Dynamic Classifier Selection Based on Imprecise Probabilities

Meizhu Li
Ghent University

Co-work: Jasper De Bock, Gert de Cooman

Outline

Dynamic classifier selection

Strategy of selection

Experiment results

Motivation

- Normally, one classifier is used for all the instances of the data set in a classification task.

- However, a classifier may only performs good on parts of instances, whereas another classifier performs better on other instances.

Dynamic Classifier Selection

- For each instance, select the classifier that is most likely to classify it correctly
- Use the result of the selected classifier to predict the class of that instance
- The combined classifier is expected to outperform each of the individual classifiers they select from.

Dynamic Classifier Selection

- For each instance, select the classifier that is most likely to classify it correctly
- Use the result of the selected classifier to predict the class of that instance
- The combined classifier is expected to outperform each of the individual classifiers they select from.

How to select an appropriate classifier for each instance?

Strategy of selection - Robustness measure

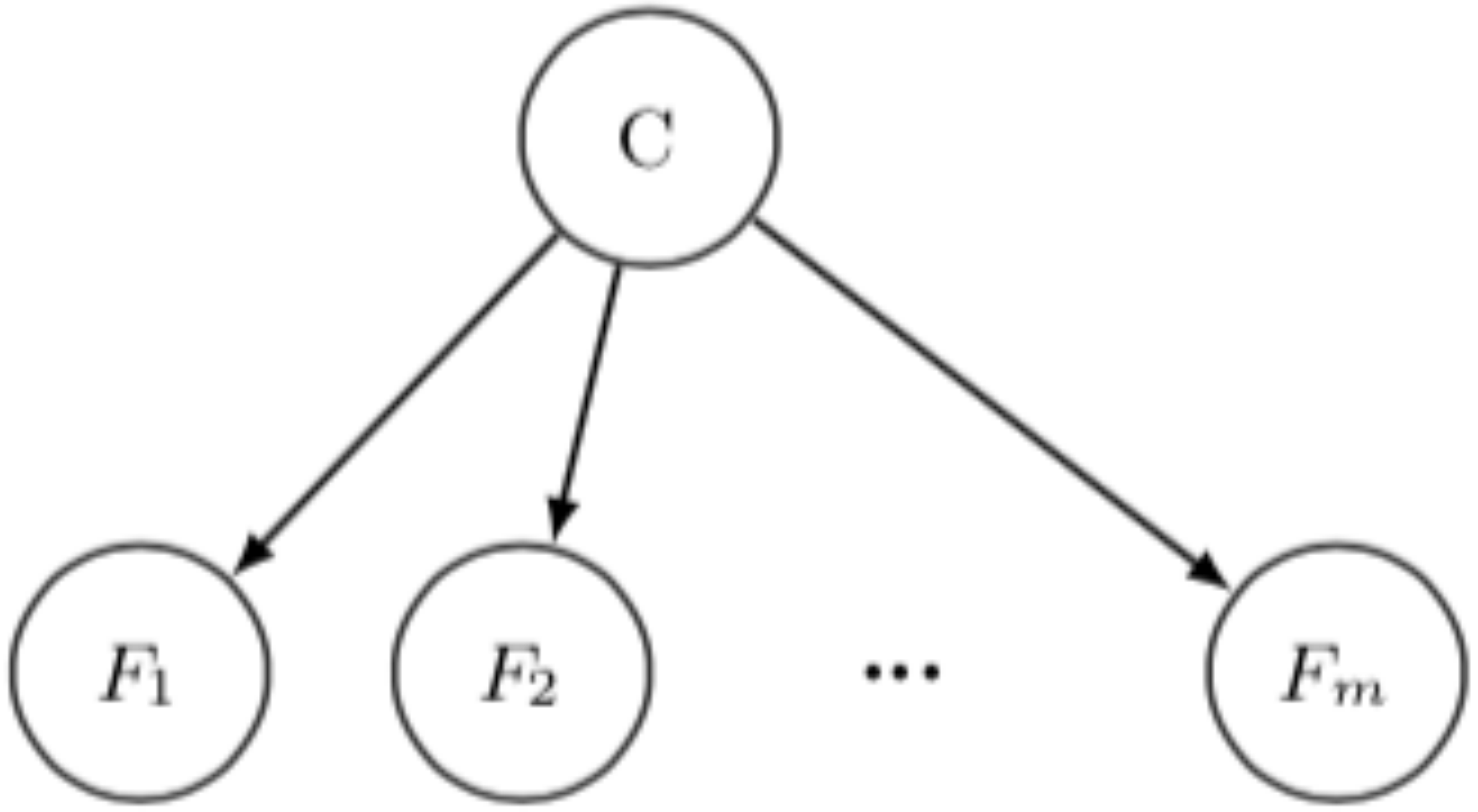


Fig.1 Example of a Naive Bayes Classifier

Let us denote C as the class variable, taking values c in finite set $\mathscr{C}$.

For each $c \in \mathscr{C}$, $\mathscr{P}(c)$ denotes a set of probability mass function P(c).

