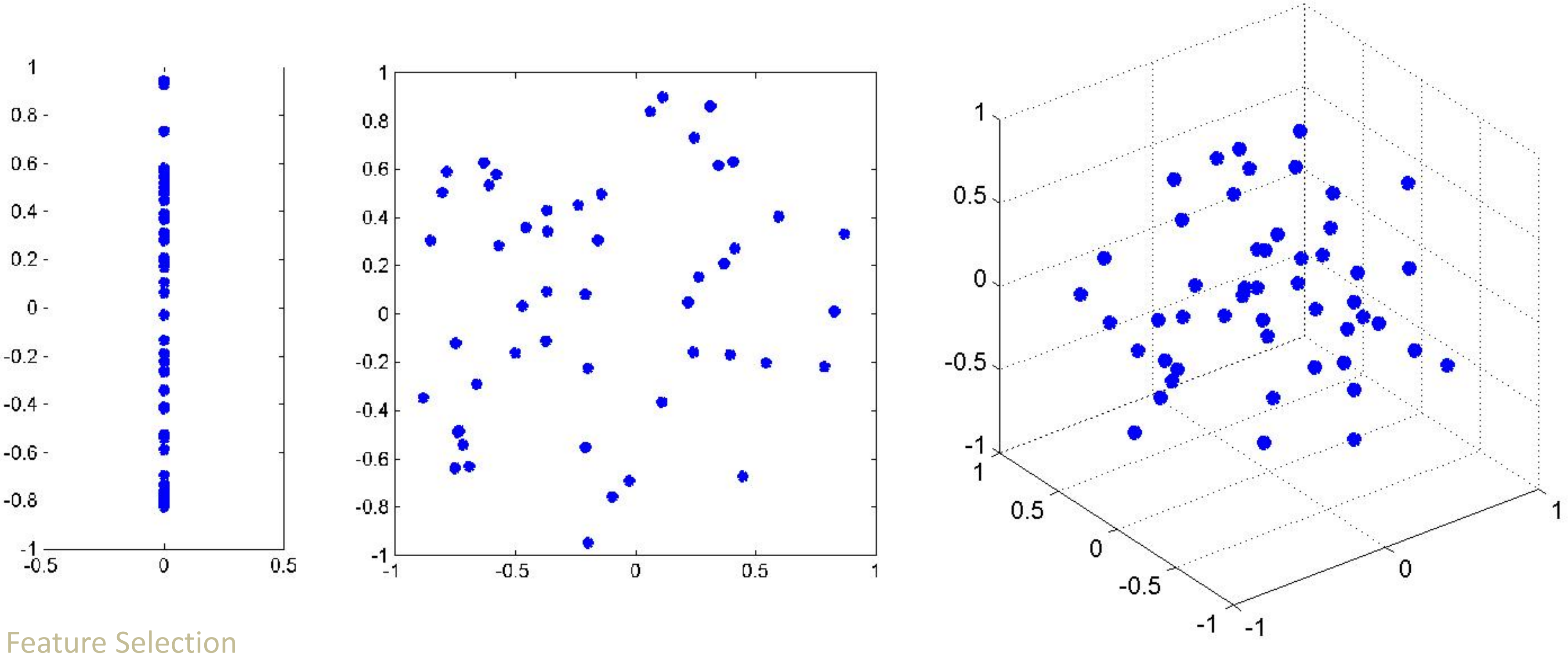
ECE 681

Spring 2016

Stacy Tantum, Ph.D.