$$P(c) = \frac{n(c) + 1}{N + |\mathscr{C}|}$$

Strategy of selection - Robustness measure

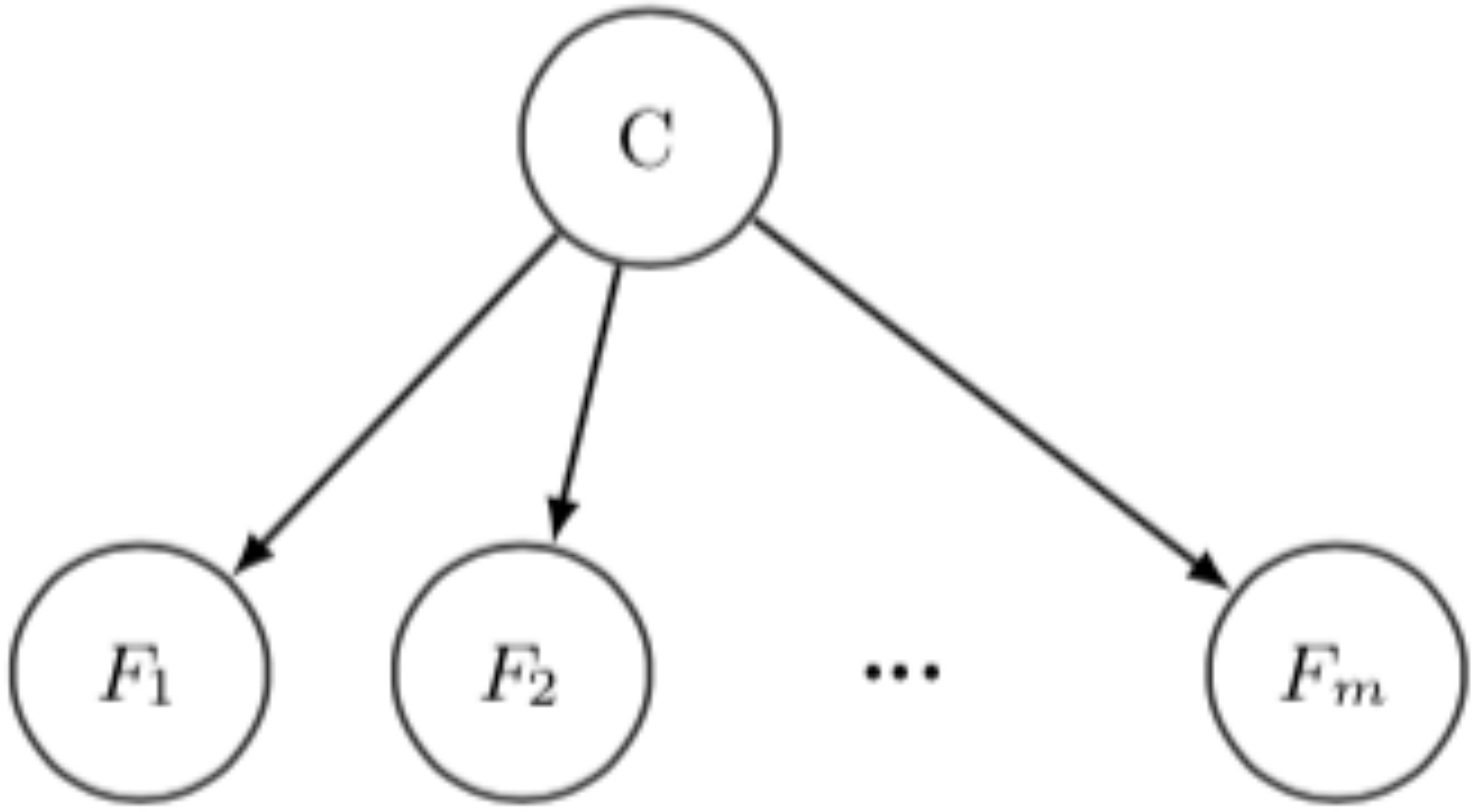


Fig.1 Example of a Naive Bayes Classifier

Let us denote C as the class variable, taking values c in finite set $\mathscr{C}$.

For each $c \in \mathscr{C}$, $\mathscr{P}(c)$ denotes a set of probability mass function P(c).

$$P(c) = \frac{n(c) + 1}{N + |\mathscr{C}|}$$

$$\downarrow$$

$$P(c) = \frac{n(c) + 1 + st(c)}{N + |\mathscr{C}| + s}$$, for all $c. \in \mathscr{C}$

where s is a fixed hyperparameter that determines the degree of imprecision, t is any probability mass functions on c.