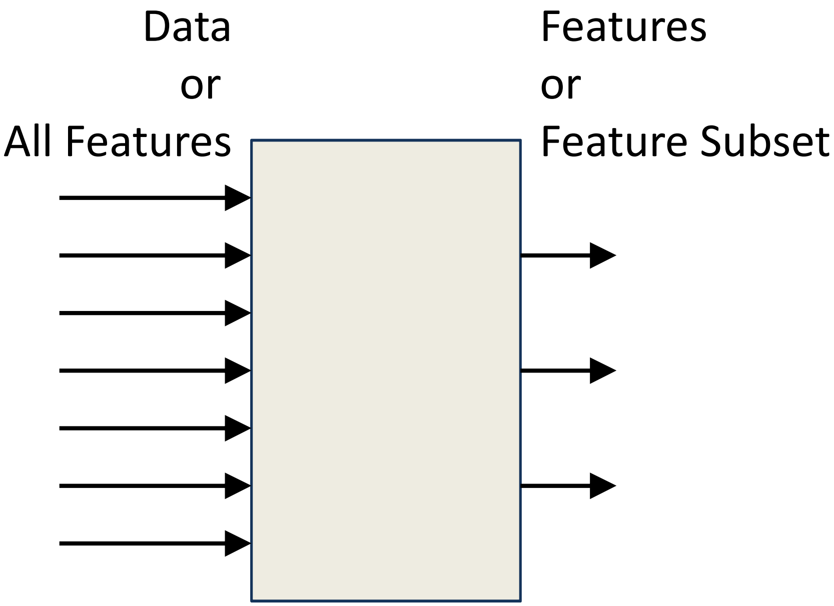
Curse of Dimensionality

50 data points, all within 1 unit of the origin (L_2-norm)



From Data to Features

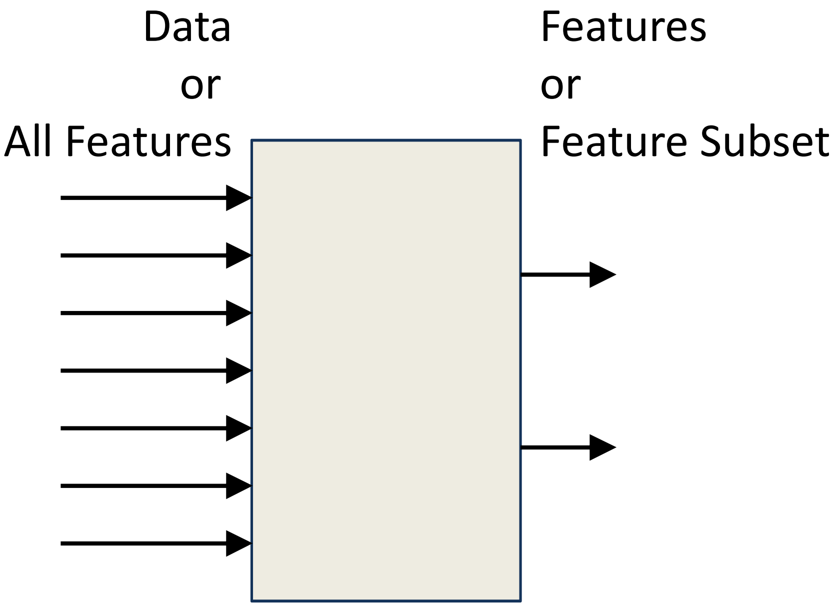
FEATURE SELECTION



Data
or
All Features

Features
or
Feature Subset

FEATURE EXTRACTION/GENERATION

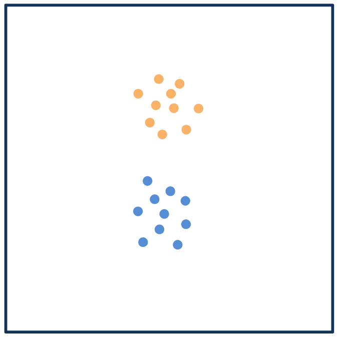


Data
or
All Features

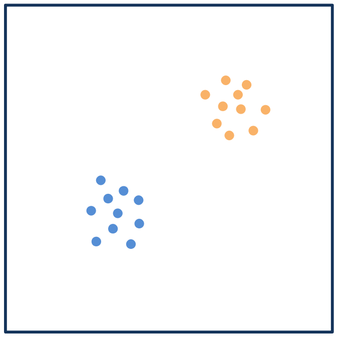
Features
or
Feature Subset

Desirable Feature Set Characteristics

MAXIMIZE RELEVANCY



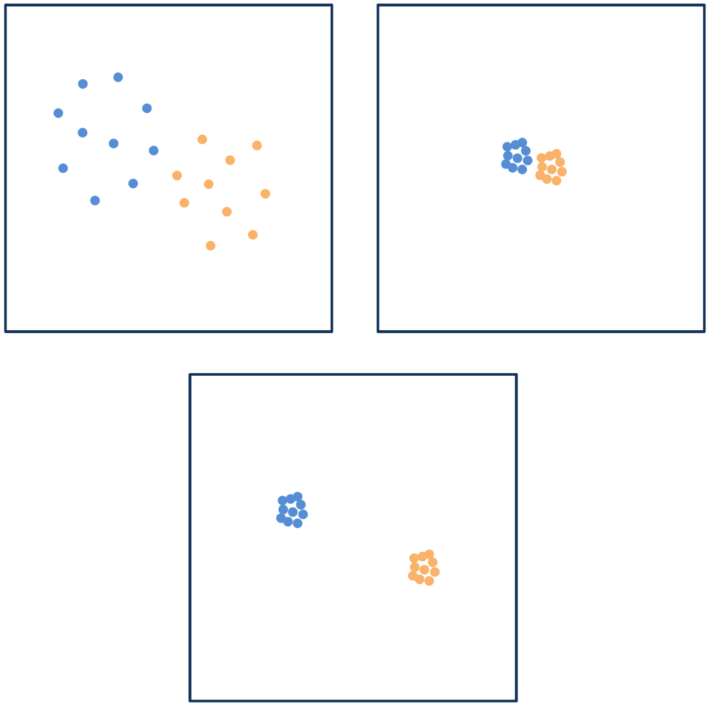