Strategy of selection - Robustness measure

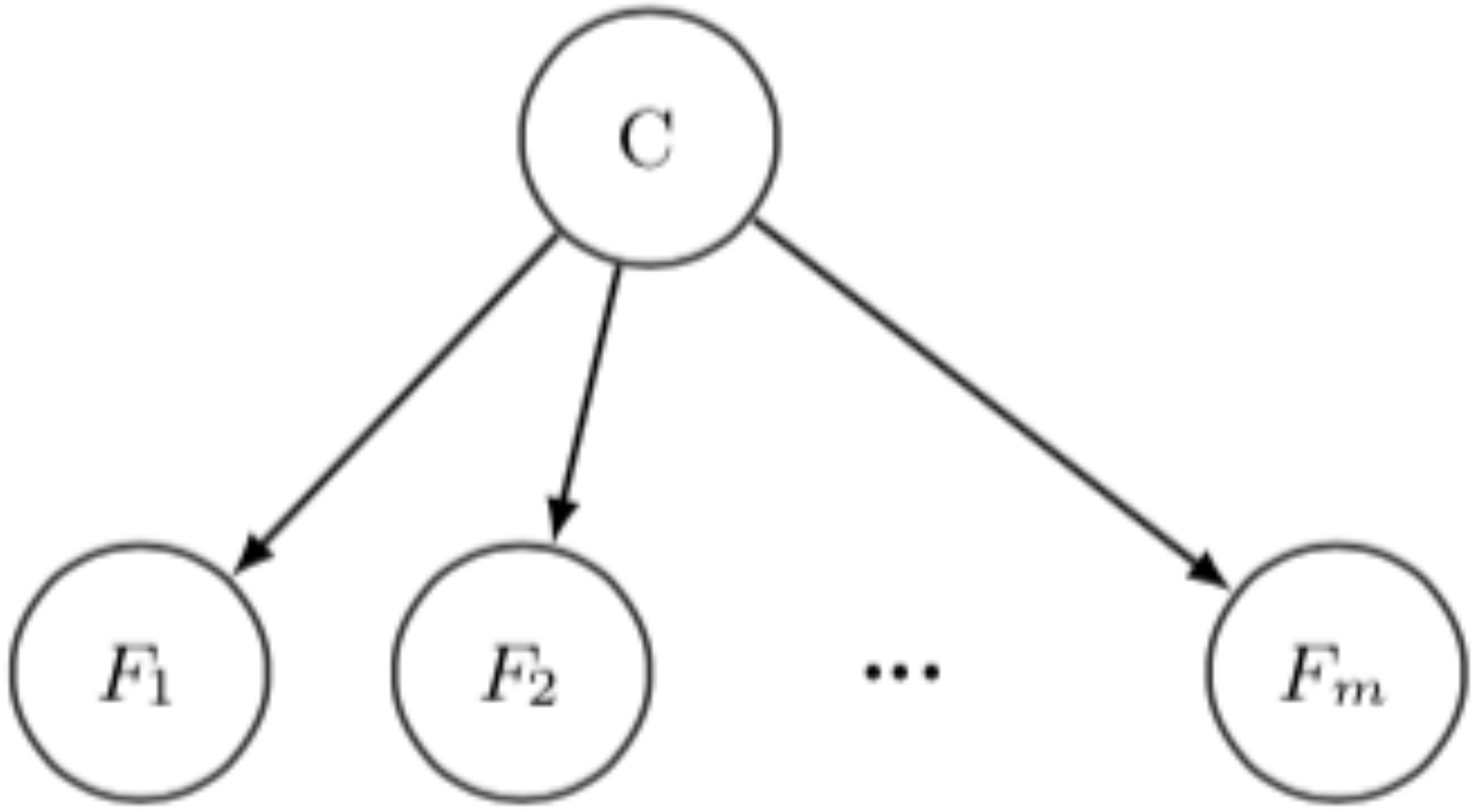


Fig.1 Example of a Naive Bayes Classifier

Let us denote C as the class variable, taking values c in finite set $\mathscr{C}$.

For each $c \in \mathscr{C}$, $\mathscr{P}(c)$ denotes a set of probability mass function P(c).

$$P(c) = \frac{n(c) + 1}{N + |\mathscr{C}|}$$