MINIMIZE REDUNDANCY



Feature Selection Philosophies

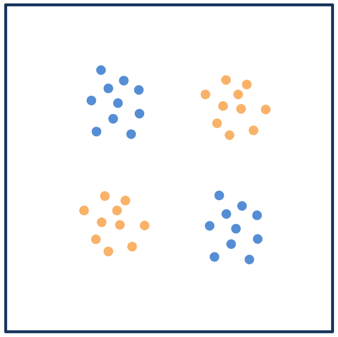
FILTER METHODS

Evaluate the class separability of the candidate feature sets independent of the classifier



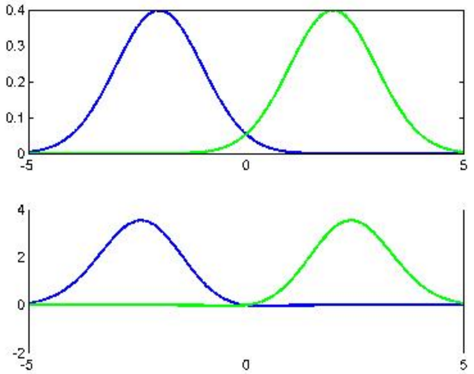
WRAPPER METHODS

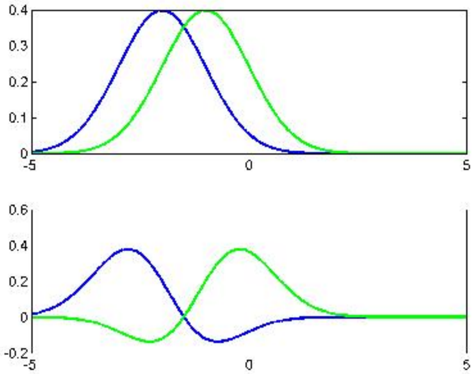
Evaluate classifier performance with the candidate feature sets

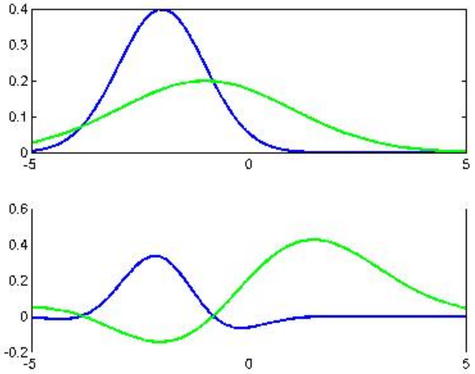


Divergence for Class Separability

Divergence compares pdfs





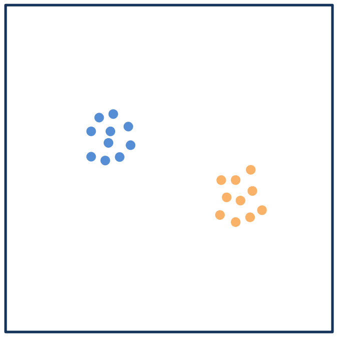


Scatter Matrices for Class Separability

Compute numerically

Goal is to simultaneously

- Minimize within-class scatter
- Maximize between-class scatter

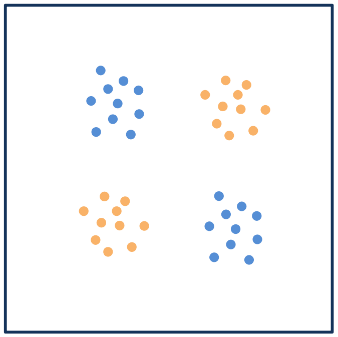


Feature (Sub)Set Selection

SCALAR SELECTION

Evaluate each feature independently

Feature	Quality Measure
Temperature	1.8
Humidity	2.6
Wind Speed	2.3



VECTOR SELECTION

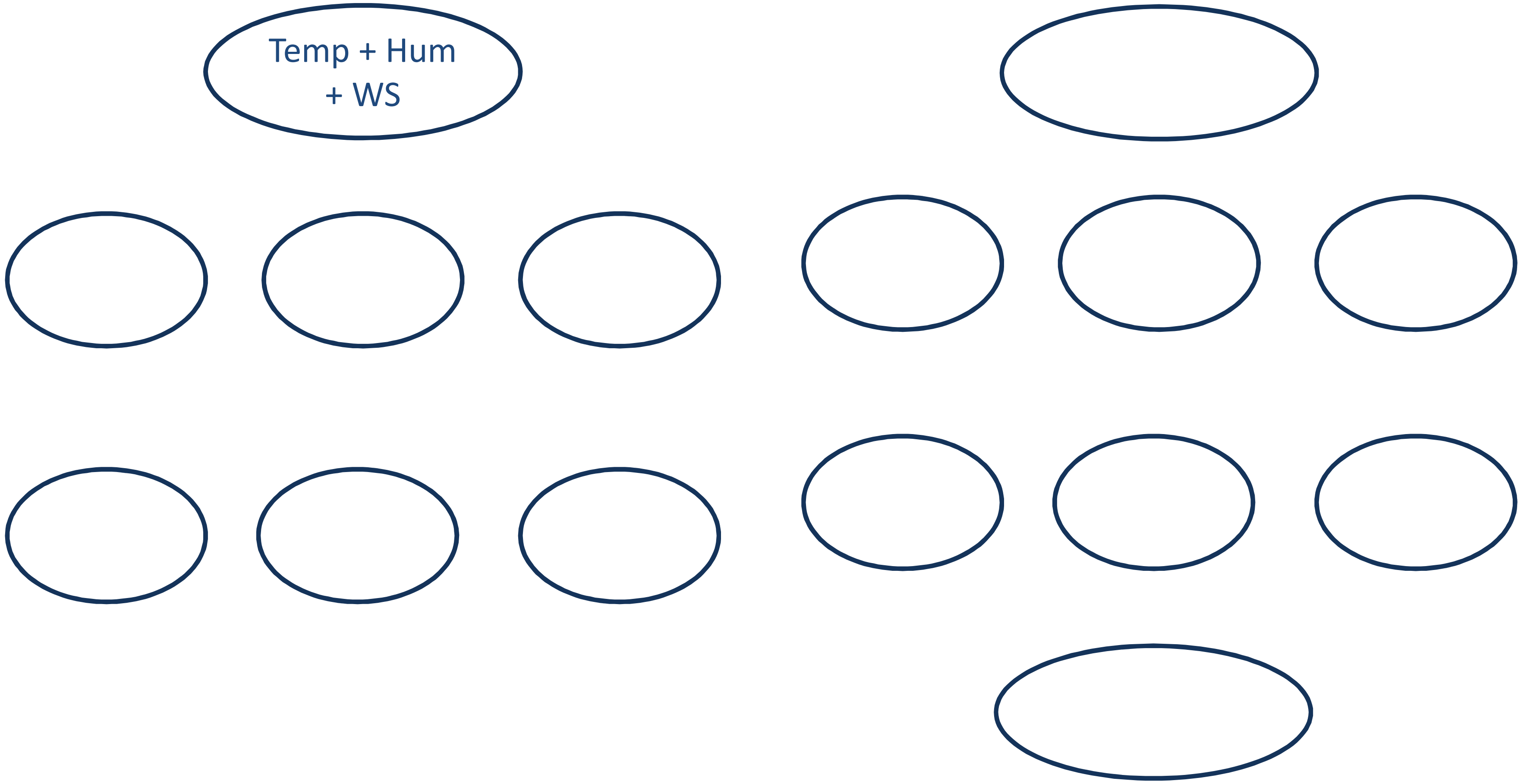
Evaluate all possible groups of features