$$\downarrow$$

$$P(c) = \frac{n(c) + 1 + st(c)}{N + |\mathscr{C}| + s}$$, for all $c. \in \mathscr{C}$

Threshold:
the largest value of s that does not induce a change of prediction result.

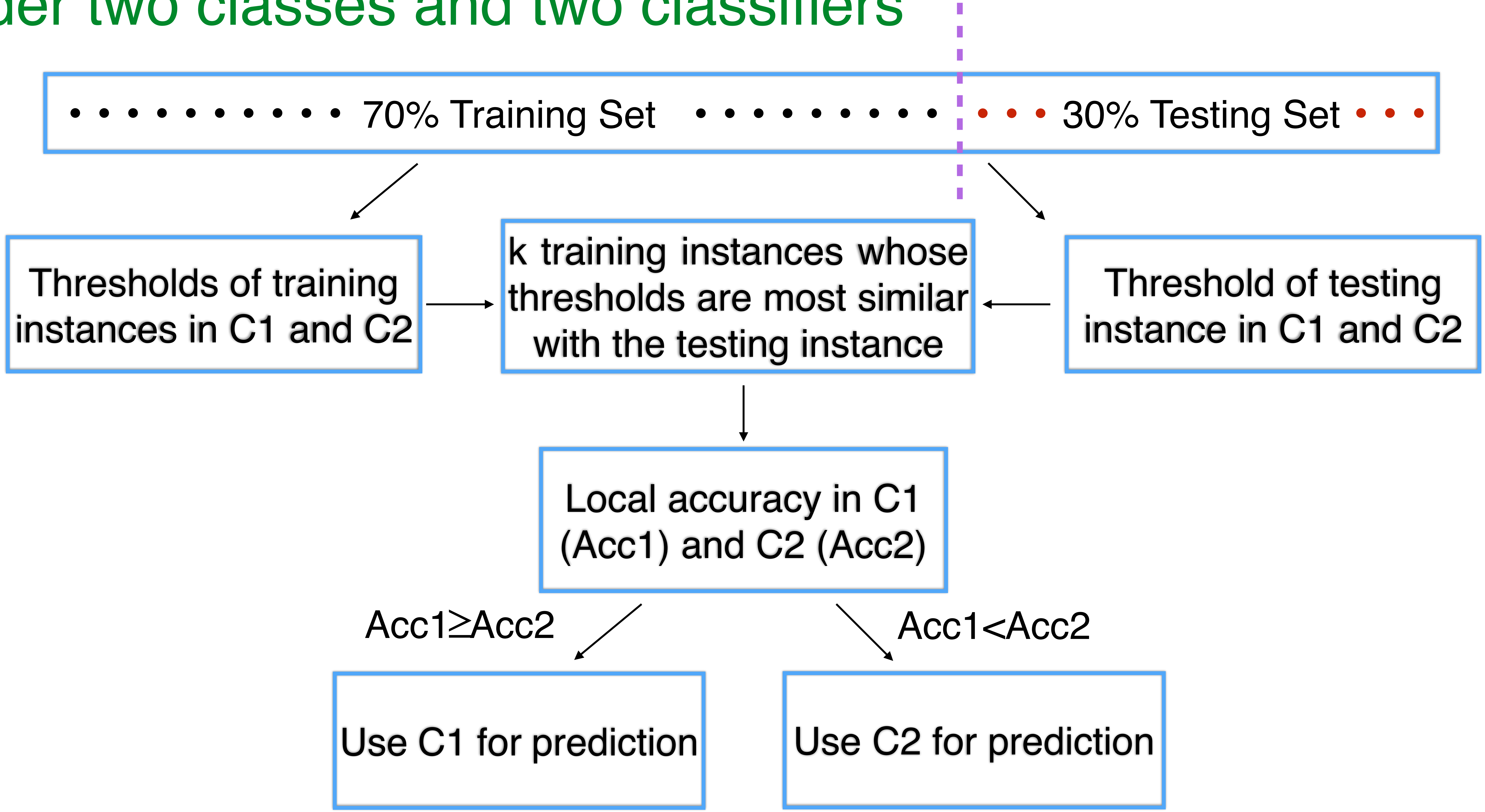
where s is a fixed hyperparameter that determines the degree of imprecision, t is any probability mass functions on c.

Strategy of selection - for the thresholds

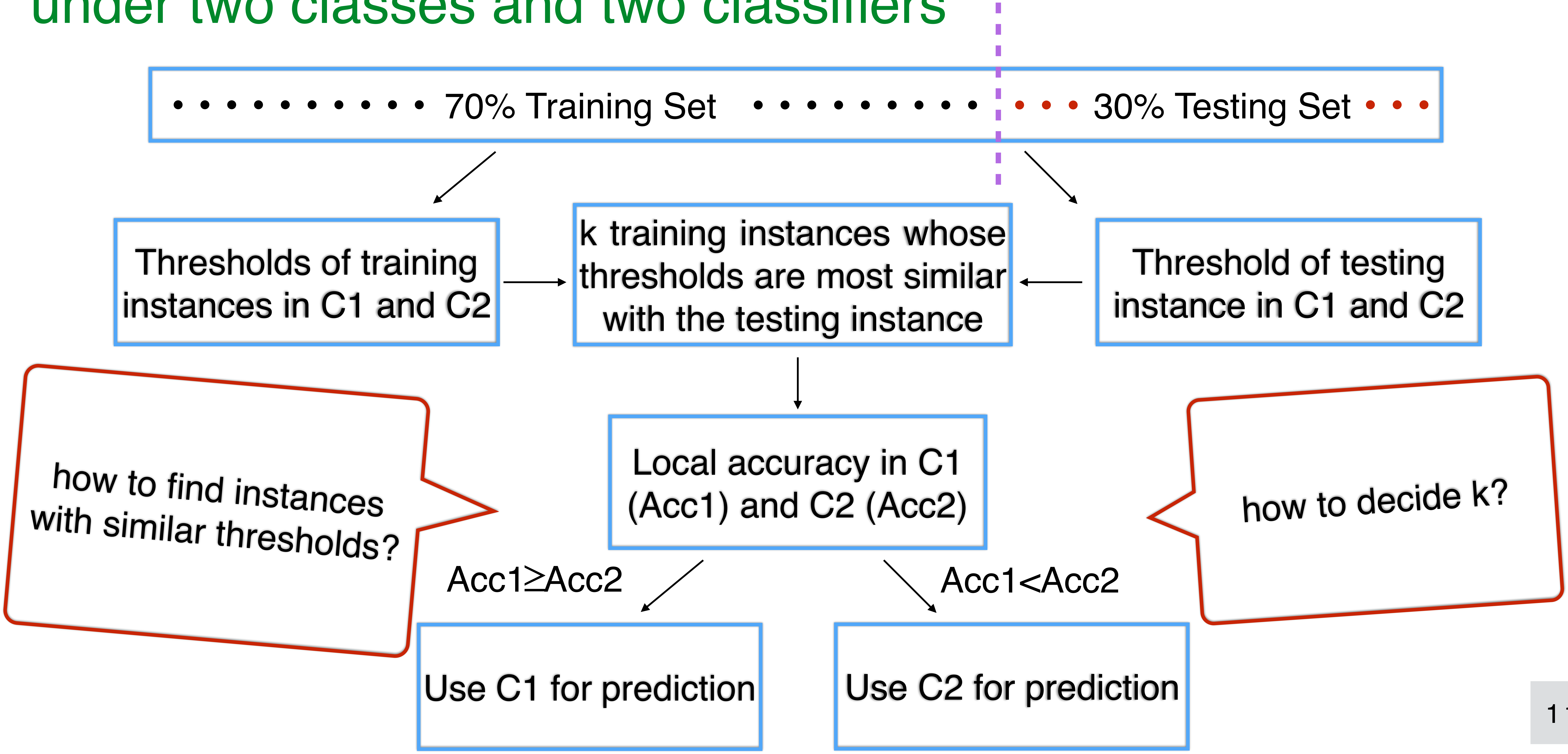
- Reference [1] provides an algorithm to calculate the thresholds by global sensitivity analysis for MAP inference in graphical models.
- Reference [1] also shows that instances with similar thresholds have a similar chance of being classified correctly.
- For every new test instance that is to be classified, we start by searching the training set that have a similar pair of thresholds.