Feature Subset	Quality Measure
Temp	1.8
Hum	2.6
WS	2.3
Temp + Hum	3.5
Temp + WS	2.4
Hum + WS	2.7
Temp + Hum + WS	3.1

Feature (Sub)Set Search

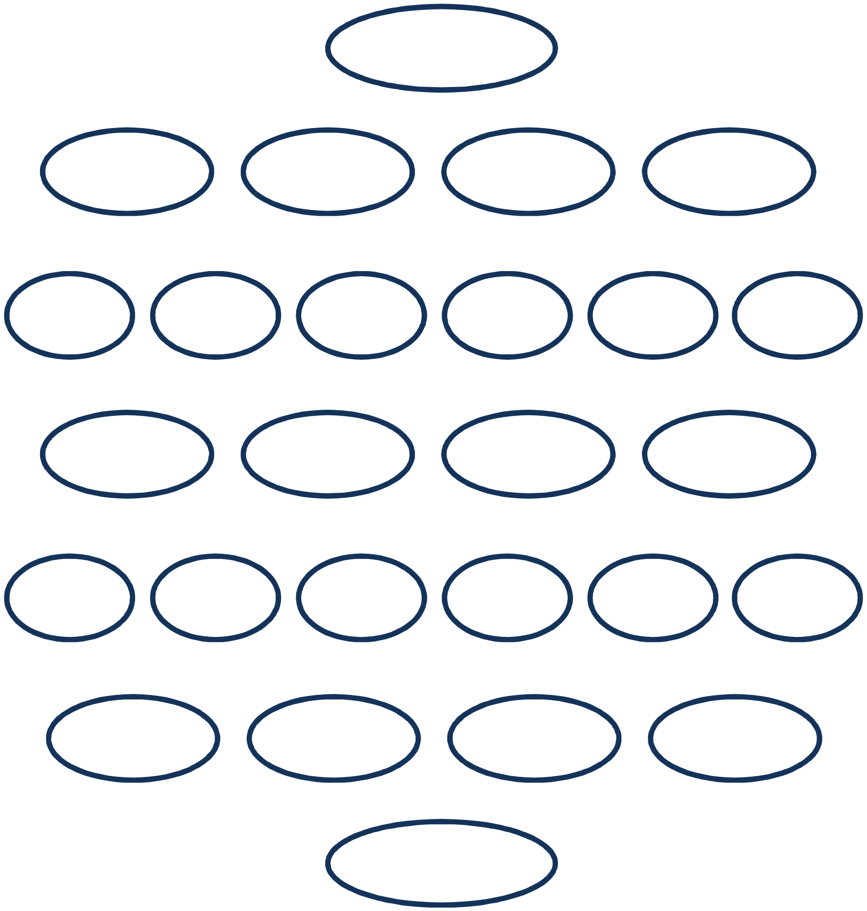
SEQUENTIAL BACKWARD SEARCH

SEQUENTIAL FORWARD SEARCH

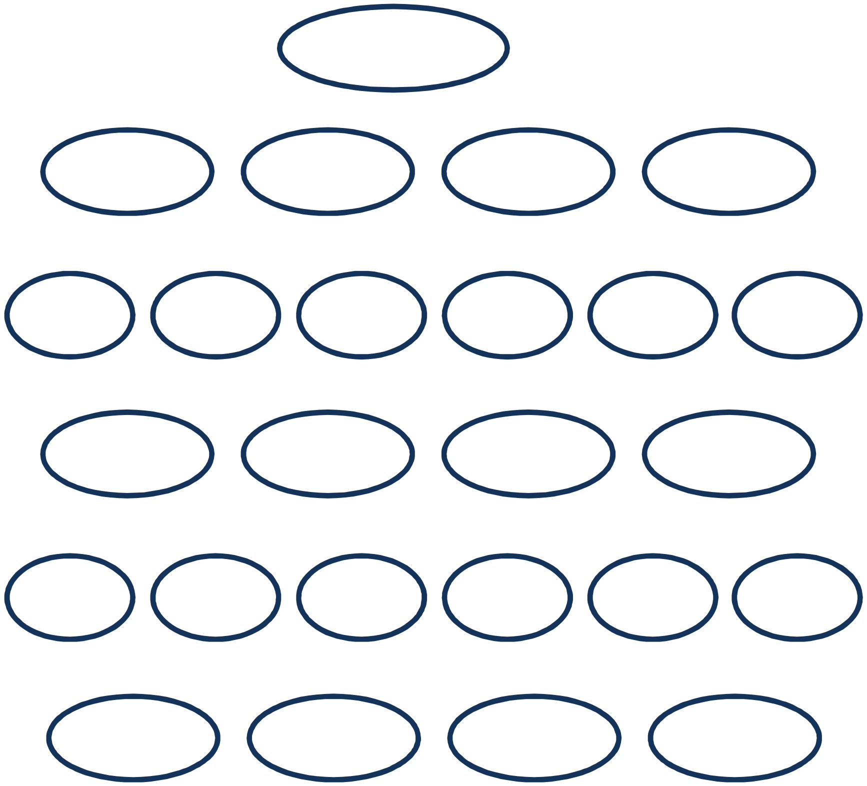


Feature (Sub)Set Search

FORWARD-BACKWARD SEARCH

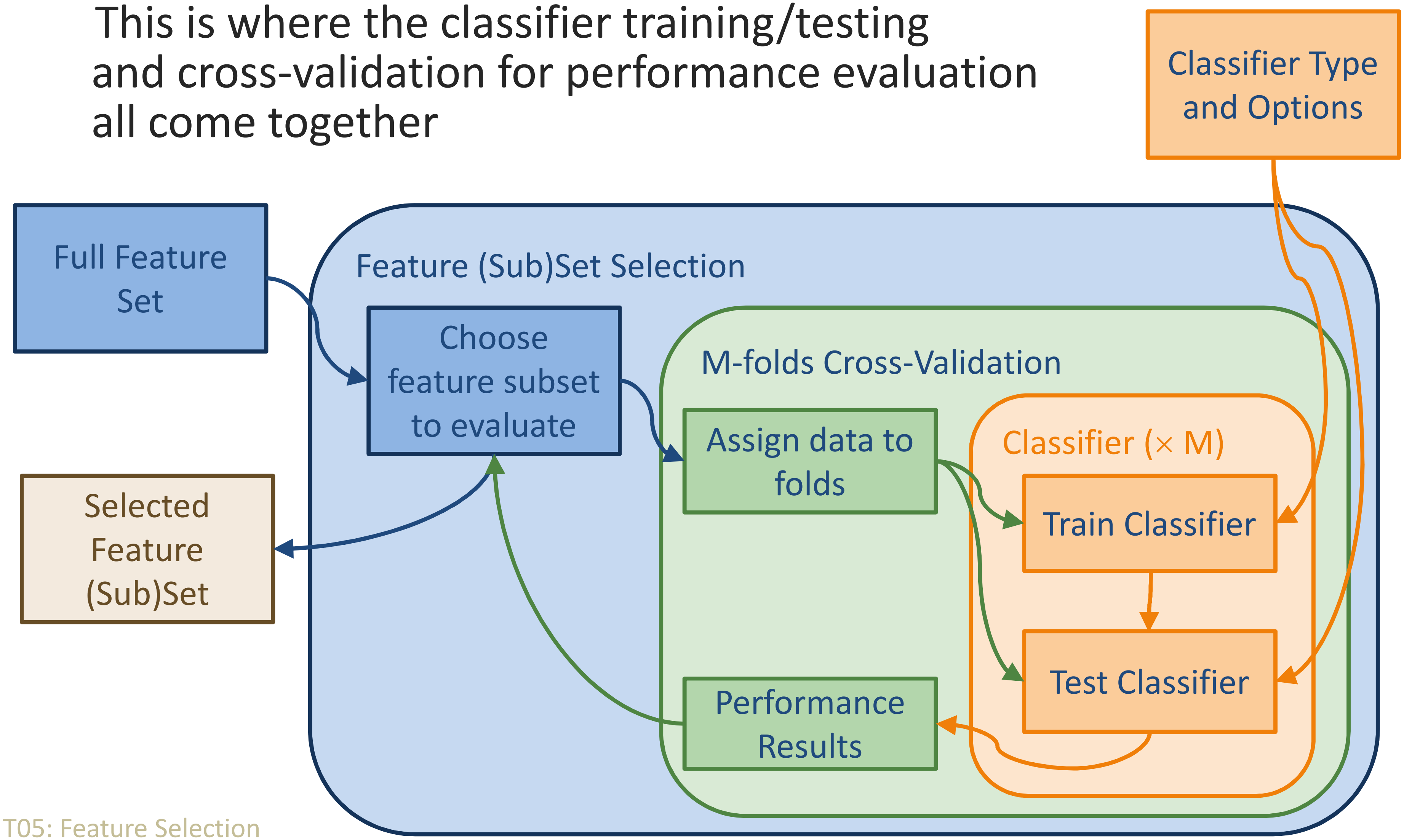


BACKWARD-FORWARD SEARCH



Feature Selection (by Search) Framework

This is where the classifier training/testing and cross-validation for performance evaluation all come together



Feature Selection (by Search) Coding Tips

The feature selection scheme is “wrapped around” classifier training and testing

Classifier training/testing should be done with cross-validation

You’ll need to decide how you are going to store performance for each feature subset

- Consider:
  - *Binary numbers are comprised of 0’s and 1’s*
  - *i.e., if features 1, 3, and 4 are tested, that is* $1101_2=13_{10}$

```
% Generate feature set to be tested
testFeatures = …;

% Select data corresponding to these features
dataSubset = data(:,testFeatures);

% Train/test classifier with dataSubset

% Select the next feature subset according to
your search scheme

% Repeat the process
```