[1] De Bock, J., De Campos, C.P., Antonucci, A.: Global sensitivity analysis for MAP inference in graphical models. Advances in Neural Information Processing Systems 27 (Proceedings of NIPS 2014), 2690–2698. (2014)

Strategy of selection:
under two classes and two classifiers



Strategy of selection: under two classes and two classifiers



Strategy of selection: distance between two instances

Thresholds	Classifier 1	Classifier 2	
● Testing instance	a_1	b_1	(a_1, b_1)
• Training instance 1	x_1	y_1	(x_1, y_1)
• Training instance 2	x_2	y_2	(x_2, y_2)
$\vdots$	$\vdots$	$\vdots$	$\vdots$
• Training instance n	x_n	y_n	(x_n, y_n)

Strategy of selection - illustration

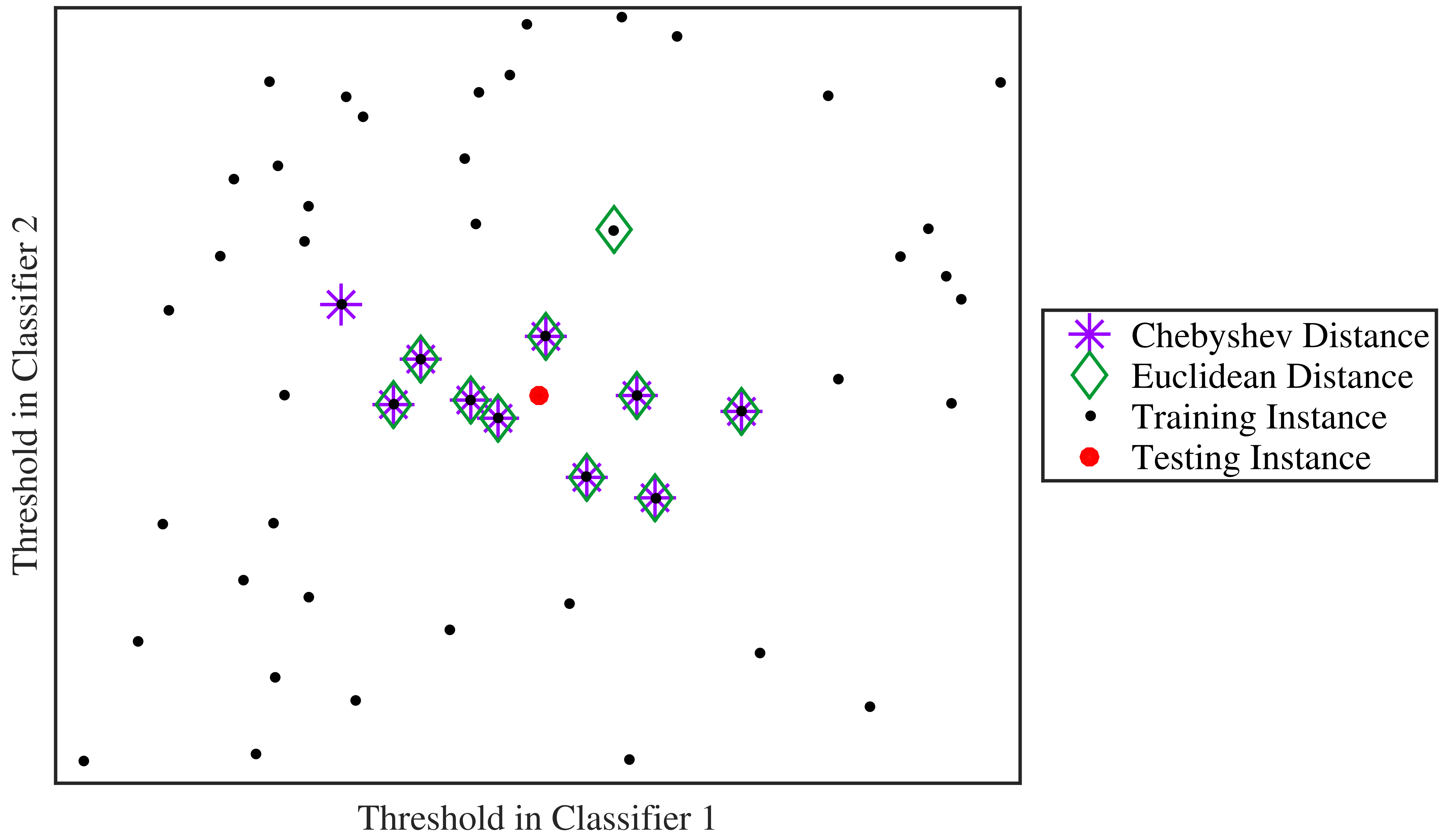


Fig. 1: Illustration of the chosen k-nearest instances, using a fictituous data set with fifty training points, and for k = 10 and two different distance measures

Experiments - Setting

- Five data sets from UCI repository[1].
- Feature selection: Sequential Forward Selection (SFS) method
  Classifier 1 (C1) and Classifier 2 (C2)
- Data with missing values were ignored.
  Continuous variables were discretized by their median

Table 1: Description of data sets

Name	# Data	# Class values	# Features	Features C1	Features C2
Balloons	76	2	4	$(1,2,3,4)$	$(1,3,4)$
BCW	699	2	9	$(1,2,5,6,8)$	$(1,2,6,8)$
ACA	690	2	14	$(3,4,6,13)$	$(1,3,4,6,7)$
MPG	398	2	8	$(1,5,6)$	$(1,4,5,6)$
TIC	958	2	9	$(1,5,7,8)$	$(1,2,5,7,8)$

[1] UCI Homepage, http://mlr.cs.umass.edu/ml/index.html.

Experiment result 1: Accuracy with different k value

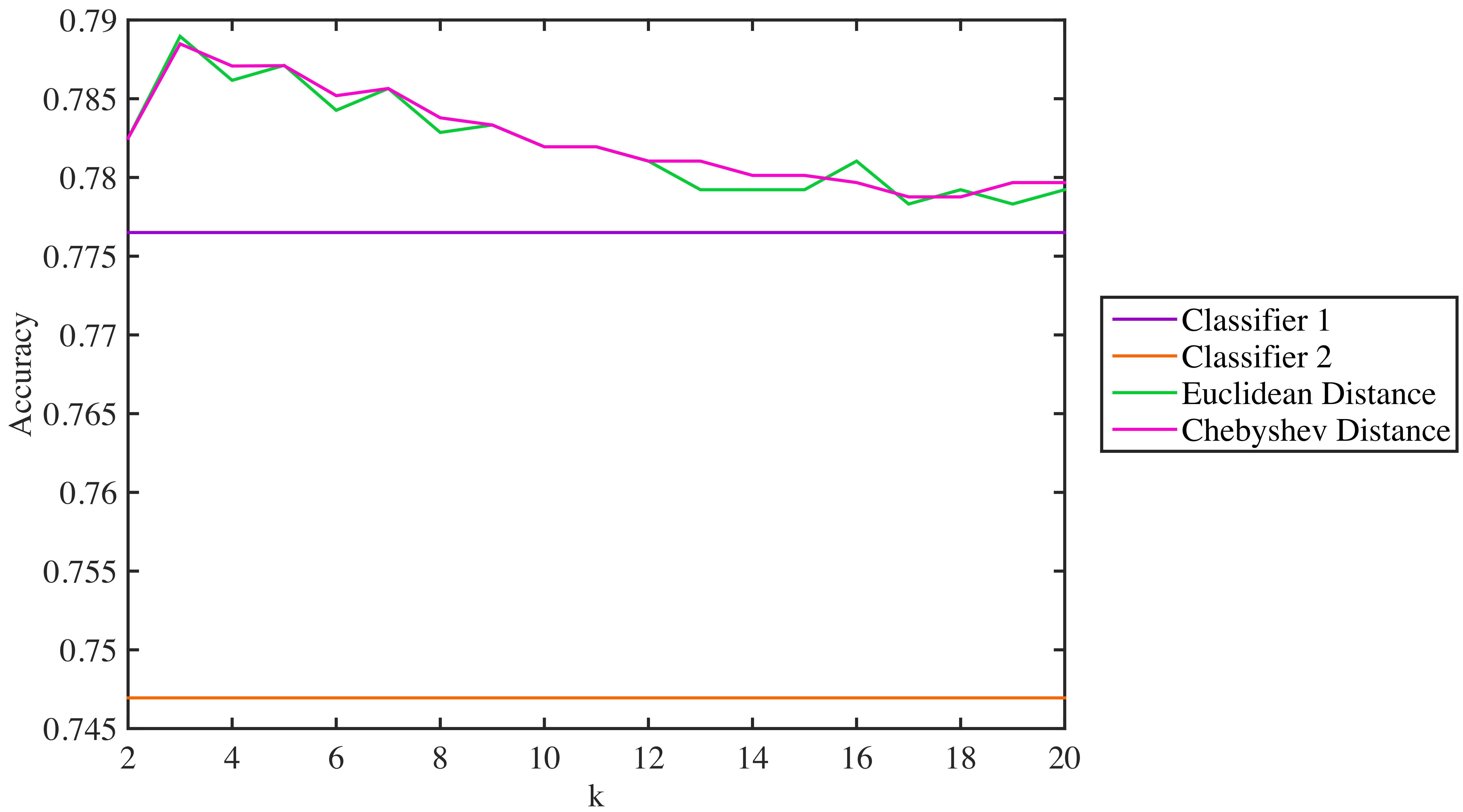


Fig. 2: The achieved accuracy as a function of the parameter k, for four different classifiers: the two original ones (which do not depend on k) and two combined classifiers (one for each of the considered distance measures)

Experiment result 2: with optimal k value

- For each run, an optimal value of k was determined through cross validation on the training set.

Table 2: A comparison of all four classifiers

Data Set	AC_{C1}	AC_{C2}	AC_{eu}	AC_{ch}
Balloons	0.776502	0.746948	0.781039	0.781970
BCW	0.974221	0.972496	0.974710	0.974710
ACA	0.723675	0.723190	0.724884	0.724884
MPG	0.920696	0.917610	0.921039	0.920697
TIC	0.724366	0.724363	0.733661	0.732731

- Our combined classifiers outperform the individual ones on which they are based.
- The choice of distance measure seems to have very little effect.

Summary

- The imprecise-probabilistic robustness measures can be used to develop dynamic classifier selection methods that outperform the individual classifiers they select from.

Future work

- Deepen the study of the case of the Naive Bayes Classifier.
- Other strategy of selection: weighted counting…
- Compare our methods with other classifiers such as Lazy Naive Credal Classifier.

Thank you!

MEIZHU LI
GHENT UNIVERSITY

meizhu.Li@ugent.be